**NTNU – Trondheim**
Norwegian University of
Science and Technology

# An Approach to planning Data Collection for IAM in VMW

D1-2012-27

## Marius Møller Rokstad

Civil and Environmental Engineering
Submission date: June 2012
Supervisor: Rita Maria Ugarelli, IVM
Co-supervisor: Sveinung Sægrov, IVM

Norwegian University of Science and Technology
Department of Hydraulic and Environmental Engineering

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF HYDRAULIC AND ENVIRONMENTAL ENGINEERING

<table>
<tr><td rowspan="2">Report Title:<br>An Approach to planning Data Collection for Infrastructure Asset Management in VMW (D1-2012-27)</td><td colspan="4">Date: June 10th 2012<br>Number of pages (incl. appendices): 149</td></tr>
<tr><td>Master Thesis</td><td>x</td><td>Project Work</td><td></td></tr>
<tr><td colspan="5">Name: Stud.techn. Marius Møller Rokstad<br>Contact information: marius.rokstad@gmail.com or + 47 980 36 599</td></tr>
<tr><td colspan="5">Supervisor: Adj. prof. Rita Maria Ugarelli (PhD, SINTEF/NTNU)</td></tr>
<tr><td colspan="5">Co-supervisor: Prof. Sveinung Sægrov (PhD, NTNU)</td></tr>
</table>

# I Abstract

Ageing infrastructure and changing external conditions has invoked the need for performing proactive and predictive asset management (AM) in water utilities through the utilisation of AM tools. AM tools can act as decision support for maintenance and renewal decisions in a water utility, through assessing decision-influencing factors such as criticality, performance, condition, failure rates, hydraulic and economic risk, life cycle cost etc. The tools for AM are dependent on data about the characteristics of the assets (inventory) as well as record data from the assets' life cycles.

The Flemish Water Company (Vlaamse Maatschappij voor Watervoorziening, VMW) is currently in a transition from a reactive to a predictive AM strategy, and is performing this shift by the implementation of several AM tools. However, VMW is a company in which data collection has been executed to a very limited extent in the past, and will therefore be challenged by the lack of available data when implementing these tools for AM.

The project described in this document has endeavoured to assess how VMW should (1) ensure that their AM tools are not impeded by data availability and low data quality, and (2) how the collected data can be utilised optimally for AM. These two assessments were made through five work areas (WA): a comprehensive literature study on which factors influence data quality, an analysis of the data needs for buried assets followed by an analysis of the technical and organisational factors in VMW, a case study in *Trondheim Vann og avløp* and a cost-benefit analysis of data collection. The scope has been focused on data that is relevant for describing the life cycle of buried assets.

It was found that the quality of data is likely to be impeded by the dispersed storage structure that is imposed by the AM tools' proprietary databases. Therefore, it has been suggested to create one central data repository for all life cycle data (a diary database). The database should work as a distributor of data to the different data-consuming applications. Suggestions for the structure and development of this database were made.

Further was a comprehensive list of measures for how to stimulate operational personnel to collect data produced. Among the measures were the establishment of data quality indicators, data quality monitoring and communication routines, training, and early data collection methods.

Lastly, the cost-benefit analysis showed that VMW could achieve much higher informational benefits if they invest 7.5 % more in data collection (compared to current long-term plans), allowing for tools that assess hydraulic and economic risk of failure to be used (hydraulic criticality, expected unmet demands, combined with cost of failures). The cost-benefit analysis also showed that VMW should search for tools that utilise inspection data better, and that the benefits of extended data about asset interventions should be tested (in selected "trial" service centres) before being collected in full scale.

For more details about the results and conclusions, please confer with the Executive summary (chapter VII).

A paper was made on the cost-benefit analysis of data collection, and is enclosed in Appendix E. It is suggested to develop the cost-benefit analysis further by (1) measuring the unit costs of data collection and (2) expressing the benefits as reduced economic risk.

Keywords:

1. Infrastructure asset management
2. Data collection
3. Cost-benefit
4. Buried assets

_______________________________________

# II Sammendrag (norsk/Norwegian)

Aldrende infrastruktur og forandring i rammebetingengelsene for drift av VA-anlegg har økt behovet for proaktiv og prediktiv forvaltning (asset management, AM) i vannforsyningssektoren, gjennom bruk av AM-verktøy. Slike verktøy kan bidra som beslutningsstøtte for VA-anleggseiere ved å modellere og kvantifisere faktorer som kan ha innflytelse på avgjørelser om vedlikehold og fornyelse, for eksempel kritikalitet, tilstand, ytelse, bruddfrekvens, hydraulisk og økonomisk risiko, levetidskostnader etc. Verktøy for AM er avhengige av data om egenskaper til komponentene i ledningsnettet, samt en historikk av data som beskriver hendelsene komponentene har gjennomgått i løpet av livssyklusene deres.

Det Flamske Vannforsyningsselskapet (Vlaamse Maatschappij voor Watervoorziening, VMW) er nå i en overgangsfase fra en reaktiv til en prediktiv forvaltningsstrategi. Denne overgangen gjennomfører VMW ved å implementere forskjellige AM-verktøy. VMW er en bedrift hvor datainnsamling kun har vært gjort i en begrenset grad, og suksessen av AM-verktøyene vil derfor være begrenset av den manglende tilgangen til data som beskriver livssyklusen til ledningsnettet.

Dette prosjektet har forsøkt å vurdere hvordan VMW burde (1) sikre at resultatene fra AM-verktøyene deres ikke blir begrenset av datatilgjengelighet eller data av lav kvalitet, og (2) hvordan VMW kan sørge for at innsamlet data fra ledningsnettet blir brukt optimalt. Disse to spørsmålene har blitt vurdert i fem arbeidsområder (work areas, WA): en litteraturstudie med fokus på hvilke faktorer som påvirker datakvalitet, en analyse av livssyklusen til en nedgravd komponent (buried asset) fulgt av en analyse av tekniske- og organisasjonsfaktorer, en tilfellestudie av *Trondheim Vann og avløp*, og en kost-nyttevurdering av datainnsamling. Studien har begrenset seg til data som er relevant for å beskrive livssyklusen til nedgravde komponenter.

Studien har vist at datakvaliteten i VMW er forventet til å bli begrenset av en spredd datastruktur, fordi AM-verktøyene har egne databaser i forskjellige format. Derfor har det blitt foreslått å opprette én sentral database for all livssyklusdata (en dagbok). Forslag for innhold og struktur i denne dagboken er også blitt foreslått.

Videre har det blitt utarbeidet en omfattende liste med tiltak for hvordan VMW kan stimulere driftspersonell til å samle inn nødvendige data. Blant disse tiltakene er utarbeidelse av datakvalitetsindikatorer, datakvalitetsovervåking, opplæring, og metoder for datainnsamling og databruk i oppstartsfasen beskrevet.

Til slutt viste kost-nyttevurderingen at VMW kan oppnå et mye høyere nivå av informasjon dersom det blir investert 7.5 % mer på datainnsamling (i forhold til nåværende langtidsplan) som gjør det mulig å modellere hydraulisk og økonomisk risiko for brudd. Kost-nyttevurderingen viste også at VMW burde undersøke muligheter for å øke nytteverdien av inspeksjons- og lekkasjesøkedata, og at innsamling av «utvidet» data (eksterne faktorer m.m.) bør testes for nytteverdi ved enkelte driftssentraler før det samles inn i hele virksomheten

For mer detaljer om resultatene og konklusjonen, se Executive summary (kapittel VII).

En artikkel om kost-nytteverdivurderingen av datainnsamling er også blitt skrevet, og er vedlagt i Appendix E. Det er foreslått å utvikle kost-nyttemodellen videre ved å uttrykke nytteverdien som redusert økonomisk risiko.

# III Preface

The project described in this text was performed in January to June 2012 as a collaboration project between Vlaamse Maatschappij voor Watervoorziening (VMW), The Norwegian University of Science and Technology (NTNU) and the candidate.

The candidate, for the degree Master of Science, is Marius Møller Rokstad. The degree is the final component of the NTNU integrated Master's programme in Civil and environmental engineering. The thesis is administered by the Department of Hydraulic and Environmental Engineering at NTNU, where the candidate has been specialising in water and sewerage engineering.

The main adviser for this project has been Rita Ugarelli (PhD), researcher at SINTEF and adjunct professor at NTNU. Co-adviser has been Sveinung Sægrov (PhD), professor at NTNU.

VMW is a drinking water production and distribution utility, providing drinking water to approximately 2.78 million customers in the Flemish region of Belgium. VMW has 30 496 km of water distribution pipes, and provide water from 77 water production centres, mostly groundwater. VMW is active in the provinces of West-Vlaanderen, Oost-Vlaanderen, Vlaams-Brabant and Limburg. (VMW, 2010, VMW, 2012a).

# IV Acknowledgments

The advisers, Rita Ugarelli and Sveinung Sægrov, are thanked for following up the work in the project on a continuous basis, and for returning constructive and necessary feedback through the progression of the project.

The work group in VMW has consisted of Mr Luc Keustermans, Mr Edward Diericx, Mr Frederik Looten and Mr Gert Danckaers. They are all thanked for participating at meetings, discussing, giving feedback on the work, and for providing the necessary background data for the project.

Ingrid Selseth from SINTEF is thanked for providing reports that have been used in the Trondheim case study. Also, Trond Ellefsen, Rannveig Høseggen, Halvard Kierulf and Olav Nilsen from *Trondheim Vann og avløp* are thanked for answering the questionnaire made for the Trondheim case study.

Marius Møller Rokstad
June 2012, Leuven, Belgium

# V Table of contents

# VI List of figures

# VII List of tables

# VIII Executive summary

## i) Purpose

This project has had two overall purposes. The first was to examine how VMW should collect data about their buried assets in order to be able to manage them. The second was to examine how the data could be utilised optimally.

The main questions to be answered in this project were: What possible sources of data could be used to monitor the life cycle of a buried asset? How can a utility improve and monitor quality data? What cost-benefit relationships exist between cost of collection data and informational benefits?

## ii) Methods

The project has been divided into five work areas. Each work area has had its own methodology and sub-purposes. The different purposes and methods have been listed in Table 1. In the three first work areas, the question of how VMW should collect data is discussed. In the two last work areas, the question about data utilisation is discussed.

**Table 1: Summary of work packages' purposes and methods**

| WA | Purposes | Methods |
|---|---|---|
| **1** Theory review | • Establishment of AM nomenclature<br>• Review of perspectives on data quality | • Literature study |
| **2** Data collection: technical perspective | • Assess tools in VMW with respect to life cycle monitoring capabilities<br>• Identify measures for improvement | • Analysis of buried asset's life cycle<br>• Analysis of tools used in VMW<br>• Application of literature suggestions (WA1) |
| **3** Data collection: organisational perspective | • Describe current data collection situation in VMW<br>• Identify measures that can improve data collector compliance | • Analysis of current data collection practice<br>• Application of literature suggestions (WA1) |
| **4** Case study Trondheim | • Gathering experiences about data collection<br>• Identifying possibilities and threats | • Report study<br>• Personal enquiry/questionnaire (based on WA1-3) |
| **5** Cost-benefit analysis | • Establish relationship between cost of collecting data and informational benefits | • Development and calibration of cost-benefit model |

## iii) Results and recommendations for VMW

The results are organised according to the work areas.

**WA1:** The theory review consisted of four main themes:

- Perspectives on what infrastructure asset management (IAM) is, and the terminology used
- The role of data in the IAM processes
- The use of models as a transformer of data to information
- Perspectives on data quality

All these perspectives have been used to various extents throughout the thesis. The most comprehensive element of the theory review was the review of data quality perspectives; both technical and organisational perspectives on what data quality is, how data quality may be impeded, and how data quality can be improved, quantified and monitored, were discussed in the review. Perspectives on database design and structure, data quality documentation, the data-collection-utilisation interaction, motivation, training, work process design, perceived behavioural control etc. enabled a holistic discussion of data (quality) throughout WA2 to WA5.

The perspective on data quality that has been used in this thesis is also hoped to provide new ideas on data quality management within water utilities.

**WA2:** In the technical perspective of data collection it was focused on answering the questions of which data that needs to be collected, how well the planned tools in VMW will oblige these needs, and which improvements could be made. The data sources were identified by applying the definitions of IAM from the Canadian National Research Council and analysing the life cycle of a buried asset. The main step towards predictive AM for VMW is the implementation of LCC software (LCC-AM/QM), USTORE and the development of the data integrator project Octopus. Some findings were:

- Even though LCC-AM/QM is a step in the right direction, this software does not address all the dimensions of interest for IAM. LCC-AM/QM only expresses the effect of maintenance and repair as reliability. It is necessary to collect other data in addition to the data required by LCC-AM/QM and USTORE – special concern is raised about utilisation of inspection (leakage), complaints and water quality data. Other tools should also be evaluated in order to use the data more optimal (see WA5).
- VMW should avoid the dispersed data storage structure that is imposed by the proprietary software solutions. A central diary database has been suggested to (1) avoid a dispersed and incompatible storage structure, (2) facilitate storage of additional data, and (3) enable data quality monitoring.
- A technical quality ensuring scheme was suggested for the diary database, including verification and quality tagging metadata.

**WA3**: In the organisational perspective of data collection the question to be answered was: How can VMW, an organisation where data collection previously has not been a high priority, plan the implementation of data collection schemes in such a way that personnel across the organisation will be inclined to comply with the new data requirements?

The initiation phase of data collection initiatives is difficult. The organisation needs data to demonstrate that data generates useful information, but the useful information is dependent on the collection of data. The challenge is to produce relevant informational outcomes from scarce data resources, with relevant enough results to perpetuate the attitudes towards data collection. The most important measures for the initiation phase of data collection were found to be:

- Improving attitudes by including representatives from all affected parties (at strategic, tactical and operational level in the organisation) in the decision-making during the early phase of establishment of new data collection schemes. It is suggested to assign a data quality responsible in each service centre.

- Communicating that data collection is an expected part of the job: Members of the local service centres are suggested to undergo a voluntary training program, and initially collect data collectively (verbal inquiry in meetings). Successful data collection should be rewarded, thus demonstrating that data collection is a sign of personnel professionalism.
- Gradual change of existing paper-based reporting, so that personnel working in the field are better prepared for long-term computer-based data collection.
- Monitoring and communicating data quality by constructing data quality indicators (analogous to performance indicators, PI's). Specific indicators have been suggested.
- Early-stage utilisation of the data at the local service centres, through display of PI's, and scheduling of planned activities (maintenance, inspections, quality sampling, investigation of complaints etc.), thus demonstrating functionality of the collection system and the value of data.

These measures are supposed to substantiate that data collection has a positive effect on a personal, group and organisational level.

**WA4:** A case study of *Trondheim Vann og avløp*, with focus on manual data collection, utilisation and quality was carried out through personal inquiry of personnel in the utility and study of relevant reports from SINTEF. The study resulted in some interesting perspectives, and confirmed several of the problems that were identified in the preceding work areas. Three very important "take-home messages" for VMW remain standing:

- *Trondheim Vann og avløp* has been able to utilise and transform data collection «traditions» from before the digital age – data collection for new IAM processes has been perceived as minor transitions. VMW should strive to start from existing work processes and data collection routines when introducing new data collection systems such as Octopus (see section 4.1.2), and transform these processes over time in order to minimise the perceived change.
- There are three main reasons for data quality problems in *Trondheim Vann og avløp*: (1) a dispersed and disparate storage structure, which impedes the data utilisation capabilities, (2) lacking data definitions that make data inconsistent over time, and (3) lacking capabilities of data quality documentation. These three quality impediments reduce the utility's confidence in conclusions from the analyses based on the data. If measures are not taken to avoid the same problems in VMW, it is likely that similar problems will emerge in VMW also.
- IAM can be carried out with tools that require only a minimum of data, as well as advanced tools that require a vast amount of data. The results are used on different aggregation levels and stages in the planning processes. Tools with low data demands can also be useful: VMW should focus on simple strategic high-aggregation level data utilisation tools the first following years, and gradually implement more advanced data-consuming tools.

**WA5:** In order to be able to answer questions about the value of data under different collection and utilisation alternatives, a cost-benefit model was developed. In this model, costs were expressed as the work hours spent to collect and treat data. The benefits were expressed as information items, strategic, tactical and operational benefits the information items together yield, and the planning strategies the information allow the utility to realise. The benefits were all weighted and summed in one grand measure: The total informational completeness. The model was calibrated for VMW East

Flanders. The methodology that was developed made it possible to assess relationships between costs and benefits for different tool combinations, as well as independently of tool combinations.

The most important results from the analysis of VMW East Flanders were:

- In addition to investing in USTORE, LCC-AM/QM and Octopus, implementing tools for calculating hydraulic criticality, economical risk of failure and expected unmet demands will only require 7.5 % additional increase in the total data costs, but bring the level of total informational completeness from 76 % to 99 %.
- VMW should search for innovative solutions that could increase the utilisation degree of inspection (leakage) data.
- Collection of *extended* data classes should be considered carefully. It is suggested that different service centres collect different extended incident data in a test period, which can be compared in order to make an informed decision about what extended data is most useful.

A paper was made to describe and demonstrate the cost-benefit methodology. This paper is, at the time of writing, submitted to *Journal of Water Resources Planning and Management* and is awaiting review. In the paper the calibration was made with analysis of a fictive water utility, in order to demonstrate the methodology in the model without revealing sensitive details about VMW. The paper is enclosed in Appendix E. A paper based on the contents of chapter 6 is also under development.

## iv) Conclusions and suggestions for further work

This project has endeavoured to establish a broad theoretical framework in which data quality issues for IAM can be evaluated and improved. The set of theories and perspectives that have been reviewed, have been used to evaluate and suggest improvements for VMW. The most important suggestions are the diary database, the list of measures for data collection implementation and the cost-benefit analysis. The case study of Trondheim was also useful to reveal threats and opportunities for VMW.

The suggested methodology for cost-benefit analyses made in this project represents a step in the direction of rationalisation of the selection process of AM tools. The spread sheet model allows the utility to evaluate its current cost-benefit position, identify attractive alternative tool additions, and to identify data classes that yield low benefits compared to their costs, and thus need to be utilised more effectively. The results show that it is both possible and useful to assess the costs and benefits of different data collection scenarios within a systematic framework.

Several new ideas have appeared throughout the project, which still needs to be developed further. Four main innovative points are emphasised:

**Data quality documentation and monitoring:** Data quality documentation is neglected in water utilities (especially for manually collected data), and the idea of quality documentation should be developed and adapted to water utilities by (1) investigating which data quality indicators are the most useful for analysts and for stimulating improved data collection, (2) how raw data quality documentation can aid the assessment of the quality of informational outcomes.

**The organisational factor of data collection:** A comprehensive list of suggested measures has been made for how data collection can be developed within an IAM organisation. These suggestions have not been tested in practice. The ideas should be developed further by (1) identifying which suggestions are more effective and hence should be part of a data quality development plan, and (2) investigating how feedback and communication between data consumers and data collectors can become more effective through e.g. data quality reports and data quality indicators.

**The cost-benefit methodology:** The cost-benefit model should be developed by (1) obtaining measured values for the unit costs of data collection, (2) including a wider array of AM tools, (3) including temporal change effects, and (4) linking informational outcomes to monetary or risk benefits. This could be done in a case study where the reduced economic risk of realised renewal projects is assessed under different decision rationales (which require different data), and the value of data is calculated based on the variation in reduced economic risk between the decision rationales.

**Improving data utilisation:** The cost-benefit analysis showed that inspection data score a low benefit-cost ratio. It is suggested to develop AM tools that can utilise inspection data better by extrapolating inspection results from assets to cohort level, and by forecasting inspection results, much like reliability models do today.

# IX Glossary

**Censoring:** Historical data about infrastructure is often censored, meaning that one does not have a complete record of the events that has occurred throughout the asset life cycles. Left censored data lack data before the temporal observation window, whilst right censored lack data after the observation window.

**Condition:** The term condition will be used relatively loosely in this text. It will be used as a collective description of an asset's state (Vanier, 2001), or "health", and may encompass structural or hydraulic condition, reliability etc. depending on the context in which it is used.

**Data dictionary:** "…a centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format." (IBM, 1993)

**Data integration:** "… the process of combining or linking two or more datasets from different sources" (FHwA, 2001)

**Data aggregation level:** In this text, there is made a distinction between data, depending on what level they are aggregated on. If data is linked to an asset, it is said to be on asset level. If data can only be linked to a zone or a district metering area, it is said to be on district metering area; likewise for cohort level and network level data. The aggregation level will refer to the most specific level a data point can be linked to.

**District metering area:** A sub-section of a water distribution network that is demarcated by closure valves and water flow meters (Morrison, 2004)

**Hydraulic criticality index (HCI):** A measure of the relative importance of a water distributing link. The HCI of a component is defined as the ratio between the demand the system can satisfy without that component in function, and the demand it can satisfy with the component in function (Andralanov, 2012). If a water main has a HCI of 0, then all water demand in the system will be satisfied regardless of whether or not the main is working. If a water main has a HCI of 0.25, then a quarter of the water demand will not be satisfied if the main fails.

**Infrastructure value index (IVI):** The ratio between the current value of an infrastructure asset (or set of assets), and the replacement cost.

**Inspection:** In this text is an inspection defined as a test in which the condition or performance of an asset (or a group of assets) is assessed using a more or less objective assessment method.

**Life cycle cost:** The total cost of ownership, including acquisition, maintenance, inspection, repair and failure cost (Frangopol et al., 1997)

**Life cycle data:** In this thesis the term *life cycle data* will be used as a collective description for all data that can provide a description of performance, risk and cost evolution of the asset through the life cycle of the asset.

**Link:** A network link is a component in a hydraulic system that can convey unidirectional flow (Rossman, 2000). In most hydraulic models are pipes, valves and pumps considered as network links.

**Maintenance:** In this text is maintenance understood as any action that extends the useful service life of an asset, either by increasing its condition or by decreasing its deterioration rate.

**Performance (indicators):** The ability to fulfil a function (see section 2.1.2).

**Risk:** The combination of an event's severity and probability (see section 2.1.2). Economic risk is defined as the expected value of the economic loss.

**Survival function:** A function that represents the probability of an asset still being within its useful service life as a function of its age.

# X Abbreviations

| | |
|---|---|
| ABAO | As bad as old |
| AGAN | As good as new |
| AM | Asset management |
| CAS | Condition assessment system |
| CMMS | Computerised maintenance management system |
| DMA | District meter area |
| DSS | Decision support system |
| DQ | Data quality |
| GIS | Geographic information systems |
| HCI | Hydraulic criticality index |
| IAM | Infrastructure asset management |
| IVI | Infrastructure value index |
| LCC | Life cycle cost |
| LEYP | Linear Extension of The Yule Process |
| LIMS | Laboratory information management system |
| NRCC/NRC | National Research Council Canada |
| NTNU | Norwegian University of Science and Technology |
| PI | Performance indicator |
| PMO | Project Management Office |
| RSL | Remaining service life |
| SRB | Sulphate reducing bacteria |
| VMW | Vlaamse Maatschappij voor Watervoorziening |
| WTO | Worse than old |

# 1 Introduction and project scope

## 1.1 Motivation for the project

Traditionally, it has been economically feasible for water utilities to maintain a desired level of service by repairing or replacing an asset whenever it appears to provide a lower level of service than desirable (D'Água et al., 2007). However, the peak of capital investment intensity has passed for most water utilities, which implies that the infrastructure in the urban water utilities is ageing. At the same time is the expected level of service growing and the economic constraints of the utilities are becoming increasingly stringent, while urbanisation and population growth is increasing (Alegre and Matos, 2009). The modern society is systemically dependent on water infrastructure, therefore is the need for utility owners to consider and minimise risk of service failure growing (Aven, 2010). For these reasons the traditional reactive approach will not be feasible in the future – a predictive management approach will be needed to face the future challenges.

There exists a wide array of tools aiding asset management (AM) strategies. Software and analytical procedures can, through knowledge about economy and engineering, assist more rational decision-making processes within the water utilities. However, these tools have one thing in common: They rely on data about the assets. A lot of the data that is required by the new decision-support tools have traditionally not been collected by the water utilities, because they have not been needed, thus often making data availability an impeding factor for the capabilities of these tools (Halfawy and Figueroa, 2006). Water utilities that are in the transition from reactive to predictive utilities are challenged with the question of how they should ensure that they will be able to provide the data that is needed to make this transition. It is this question that this project is embracing.

## 1.2 Background for the project work

### 1.2.1 VMW's developments towards AM

Like many other water utilities, VMW has been very capital investment intensive in the period from 1960-1980 (Rokstad, 2011a), where maintenance has not been a primary concern. Nowadays VMW invests less in new transport systems; the length of the VMW distribution network had a growth of approximately 1 % in 2010 (VMW, 2010). With an expected service life of 45-75 years on the buried assets, depending on the material (VMW, 2010), it is apparent that VMW's infrastructure is ageing. Hence must VMW be experiencing a paradigm shift from an investing to a maintaining utility (Murphy et al., 2008), where the main concern of the utility will be to maintain its asset portfolio as it ages – from a reactive to a predictive maintenance strategy. VMW is now actuating this shift by implementing tools that facilitate infrastructure asset management (IAM). The portfolio of tools is expected to help VMW to keep more detailed control over the resources spent on managing their assets, and facilitate informed decisions about the allocation of resources needed to manage and maintain their assets (minimising the life cycle cost whilst maintaining the desired level of service). Among the tools that are currently being implemented are software for life cycle costing and maintenance planning, tools for accessing operational data during field work, structures for storing failure data etc.

### 1.2.2 Results and conclusions from previous project work (TVM4510)

This thesis is a continuation of the work described in the project report *Statistical failure analysis of the VMW water distribution network in East Flanders* (Rokstad, 2011a), which is part of the NTNU course *TVM4510 Water and wastewater engineering, specialization project*. In that project, failure data from VMW's water distribution system in the province of East Flanders (Oost-Vlaanderen) was analysed in order to reveal whether or not readily available explanatory factors in the VMW databases could explain the variability in the failure data to a degree that would justify the implementation of advanced failure models. The analysis was carried out with analysis of variance, regression analysis and the tailor-made model LEYP (Cemagref, 2010, Cemagref.fr, 2011, Le Gat, 2009). The conclusion from the project was that although it is possible to explain a significant amount of the observed variability in the failure incidents through explanatory factors from the central GIS database, the data quality was so low that VMW was advised to invest resources in improving collection of failure and maintenance data before implementing advanced models for prediction of deterioration and failures. (Rokstad, 2011a)

Now that VMW is planning to implement different tools for AM, one should seize the moment and plan the data collection that is needed to utilise the planned tools as best as possible.

## 1.3 Research goals and desired results

The overall research goal for this thesis project is to evaluate how VMW should collect data to monitor the life cycle of their buried assets in order to ensure that the organisation will have the necessary information to make informed decisions about their assets in the future.

Results from the project should include:

I. A theory review about data and data quality for AM, and an evaluation of how data collection could help describe the life cycle of a buried asset
II. An evaluation of the technical and organisational perspectives of data collection and utilisation in VMW, with suggestions for how the data collection can be improved.
III. An evaluation of expected performance of the system, through a cost-benefit analysis

The studies embedded in this thesis will be concerned around the buried water distribution assets of VMW. Water production units, water tanks, and pumping stations will not be discussed explicitly (energy costs for pumps etc. will neither be included). Even though VMW owns a small portion of waste water networks, this will neither be discussed explicitly within this project.

## 1.4 Research plan

The research plan for the project is outlined in Table 2, where the work areas (WA) represent main deliverables, and the work packages (WP) represent main research elements in the deliverables.

**WA1** encompasses a literature study on AM, with focus on data and data quality for IAM. Following the literature review, the life cycle of a buried asset is analysed in **WA2**, with respect to what has been uncovered in the theory review. The analysis of data through the asset life cycle, combined with an analysis of the VMW initiatives for AM and data acquisition systems, will result in an evaluation of VMW's technical capability to feed their AM tools with data, and if this is sufficient to "monitor the life cycle" of a buried asset. **WA3** is similar to WA2, but focuses on the organisational perspective of data acquisition. **WA4** is a study of *Trondheim Vann og avløp* (water and sewerage department), where their implementation of AM tools is studied and discussed, with focus on (lack of) data collection, data quality and data utilisation. The case study should yield some pointers about opportunities and threats for VMW. In **WA5** the expected performance of the AM tools in VMW are assessed through a cost-benefit analysis.

The cost benefit-analysis (WA5) is calibrated with VMW's assets in East Flanders, since this has been the case study area in previous works. In the Trondheim case study, Trondheim will be compared to the East Flanders province as well.

**Table 2: Research plan**

| WA | WP | Description | Section |
|---|---|---|---|
| | *1.1* | *Information collection about VMW's initiatives (vision and strategy)* | *1* |
| | *1.2* | *Literature study (AM, data quality)* | *2* |
| 1 | | Introductory literature study and information collection about VMW AM plans | 1-2 |
| | *2.1* | *Analysis of buried asset life cycle - identification of data requirements* | *3* |
| | *2.2* | *Information collection and analysis of AM tools in VMW* | *4* |
| 2 | | Technical evaluation of VMW data collection with suggestions for improvement | 5 |
| | *3.1* | *Collection of information about data collection procedures and attitudes in VMW* | *6.1-6.2* |
| | *3.2* | *Analysis of organisational context for data collection in VMW* | |
| 3 | | Organisational evaluation of VMW data collection - suggestions for improvement | 6 |
| | *4.1* | *Study of Trondheim municipality - steps towards SAM (data perspective)* | *7.1-7.2* |
| | *4.2* | *Study of Trondheim municipality – lessons learned in Trondheim* | *7.3* |
| 4 | | Case study: Trondheim – what can VMW learn from Trondheim? | 7 |
| | *5.1* | *Development of cost-benefit spread sheet model (calibration in East Flanders)* | *8* |
| | *5.2* | *Paper on cost-benefit* | *Appendix E* |
| 5 | | Cost-benefit of data collection: Selection of tools/models for future use | 8 |

# 2 Theory review: Quality data for asset management

The purpose of this chapter is to provide the necessary theoretical background on:

- Infrastructure asset management (IAM)
- The role of data in IAM processes
- The use of models in IAM
- The challenges regarding data quality in IAM

This theoretical background is the basis for all further evaluations and recommendations for improvements that are made later in this project.

## 2.1 Infrastructure asset management

### 2.1.1 What is asset management? Overall definitions

Many definitions of the term *asset management* (AM) exist. USEPA (2008) defines AM:

*"Asset management is maintaining a desired level of service for what you want your assets to provide at the lowest life cycle cost."*

This definition states thus that AM is a governance process that balance two conflicting objectives – maximal level of service, and minimal life cycle cost. *An asset* is understood as any property of a person or organisation that has a value and is revenue-generating (Amadi-Echendu et al., 2010). Ugarelli et al. (2010) define AM:

*"…asset management can be recognised as a set of management, financial, economic, engineering activities, systematic and coordinated, to optimally manage the physical assets and their associated performance, risks and expenditures over their life cycle with the objective of ensuring level of service in the most cost-effective manner."*

This definition goes further than the definition from USEPA, since it also includes *risk* as a factor for optimization, and also defines a set of processes necessary to implement AM. Cromwell and Speranza (2007) state that AM inherently *is* risk management.

It is often referred to three levels of planning in AM (Alegre and Covas, 2010):

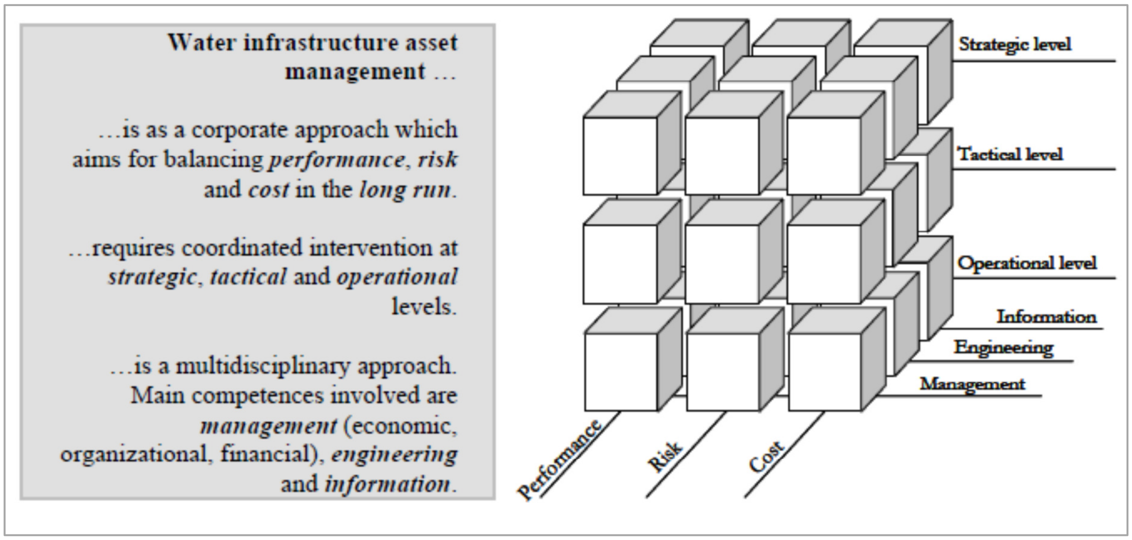
1. At the **strategic level** is the general direction of the planning set through defining the goals and objectives of the utility. The desired level of service, overall performance and financial needs of the utility are determined at the strategic level; how the objectives should be benchmarked should also be decided on the strategic level. Strategic planning occurs at a network level, and typically has a long time horizon (10-20 years).
2. At the **tactical level** are different alternatives to achieve the strategic goals evaluated and compared. In a tactical planning situation are individual assets evaluated with respect to the objectives defined at the strategic level. A tactical plan yields a set of prioritised assets that need

to be renewed or rehabilitated (projects). The time horizon for tactical planning is typically 3-5 years.

3. At the **operational level** are different technologies to realise the projects, which have been selected on the tactical level, evaluated and selected. The operational plan is a short-term plan (1-2 years) and contains the description of how the projects should be implemented.

An organisation that has a *strategic AM plan*, plans on all these three levels (Ugarelli, 2008).

Alegre et al. (2006a) acknowledges that AM is a multidisciplinary pursuit, requiring competences about *engineering, management and information*. If the different AM disciplines are included, one may illustrate AM along three axes, as in Figure 1 – performance risk and cost are balanced on the strategic, tactical and operational level, by the aid of information, engineering and management disciplines.



**Water infrastructure asset management** …

…is as a corporate approach which aims for balancing ***performance***, ***risk*** and ***cost*** in the ***long run***.

…requires coordinated intervention at ***strategic***, ***tactical*** and ***operational*** levels.

…is a multidisciplinary approach. Main competences involved are ***management*** (economic, organizational, financial), ***engineering*** and ***information***.

Figure 1: Three dimensions of IAM (copied from Alegre et al. (2006a))

As presented in Figure 1, Alegre et al. (2006a) provides a quite comprehensive definition of IAM. The reader may especially note the incorporation of the phrase "… in the long run", putting emphasis on the fact that AM is long term planning, and indeed an asset life cycle consideration.

In this project it is regarded as suitable to utilise a more operational definition of IAM. The National Research Council in Canada (Vanier, 2001) regards IAM as the "successful implementation of data collection" related to six questions about the assets:

- What do you own?
- What is it worth?
- What is the deferred maintenance?
- What is the condition?
- What is the remaining service life?
- What do you fix first?

Vanier (2001) and the "six whats of AM" will be used as a basis to evaluate how a life cycle data collection system can help to answer these questions. A further explanation of what is meant by the six questions is provided in section 2.2.1. Implementation of municipal infrastructure management is described in Vanier et al. (2009) and encompasses eight processes and six facets that need to be evaluated in each process. The processes and facets are listed in Table 3.

**Table 3: Eight processes and six facets for implementation of IAM, according to Vanier et al. (2009).**

| Processes | Facets |
|---|---|
| 1. Select protocols | 1. Inventory |
| 2. Itemise assets | 2. Performance |
| 3. Inspect assets | 3. Service life |
| 4. Rate assets | 4. Life cycle cost |
| 5. Forecast needs | 5. Criticality |
| 6. Integrate needs | 6. Alternatives |
| 7. Recommend resources | |
| 8. Optimise investment | |

From the definitions above, one can state that IAM is a way of managing information and engineering knowledge, on several levels, in order to balance the resource use with the desired level of service and acceptable level of risk. In the following subsections, a more in-depth review of the concepts mentioned here will be presented.

### 2.1.2 Why is IAM important?

The need for strategic IAM of water utilities has been pointed out in the literature (Alegre, 2009, FHwA, 2001, Halfawy, 2008, Lemer, 1998, Vanier and Rahman, 2006). As previously mentioned, the capital investment era for water utilities is over: as the infrastructure ages and deteriorates, the risk of failing to provide the required level of service increases, and the need for a predictive maintenance strategy emerges. Increased requirements to the level of service, cost efficiency, economical control, population growth and accountability for the risk imposed from the water utilities are also drivers for implementation of predictive IAM strategies in water utilities (Alegre, 2009, Vanier and Rahman, 2006).

With the changing environment in which water utilities operate, where requirements are becoming increasingly stringent, it is becoming increasingly important that water utilities are managed as effective as possible through the aid of AM. Prolonging asset life (through efficient and correct decisions about rehabilitation, repair and replacement), sustainability, sound financial planning, improved emergency response and asset security are benefits of AM planning identified by USEPA (2008). Alegre (2009) also identifies ability to plan for climate change, possibility for risk management, promotion of investment and operational efficiency, and clear justification and transparency of investment priorities as drivers for and benefits of IAM.

### 2.1.3 Performance, risk and cost

In the definition of AM from Alegre et al. (2006a), the concepts of performance, risk and cost are present. These notions are central in AM, and it is therefore important to define them properly:

**Performance** is the capability to perform required functions (Vanier, 2006). An asset that performs well fulfils its functions and complies with the desired level of service. Performance can be monitored by *performance indicators* (PI’s) (Alegre et al., 2006b). Alegre et al. (2006b) defines water resources, personnel, operational, physical, quality of service, economic and financial PI’s for water utilities. PI’s can be used to benchmark and compare performances between different utilities are utility sub-units.

**Risk** is “an uncertain consequence of an event or activity with respect to something humans value” (IRGC, 2008). Aven (2010) defines risk as *“a combination of an event’s severity and probability”*, which is a commonly applied definition when assessing risk in engineering contexts. Risk severity can incorporate many dimensions, such as human life or health, environmental indicators or monetary value (Aven, 2010). (*Severity* is equivalent with *criticality*.)

**Cost** in the AM context usually refers to the life cycle costs (LCC) of the assets, which incorporate all costs of asset ownership, including the direct and indirect cost of acquiring, insuring, inspecting, maintaining, repairing and decommissioning the asset, in addition to the cost of asset failure (Frangopol et al., 1997).

### 2.1.4 Different strategies to IAM

IAM can be performed at different levels, with different time horizons, but also with different strategies. Ugarelli (2008) classifies four maintenance strategies for AM in water utilities, summarised in Table 4.

Table 4: Four asset management strategies (summarised from Ugarelli (2008))

| Strategy | Description |
|---|---|
| Operative – reactive | Repair only after failure. Ad hoc-decisions; may be based on practical experience. Difficult to plan due to uncertainty about coming failure incidents. |
| Inspection – condition based | Performance and rehabilitation prioritisation is determined through (periodic) inspections. No evaluation of consequence of failure. |
| Proactive – preventive | Rehabilitation prior to failure. Usually involves utilisation of decision support systems (DSS), wherein hydraulic, environmental and structural conditions of the water system are considered. |
| Predictive - advanced | Optimising performance and reliability at lowest possible cost, using a life cycle perspective. A predictive approach involves an analysis of whether or not an asset’s service life can be extended through operation or maintenance, or if rehabilitation is required. |

A predictive maintenance approach is much more advanced than an operative (reactive) approach, and there is a progressive advancement down through Table 4. However, Ugarelli (2008) emphasises that one of these approaches is not better than the rest, but that “each approach has a specific role within the AM methodology”. To follow only a reactive or only a preventive approach will be equally exhausting on the utility’s resources. Here, as with other aspects of AM, is it a question of *balance.*

Incidentally, Schneider et al. (2006) presents a somewhat similar classification of maintenance strategies, as displayed in Figure 2. Schneider et al. (2006) organise the maintenance strategies based on whether or not condition and/or importance are considered in the different strategies. The reliability centred maintenance strategy is similar to what Ugarelli (2008) classify as “proactive”. The classification in Figure 2 includes a “time based maintenance” strategy, wherein inspections of assets follow a fixed time schedule, depending on asset characteristics and importance[1] (Schneider et al.,

---

[1] Consequence of failure (criticality/severity)

2006). As one may see, the classification in Figure 2 does not contain an equivalent to the predictive strategy from Table 4; this is because the predictive approach also involves an economic dimension, in which the expected change of service life also is incorporated.

| Condition | Importance: Not considered | Importance: Considered |
|---|---|---|
| Considered | **CBM**<br>**Condition based maintenance**<br>➔ Continuous or occasional monitoring<br>➔ Maintenance when required | **RCM**<br>**Reliability centred maintenance**<br>➔ Priority list<br>➔ Connection of condition and failure effect<br>➔ Risk management |
| Not considered | **CM**<br>**Corrective maintenance**<br>➔ No inspection or maintenance until breakdown | **TBM**<br>**Time based maintenance**<br>➔ Fixed time intervals for inspections and maintenance |

Figure 2: Classification of maintenance strategies according to Schneider et al. (2006)

The purpose of the review of these different maintenance strategies is to enlighten the reader to the fact that AM simultaneously occurs with different strategies within a utility, and that optimisation is about balancing different strategies. Assessing to what extent each of the strategies are utilised in a utility, may be one way of determining the utility's level of AM implementation. Following only a reactive strategy (repairing failures as they come) is not sustainable (Vanier et al., 2006), but a long term strategy for maintaining expected reliability and maximising return value from one's assets on one hand, and effectively repairing (seemingly) unexpected failures on the other, may very well be a sustainable strategy.

## 2.2 What role does data partake in AM?

In section 2.1 different definitions of AM were presented, and it was acknowledged that AM is a data intensive activity. In order to explain what this actually means, the "six whats" of AM will be used.

### 2.2.1 AM as a set of data-intensive processes

When applying the operational definition of AM from The National Research Council in Canada (Vanier, 2001), one has already acknowledged that AM is a set of several processes that require data. These six questions will be reviewed below in order to explain what they mean, and perhaps also why Vanier (2006) characterises them as "six simple questions with six difficult answers":

1. Water utilities usually keep track of **what they own** through an asset registry (an inventory), for instance a GIS database. An asset database can contain spatial data (the positions of the assets), material, producer, dimensions, time of installation etc. (Lemer, 1998, USEPA, 2008, Vanier, 2001).
2. Asking **how much an asset is worth** is a somewhat ambiguous question. Many water utilities use the historical cost of acquiring the asset (appreciated from year of installation to current value); nevertheless, the *value* also be expressed as the cost of replacing it (capital replacement value), or the cost of the loss experienced if the asset was not to fulfil its

function (deprival cost[2]) (Lemer, 1998). Vanier (2006) suggests that a comprehensive asset management system should have at least two different valuation methods of the assets. If the appreciated historical cost is used, the data required from the asset is the time of installation, the installation cost and the depreciation period (service life). The other valuation methods require more data.

3. The **deferred maintenance** is defined as the "cost of bringing the asset back to its original potential" (Vanier, 2001). In order to assess the deferred maintenance, one must both keep track of the cost of maintenance that has been executed *and* the effect the deferred maintenance has on the value of the infrastructure. Kong and Frangopol (2004) suggest a method for calculating the interaction between cost of maintenance and reliability of an infrastructure asset, which could also be applicable for calculating deferred maintenance. Deferred maintenance can be evaluated with respect to several dimensions, such as infrastructure value index and functionality (performance). Vanier (2006) states that a utility should assess deferred maintenance along at least two dimensions.
4. Assessment of an asset's **condition** can be viewed as an extension of the assessement of deferred maintenance; where the deferred maintenance is expressed through the deterioration of the condition (instead of through cost). Vanier (2001) and USEPA (2008) suggest that the condition of assets should be detected and recored through a condition assessment system (CAS). Non-destructive ways of assessing the condition of water distribution infrastructure exist (Liu et al., 2012); most widely used are leakage detection methods. *Reliability* is often used as a measure of water distribution system's condition (Burn et al., 2010), since the *condition* term is used in a very broad sense in this context (Vanier, 2006), both physical and functional (Bhagwan, 2009). The current and future reliability of water distribution assets can be extrapolated by using failure records and statistical models (Constantine, 1993, Davis et al., 2007, Goulter and Kazemi, 1988, Gustafson, 1999, Kleiner and Rajani, 2001, Le Gat, 2009, Rogers and Grigg, 2006). It is important to note that the development of new inspection techniques for water distribution networks might increase the importance of the structural condition concept in water distribution AM. In order to assess the condition of infrastructure assets, one needs a record of failures (reliability) and/or a record of inspections.
5. Similarly to the asset value, the **remaining service life** is an ambiguous term: in the literature it is referred to the *technical* and the *economic* service life of an asset (Lemer, 1998), both requiring different data to assess. *Remaining technical service life* means the remaining period the asset can fulfil its function (Vanier, 2001), and can be assessed through statistical or mechanical modelling (Burn et al., 2010, Kleiner and Rajani, 2001, Rajani and Kleiner, 2001), or estimated through life expectancy tables. Remaining economic service life means the remaining period in which asset maintenance is less costly than capital renewal (Vanier, 2001), and can be estimated through LCC considerations, where data about repair, maintenance and inspections are balanced with cost of capital renewal (Vanier, 2001). It can be argued that the technical service life of a buried asset is of little relevance: Le Gat (2009) states that the service life of a pipe ends when the owner decides to decommission it. If this is true for all buried assets, assessing the service life of an asset is still important, because it allows the asset owner to assess which relationship there is between the decommissioning

---

[2] It is usually not suitable to assign *deprival costs* to water distribution assets, due to the interdependencies between the assets

decision and the actual service life of the assets. Assessing the economic service life of an asset requires data about all costs related to the asset.

6. Vanier (2001) argues that the better answers (information) one has obtained for the five previous questions, the better one is able to answer the final question: **What do you fix first?** If the asset manager knows what the utility owns, decisions will be made based on the complete array of assets, but still on subjective evaluations of the individual asset's need for renewal; if the value and condition is known, the decisions will be made on these rationales, and so forth. In order to manage assets optimally, with respect to maximal utilisation of capital (Lemer, 1998), one needs to be able to answer all the "six whats" of AM especially crucial for economical optimisation, is the consideration of remaining economic service life, which accumulates data from all the previous questions (Burn et al., 2010).

### 2.2.2 Data – the most challenging issue of IAM?

Reviewing the "six whats" of AM from Vanier (2001), it is clear that the level of AM implementation one is able to achieve is dependent on what data that is available. A low level of implementation only requires an asset registry, accompanied by some method of asset valuation that does not necessarily account for the asset condition (Lemer, 1998, Vanier, 2001, Vanier and Rahman, 2006). A high level of AM, on the other hand, requires an array of different data, keeping track of the physical condition, the value and the cost associated with the asset. The models that exist for answering the six questions are also very much dependent on data over a certain observation period, and very often are the quality of the model results impeded by the quality and amount of the input data. Vanier (2001) argues that "efficient information management is the key to better decision-making for municipal infrastructure". Several other authors identify lack of data (or sufficient data quality) as one of the main challenges for implementation of high-level AM (Halfawy and Figueroa, 2006, Halfawy et al., 2006b, Lemer, 1998, Lin et al., 2006, Lin et al., 2007) and identifies data as the foundation for operational, tactical and strategic planning and resource management – complex models for (life cycle cost, condition, reliability, remaining service life) cannot be successfully implemented without adequate data (Wood and Lence, 2006). This statement from Halfawy and Figueroa (2006) summarises AM's dependence on life cycle data:

*"Successful implementation of asset management strategies largely depends on: (1) the efficiency to share, access, and manage the asset life-cycle data; and (2) the ability to efficiently support and coordinate the multi-disciplinary work processes at the operational and strategic levels." (Halfawy and Figueroa, 2006)*

In the following subsection will the role of models as data utilizers be discussed.

## 2.3 The use of models in IAM

### 2.3.1 In general

The use of models to aid the decision support processes that are inherent in IAM has become increasingly more important the past decades. Models for assessing and predicting reliability of buried assets, based on statistical and/or physical principles, have been utilised as decision support for deciding project prioritisation ("what do you fix first?") (Rajani and Kleiner, 2001, Kleiner and Rajani, 2001). Models can be used to assess performance, condition, criticality, failure rates, risk, life cycle costs and other decision-influencing factors, and are hence important tools on the tactical AM level.

Models for IAM can be understood as tools that transform data into useful information. The models vary greatly in what underlying principles they are built on, what data they require, and in which form the results emerge. One way of distinguishing model principles is described here:

- On one hand, a utility will usually rely on a hydraulic model portfolio, which is used for assessing the importance of components, hydraulic performance, the effect of changes in the hydraulic conditions etc. A hydraulic model is based on simplified physical laws and basic characteristics of the components in the water distribution network.
- On the other hand, the utility may utilise models that assess the deterioration of the assets, in the form of condition, reliability etc. – such models utilise asset characteristics data and historical data (repairs, maintenance, inspections, decommissioning) to forecast the "health" of the components, either by using mechanical or statistical principles.
- Both models for assessing importance or performance, and deterioration models may be the basis for decision-making, which can be aided by decision-support systems. Decision-support models may be based on criticality, reliability, hydraulic or economical risk, life cycle costing etc. Decision-support may be based on the aforementioned model classes, as well as data about costs and the different alternatives that are relevant to choose between. (Hadzilacos et al., 2000)

There are many different model principles within these three classifications, which will not be discussed in this thesis. However, for a review of the general principles and data need for hydraulic models it is referred to Rossman (2000); for a review of principles and data needs for failure and deterioration models it is referred to Kleiner and Rajani (2001), Rajani and Kleiner (2001), Røstum (2000) and Kong and Frangopol (2003); and an example of an integrated decision support system it is referred to Hadzilacos et al. (2000).

### 2.3.2 The Aware-P toolbox

As an example of the range of different modelling tools that can be used for AM, the *Aware-P infrastructure asset management software* will be used (from here on denoted *Aware-P toolbox*) (baseform.org, 2012a). This toolbox contains both a hydraulic component, for assessing component importance and performance, a deterioration component for assessing reliability and service life, and a decision-support system for assessing different investment scenarios. The Aware-P toolbox is thus a broad example of how models can help to provide information necessary for rational and informed planning.

The Aware-P tool integrates the software tools developed in the Aware-P project (aware-p.org, 2011). The software platform allows the user to combine sources of data, channel these into to portfolio of tool modules in order to evaluate and forecast performance, risk and cost aspects under given management scenarios (Coelho and Vitorino, 2011). It is available as a web-based application on **www.baseform.org**. The tools in Aware-P (modules) are described in Table 5.

Table 5: Tools in the Aware-P platform

| |
|---|
| **Baseform Core**<br>All the tools in Aware-P rely on the *Baseform Core* platform, a common data manager and user interface platform. The Core handles all data import, and manages all data commonly, which allows for a more seamless interaction between the tools. (baseform.org, 2012b) |
| **PLAN**<br>The *PLAN* tool is created to be able to compare different renewal alternatives. The tool consists of three axes; the alternatives axis (for different alternatives), the metrics axis (where different alternatives are compared) and the time axis. In other words, the set of chosen performance metrics can be assessed for each alternative strategy and each future time step, allowing a transparent and informed decision process, by comparing alternatives side by side. (baseform.org, 2012g) |
| **EPANET**<br>The *network simulation* tool is based on the EPANET hydraulic model (Rossman, 2000). The tool includes all the functionalities of the standard EPANET model, like dynamic pressure and flow considerations. However, the Aware-P network modeller has expended visualisation capabilities, with the possibility of displaying results in 2D/3D maps as layers on top of third-party maps (such as Google Maps). Modelled parameters from other modules, such as failure rates, can be displayed in the same maps. The tool also allows the export of network data and model results as spread sheet files. (baseform.org, 2012d). |
| **PI**<br>The *PI* tool allows the user to define PI's or select PI's from comprehensive lists. *PI* then produces the spread sheets where necessary data must be filled in – what data is necessary, depends on which PI's have been selected. Data can be filled in through the web-browser, or spread sheet files can be downloaded and filled in through spread sheet applications. (baseform.org, 2012f) |
| **PX**<br>The *Performance Indices* tool (*PX*) is similar to the *PI* tool, but the *PX* calculates technical performance metrics based on aggregated and/or modelled results (hence can performance indices be modelled for future scenarios). Examples of performance indices are for instance minimum pressures; PX calculates the minimum pressures by the help of the network modeller, compares these to the user's reference values, and returns a graded result to the user in the form of an aggregated result (globally, and for each time step), and a map of the network showing the variations in the results between each asset. The results are graded as good (2-3), fair (1-2), or bad (0-1). (baseform.org, 2012h) |
| **FAIL**<br>*FAIL* calculates current or projected failure rates, based on the water mains' characteristics and failure records, either utilising the LEYP model or a Poisson process (depending on the user's choice). In LEYP failures are predicted based a linear extension of the Yule process, with number of previous failures, elapsed time and a set of explanatory covariates as input (Cemagref, 2010, Cemagref.fr, 2011, Le Gat, 2009, Rokstad, 2011a). *FAIL* produces a table of average failure rate estimates for each material and each pipe (which can be displayed in a map) for a given year. The input data for each network main are ID, material, length, installation date, and decommissioning date. For each failure, the data required are failure date, type, duration, and link ID. |
| **LLIFE**<br>*LLIFE* is a failure analysis tool, based on the LEYP model. *LLIFE* will go one step further than *FAIL*, by introducing costs of failures to the failure forecasting; if the user provides a certain economic limit, *LLIFE* should then be able to estimate the economic service life of network components (Vitorino, 2012). This tool is still under development, and not yet available (baseform.org, 2012c). |
| **CIMP**<br>The *Component Importance* tool (CIMP) uses EPANET to calculate the importance (hydraulic criticality) of each link (pipe, valve or pump) in a water distribution system. This tool requires the input from EPANET and a minimum pressure limit. |

| |
|---|
| **UNMET**<br>*UNMET* calculates the expected unmet demand from the system, based on the failure rates obtained from FAIL and CIMP. The unmet demand is calculated as the product of the component importance, failure frequency and the average outage time (time between failure and completed repair). The outage time must be provided in the failure logs; otherwise it has to be estimated. (baseform.org, 2012i) |
| **IVI**<br>The *Infrastructure Value Index* (IVI) calculates the ratio between the current value of the assets, and the replacement costs, based on material, length, time of installation, expected useful lifespan, construction and replacement costs. When *LLIFE* is finished, it should be possible to use the service life estimates from *LLIFE* as input to *IVI* (baseform.org, 2012e) |

All the tools in Aware-P can be utilised under different maintenance- and renewal scenarios. The tools can be used to assess how different strategies or scenarios will affect the IVI, the criticality or the expected unmet demand in the future.

So far, the review of data need for IAM has addressed the different types and sources of data. However, in order to successfully implement a high level of IAM, the data must also be of sufficient quality (Lin et al., 2006, Lin et al., 2007). Therefore, a discussion about data quality for AM is conducted in the rest of the theory review.

## 2.4 Data quality – definitions, problems and solutions

In section 2.2, it has been established that IAM is indeed a data-intensive process, and that operational and economic data are necessary to objectively evaluate an infrastructure asset. Further, in section 2.3 the role models and their data needs were presented. However, the issue of the quality of the data required for AM must also be addressed (Halfawy and Figueroa, 2006, Lin et al., 2006, Lin et al., 2007, Wang and Strong, 1996). The purpose of this section is to give a theoretical overview of what is meant by data quality, review data quality impediments, and review suggestions from the literature on how high-quality data may be obtained.

### 2.4.1 What is quality data?

Several definitions of data quality exist. Lin et al. (2006) define quality data as data that is accurate, complete, timely (updated) and consistent. Other parameters may be also be used to define data quality, such as *precision*, *reliability*, *accessibility* and *interpretability* (Wang and Strong, 1996). According to Wang and Strong (1996) data quality adequacy is dependent on the use of the data, and therefore define quality data as “data that are fit for use by data consumers”. A comprehensive list of data quality parameters may be found in Wang et al. (1993). A widely applicable definition of high data quality may be a composite of the aforementioned definitions: Data is of high quality if it is readily accessible and comply with the necessary quality parameters required by the data user.

According to Weidema and Wesnæs (1996) the quality of data can be expressed through information data (metadata) that describe measures of the quality parameters.

Low data quality is disadvantageous for organisations. In addition to inhibition of the informational outcomes, Redman (1998) reports an array of impacts of low data quality on the operational, tactical and strategic level of an organisation – including low job satisfaction, increased operational costs, compromised decision making, aggravated customer relations etc. According to Vigon and Jensen

(1995) lack of control of data quality predispose data consumers (analysts) to be more cautious about the conclusions that can be drawn from analysis of the data.

### 2.4.2 What impedes data quality?

Lin et al. (2006) state that maintaining data quality in an organisation often is acknowledged as problematic, and identify the following issues as common data quality-impeding factors:

- Inadequate management structures ensuring data quality
- Inadequate rules, training and procedural guidelines
- Fragmentation and inconsistencies among the services associated with data collection
- Requirement for new management methods which utilise high quality data to support the dynamic management environment

These four quality-impeding factors apply for engineering AM in general, but to which extent do they apply to water utilities? Inadequate management structures ensuring data quality may very well be a problem in water utilities that are facing a transition towards predictive AM strategies; if a water utility has followed a reactive strategy, the need for, and strategic value of data collection, is unlikely to have been acknowledged, thus is management emphasis on data quality (or data collection in general) unlikely to be adequate.

Further, inadequate rules, training and procedural guidelines may also be a factor present in water utilities. As an example, Müller and Fischer (2007) show studies of interrater inconsistencies in sewer condition class registrations (different personnel rate CCTV observations differently following the same protocol for registration) - without proper guidelines, accompanied by training and emphasis on data collector skills (Murphy, 2009) data quality in water utilities is unlikely to reach a satisfactory level.

Fragmentation and inconsistencies among the services associated with data collection is also a relevant issue for water utilities. *Firstly*, difference in knowledge, skills and perceptions of different operators has already been discussed, but it may be worth mentioning that outsourcing of construction, repair and maintenance work may also be a contributor to inconsistencies in the data (interpretations, semantics and formats). *Secondly*, the involvement of different disciplines may involve different interpretations of the meaning of data – e.g. a data collector may have different perceptions than a data analyst.

The last issue Lin et al. (2006) identifies as a possible data quality impediment, is the requirement for new management methods which utilise high quality data in dynamic management environments. This factor may also be highly relevant for water utilities. The transition to integrated IAM strategies requires data of open standards, to be readily used for different purposes, software and models (Halfawy, 2008, Halfawy and Figueroa, 2006, Halfawy et al., 2006a, Halfawy et al., 2006b). If a utility is experienced in collecting data for a certain asset registry or management tool, the transition to data repositories where data is in a format and of a quality that may serve several purposes, may prove challenging.

One issue that is not explicitly mentioned by Lin et al. (2006), is the fact that much of the data that is related to inspection, maintenance and repair of buried assets must be collected manually. Although both training and management emphasis on data collection have been mentioned, there is also a need to consider the environment in which the data collectors work.

From this discussion, it is clear that water utilities may meet several challenges on the path towards high-quality data repositories. In the following section perspectives on how these challenges can be met are outlined.

### 2.4.3 How does one achieve high data quality?

This subsection aims to review literature that defines and describes methods for maintaining data quality. There are several theories addressing different aspects of data quality within an organisation; some theories focus on the technological aspects of data quality, whereas some also take the organisational and personnel context into account. This section will review some of the theories and approaches that have been developed on data quality, and make a summary at the end, commenting and extracting elements from all the theories, which will be used further in the thesis.

#### *2.4.3.1 The data quality framework*

Wang et al. (1995b) present a framework for data quality research, in which different functions (elements) in an organisation are identified and assigned responsibilities and work descriptions. This framework can be used to generically analyse the state of data quality management within an organisation (Lin et al., 2006, Lin et al., 2007, Wang et al., 1995b) by evaluating the points described in Table 6. The different points in the table may be used as a "checklist for data quality". (The term *data product* appear in the table; Wang et al. (1995b) use *data product* as an analogy to product manufacturing, where a raw material (data) is *processed* to form a (data) product.)

**Table 6: A data quality research framework (copied from Lin et al. (2007))**

| Element | Description |
|---|---|
| Management responsibilities | • Development of a corporate DQ policy<br>• Establishment of a DQ system |
| Operation and assurance costs | • Operating costs include prevention, appraisal and failure costs<br>• Assurance costs relate to the demonstration and proof of quality as required by customers and management |
| Research and development | • Definition of the dimensions of DQ and measurements of their values<br>• Analysis and design of the quality aspects of data products<br>• Design of data manufacturing systems that incorporate DQ aspects |
| Production | • Quality requirements in the procurement of raw data, components and assemblies needed for the production of data products<br>• Quality verification of raw data, work-in-progress, and final data products<br>• Identification of non-conforming data items and specifications of corrective actions |
| Distribution | • Storage, identification, packaging, installation, delivery, and after-sales servicing of data products<br>• Quality documentation and records for data products |
| Personnel management | • Employee awareness of issues related to DQ<br>• Motivation of employees to produce high-quality data products<br>• Measurement of employee's DQ achievement |
| Legal function | • Data product safety and liability |

As stated in the framework, the organisation management has the responsibility to develop a corporate *data quality policy* and a *data quality system*. Further, Wang et al. (1995b) argue that cost of operating, maintaining and assuring data quality should be monitored in order to evaluate it with respect to the cost (consequences) of sub-standard data. In the lower elements of the framework, definition of quality dimensions and measurements of them, quality verification, corrective action routines, quality documentation, personnel, and liability issues are addressed.

What is perhaps the most important merit of this framework is that it emphasises that there is a need for an organisation-wide data quality policy and a data quality system, which is applicable for all data quality related issues further down in the organisation. If data quality requirements are not explicitly defined at a high level in the organisation, it is difficult to assess and improve the quality of data. Further, without defined requirements and dimensions for data quality, one may perceive the data quality problems erroneously – for instance is accuracy most commonly perceived as the main data quality problem, when it in fact is erroneous data that is most often the real problem (Lin et al., 2007).

This framework may work as an “umbrella theory”, covering the overall aspects of data quality, when one is investigating or developing data quality within an organisation.

#### *2.4.3.2 Data quality in medical records*

Another, less generic data quality framework is a result of a study of data quality in medical registries (Arts et al., 2002). The data collection conditions in medical registries, as described by Arts et al. (2002), resemble the conditions in water utilities in several aspects. Firstly are data collected at disperesed localities, and used from a central common repository. Secondly are data used for different purposes depending on who use it. Lastly have the data collectors have different motivations and knowledge than the data consumers. The medical records framework identifies data quality impeding factors under these data collection conditions, and suggests methods for quality assurance and improvement.

Arts et al. (2002) identified unclear definitions, non-compliance, lack of collector support, insufficient control over correction procedures etc. as the most important quality impediments. Table 7 shows the procedures that were suggested to limit these impediments. Suggestions are made both for the onset phase, the data collection phase, and for the improvement process after data has been collected. Suggestions are made both for the local sites where data is collected and for the central treatment of data.

This framework proves to be interesting for data collection in water utilities, due to the similarities in the data collection situation of medical data, and due to the fact that it has suggested concrete solutions to the problems and challenges that have been identified in water utilities.

**Table 7: Procedures for data quality assurance and improvement (copied from Arts et al. (2002) (slightly modified))**

| Central coordinating centre | Local sites |
|---|---|
| **Proactive work during set up of registry** | |
| **At the onset of the registry:**<br>• Compose minimum set of necessary data items<br>• Make data dictionary<br>• Make a collection protocol<br>• Identify pitfalls in data collection<br>• Make data check routines<br>• Create quality assurance plan<br>**Continuously:**<br>• Motivate participants<br>• Communicate with local sites<br>• Train new participants<br>**In case of changes:**<br>• Adjust forms, software, data dictionary, protocol, training, materials, etc.<br>• Communicate with local sites | **At the onset of participating in the registry:**<br>• Assign a contact person<br>• Control software functionality<br>• Check reliability and completeness of sources<br>• Standardise correction of data items<br>**Continuously:**<br>• Train (new) data collectors<br>• Motivate data collectors<br>• Make data definitions available<br>• Place date and initials on completed forms<br>• Keep completed case record forms<br>• Data collection close to the source and as soon as possible<br>• Use the registry for local purposes<br>**In case of changes:**<br>• Adjust data dictionary, forms, software, etc.<br>• Communicate with data collectors |
| **Detection during data collection** | |
| **During import of data into the central database:**<br>• Perform automatic data checks<br>**Periodically and in case of new participants:**<br>• Perform site visits for data quality audit and review local data collection procedures<br>**Periodically:**<br>• Check inter- and intraobserver variability<br>• Perform analyses on the data | **Continuously:**<br>• Visually inspect completed forms<br>• Perform automatic data checks<br>• Check completeness of registration |
| **Actions for quality improvements** | |
| **After data import and data checks:**<br>• Provide local sites with data quality reports<br>• Control local correction of data errors<br>**After data audit or variability test:**<br>• Give feedback or results and recommendations<br>• Resolve causes of data errors | **After receiving quality reports:**<br>• Check detected errors<br>• Correct inaccurate data and fill in incomplete data<br>• Resolve causes of data errors<br>**After receiving feedback:**<br>• Implement recommended changes<br>• Communicate with personnel |

#### *2.4.3.3 The data utilisation maturity framework*

Murphy et al. (2008) suggest a framework that puts data acquisition and utilisation into an organisational context. Central in the argumentation for an organisational data utilisation maturity framework is that the basic problem with data in many organisations is that data fails to be transformed into useful informational outcomes; according to Murphy et al. (2008) this failure is due to conflicting objectives within the organisation, wherein engineering standards, the organisations' life-cycle stage, consequence of failure, and accepted technical practice determine what data is to be acquired, and on the other hand: the strategic priorities in the organisation determine how the data will be used. The data utilisation maturity framework is therefore made to "…facilitate alignment between the data acquisition process with data use requirements…" in order to assure that data is indeed collected to produce information (Murphy et al., 2008).

Table 8: AM outcomes in the data maturity framework (based on Murphy et al. (2008)).

| AM outcome | Comment |
|---|---|
| Regulatory compliance | Is often first priority of an engineering organisation, and the only priority of "young" organisations (want to be approved to be allowed to in order to produce) |
| Time-based maintenance | Is often a result of planned maintenance or "expected need for maintenance"; an organisation with time-based maintenance has not yet reached the maturity of an organisation with condition-based AM. |
| Condition-based AM | Represents high degree of data maturity, because high quality data is needed for the models |
| Capability development | Data is used to improve and develop the organisation and inform about future developments; does also represent a high level of maturity. |

The AM outcomes in the perspective of the framework are outlined in Table 8. The level of data utilisation maturity increases down through Table 8. To perform condition-based AM, must the data utilisation maturity within the organisation be more mature than to document regulatory compliance or perform time-based maintenance. The point made by Murphy et al. (2008) about the necessity of aligning data acquisition processes and data utilisation requirements becomes evident through the analysis the AM outcomes. If an organisation has a strategic vision to maintain their assets in a condition-based manner, but the *data drivers* in the organisation are not sufficient to generate the required data, the informational outcome is not likely to be sufficient. And vice versa, if the data drivers are strong (for instance if the consequence of failure is perceived as very high) high-quality data will be produced, but if the data utilisation capability in the organisation does not match the production, the informational result is unlikely to be satisfactory, which in turn may act counterproductively on the data acquiring process.

The descriptions of data utilisation from Murphy et al. (2008) are somewhat similar to the approaches to asset management described in section 2.1.4 (Schneider et al., 2006, Ugarelli, 2008). What Murphy et al. (2008) describes as *capability development* is similar to what Ugarelli (2008) classifies as a *predictive* LCC approach. Murphy et al. (2008) put these approaches to AM in a wider organisational perspective, and identifies important organisational factors that have to be aligned in order to achieve the informational outcomes needed for high-level AM.

#### *2.4.3.4 Data integration*

The need for integration of data has been acknowledged as a prerequisite for IAM by several authors (FHwA, 2001, Halfawy, 2008, Halfawy and Figueroa, 2006, Halfawy et al., 2006a, Halfawy et al., 2006b, Lemer, 1998). Data integration is "… the process of combining or linking two or more datasets from different sources" (FHwA, 2001). Halfawy (2008) state that data needed for IAM should be integrated across:

I. A "vertical departmental axis" to accommodate the integration of different processes, software, and analytical considerations carried out at operational, tactical and strategic level within the same discipline.
II. A "horizontal multidisciplinary axis" to accommodate the multidisciplinary nature of AM, and aid the emergence of holistic and coordinated plans.

The FHwA (2001) lists (among other) increased availability/accessibility, timeliness, accuracy, correctness, integrity, consistency, clarity, completeness, reduced duplication, faster processing and turnaround time, and integrated decision making as benefits of implementing data integration within an infrastructure management organisation. Even though the organisation does not increase the quality of any data element, is the process of integrating data beneficial for data quality.

Further, the FHwA (2001) suggests that a data integration procedure should involve:

I. Analysis of the data *system requirements* through analysis of the organisation's business processes, organisational characteristics, user requirements, data characteristics and information systems infrastructure
II. Data processing and flow modelling (data flow diagrams for the different work processes)
III. Definition of alternatives, evaluation and selection
IV. Outlining of database design and specifications (data models (data structure), data standards, data reference and dictionary, communication (to interfaces) requirements, software requirements and needed management resources)
V. Development, testing and implementation

It is thus evident that an organisation's transition towards an integrated data structure may be a comprehensive transition, where work processes, technical infrastructure, organisational relations and user requirements have to be analysed.

#### 2.4.3.5 Data quality modelling

Wang et al. (1993) investigate a technological aspect of data quality problems, and argues that if data quality is not modelled into a database system (meaning that if quality measures are not included in the modelling of a database structure), it will be very difficult to assess, monitor and document the data quality in down-stream processes. Therefore, Wang et al. (1993) suggest a method for modelling data quality into databases through assigning a *tag* (tagging) to data entries, where each data type is assigned a set of *quality parameters* and *quality indicators.* A *quality parameter* is a subjective description of a data quality dimension (examples: accuracy, credibility, timeliness etc.). A *quality indicator* is an objective measure to describe a quality parameter (example: *date of entry* is an objective measurement of *timeliness*).

Wang et al. (1993) describe a method for how these tags should be assigned in the database design process. This method also includes instructions for how to include data verification (inspection) metadata into the data model (Who inspected? When it was inspected? What was the inspection procedure, and what was the result?).

**Example: Inspection of pipe.** Figure 3 shows how quality issues can be addressed into the data storage from a pipe inspection work processes (the figure is simplified). The light coloured boxes represent data repositories for pipe and inspection data. The dark boxes represent quality attributes (quality parameters; and quality indicators in square brackets []) and the verification metadata (above). For each data field is it possible to assign any number of quality attributes. As an example, the position of the pipe has an accurateness quality attribute attached to it; depending on the survey method (GPS, estimate, etc.) an analyst may evaluate the accurateness of the position differently. The same way of thinking may be used to the other quality tags.

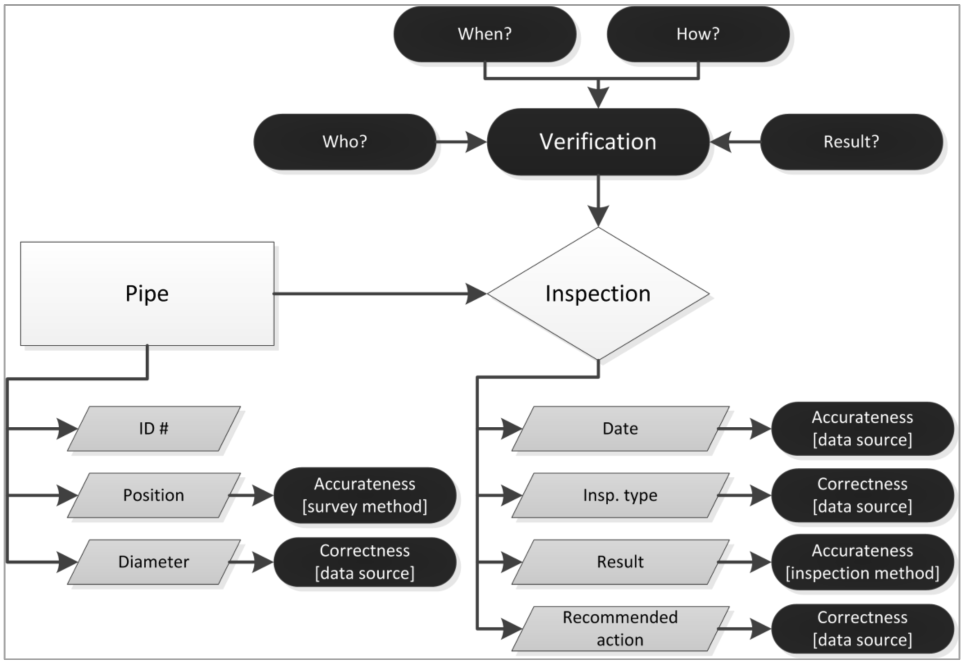


**Figure 3: Example of quality data modelling (partly inspired by Wang et al. (1993))**

The benefits of incorporating quality tagging in the design process of data structures, is that objective documentation of data quality becomes an intrinsic process when acquiring data, and this allows future users of data to assess and monitor data quality, and (maybe most important of all) filter data of unsatisfactory quality (Wang et al., 1995a). It is also necessary to store objective measurements of data quality in order to identify and react to discrepancies between perceived and objective data quality assessments (Pipino et al., 2002) – data quality may be high, but if analysts do not believe it, they will not be inclined to trust analysis results based on the data. Recording the temporal, spatial and technological (method of measurement) parameters can also help data consumers and analysts to analyse the quality of data expressed as the consistency (or correlation) along these dimensions (Weidema and Wesnæs, 1996).

Some attributes may be assigned automatically, such as *source*, *time of entry*, *method* etc. Sun et al. (2011) suggest a comprehensive set of measures for quality validation of sensor measured data, such as range detection checks, gap detection, variance check, analytical redundancy control, and tolerance band control, to detect erroneous trends in data sets. Some of these methods could be adapted to be used in validation of manually collected data as well.

#### 2.4.3.6 Quality of manually collected data

Much of the data related to buried assets is collected manually. The theory of planned behaviour (TPB) has been applied by Murphy (2009) to explain how quality of manually collected data can be improved. This theory states that a person will comply with a certain requirement (to produce quality data) if the person has the intention to comply – an intention to comply is composed of three elements:

- *Attitude toward the issue*: The data collector's attitude towards data collection is his or hers underlying belief about the importance of it. Whether or not the collector perceives data collection as important, the consequences of low quality data significant for the organisation, and/or data collection as a part of the job, will affect whether or not the collector intends to comply.

- *Subjective norm* refers to a person's perception of how people, who are significant to this person, will react to the person's behaviours. Whether or not a data collector perceives that his/hers peers or supervisors will react negatively to non-compliance, will affect the data collector's intention to comply.
- *Perceived behavioural control* refers to a person's perception of how easy it will be to execute one's intention, and is related to the confidence that one's competence (self-efficacy), capability and access to necessary resources (controllability) are sufficient to perform the intended tasks. If a data collector does not perceive to have control over own behaviour (e.g. due to time pressure), the collector is unlikely to be able to change the behaviour.

According to Murphy (2009) is the attitude an individual level issue, the subjective norm a group level issue, while the perceived behavioural control is a structural (organisation structure) level issue. Further, the role of feedback is also discussed by Murphy (2009); it is stressed that feedback is not always effective – if the perceived behavioural control is low, feedback will not serve as intended (changed behaviour), because the personnel perceive that they are not able to change their behaviour. Instead, Murphy (2009) states that all the factors of the theory of planned behaviour have to be taken into account, and suggests several initiatives (related to each of the factors) that can be implemented in order to facilitate compliance. Among the suggestions are: Utilisation of efficient technology to ease the collection process (e.g. portable computers), increase in worker autonomy (capacity to adapt behaviour), reduction of bureaucracy related to data collection, relating data collection to professional excellence, and rotation on data collection versus data treatment within work teams (consequences of low quality data are sensed by the people who collect it; increase in social pressure).

### 2.4.4 Conclusions and comparisons around data quality

Six different perspectives on how data quality can be achieved have been reviewed in this section. The succession of the different perspectives has not been chosen arbitrarily – there is an evolution from generic frameworks towards more specialised and concrete perspectives of data quality. Through the review of different data quality literature it becomes evident that the more specific perspectives may be categorised as descendants of the more generic frameworks – for instance, *data quality modelling* may be an element in the *data quality framework* related to data quality systems within an organisation. Elements from all the perspectives presented in this section may be useful for analysis, implementation and improvement of data in water utilities.

There are three aspects that return throughout the data quality literature, and these aspects are the organisational (or structural), technical and the personal aspects. The keywords of the reviewed theories have been related to these aspects, and are summarised in Table 9. As one may see, some of the terms are overlapping each other to various degrees. Regardless of this, the contents of Table 9 may be used as a reference (or checklist) for discussing and evaluating data quality issues in AM systems for engineering assets. The technical perspective will be discussed in chapter 5, while the organisational and personal perspectives will be discussed in chapter 6.

Table 9: Data quality keywords

| Aspect | Keywords |
|---|---|
| Organisational / structural | - Strategic plan/desire for informational outcomes<br>- DQ policy<br>- Organisational life cycle stage; utilisation maturity<br>- Balance between data drivers and utilisation maturity<br>- Work processes<br>- Data integration<br>- Cost of DQ versus cost[3] of low quality<br>- Corrective actions<br>- Communication |
| Technical | - Quality systems<br>- Relevant DQ dimensions?<br>- Quality verification<br>- Quality documentation<br>- Quality modelling (tagging); quality parameters, indicators, attributes<br>- Manufacturing “data products”<br>- Standards and references dictionary, data dictionary<br>- Software quality<br>- Data collection methods<br>- Data control routines and methods<br>- Consistency over time and agreement between evaluators |
| Personal / group | - DQ motivation<br>- DQ achievement measurements<br>- Attitude, subjective norm and perceived behavioural control<br>- Feedback loops<br>- Training<br>- Access to definitions<br>- Data collection close to source |

## 2.5 Theory review summary and conclusions

The theory review has encompassed four main elements, in increasing specificity. The main concepts of IAM have been reviewed, followed by a discussion about the role of data, model use in IAM, and finally have the problems regarding data quality been addressed.

Hopefully has this theory review given the reader the impression that a main fundament for AM is the availability of quality data, depending on the AM tools and models used, and that data quality management is a process that is necessary for both facilitation of production and maintenance of data quality within an organisation. Further, it is important to emphasise that both AM and data (quality) management are indeed multidisciplinary processes that require suitable technical and organisational conditions in order to flourish. An organisation’s transition towards predictive asset and data (quality) management is a process requiring review and evaluation of organisational structure, work processes, technical systems and so forth.

---

[3] Consequence

The different theoretical aspects that have been reviewed in this section will be the basis for the analyses and evaluations throughout the thesis.

# 3 The life cycle of a buried asset

The purpose of this chapter is to describe the different intervention events and work processes that can occur with a buried asset throughout its life cycle, and assess how data collection processes can be adapted to these events. This chapter may be viewed as a concretisation of section 2.2 – if asset management (AM) takes the whole life cycle of an asset in consideration, one should analyse the life cycle of the assets in question, divide the life cycle into representative events, and identify which data from these events that are relevant for AM.

As previously mentioned the analyses in this thesis focus mainly on buried assets, and the physical deterioration of these. Energy costs will not be discussed explicitly.

## 3.1 Buried asset life cycle: from planning to decommissioning

Figure 4 shows a life cycle of a generic buried asset, showing the deterioration, failure events, and possible interventions the asset owner can employ on the asset. It is important to clarify that not all of the events and interventions have to occur to all buried assets, and also that the events can occur several times. (The occurrence and timing of interventions would be dependent on the asset owner's AM strategy, for instance condition-based, reliability-centred, corrective or time-based (Schneider et al., 2006)).

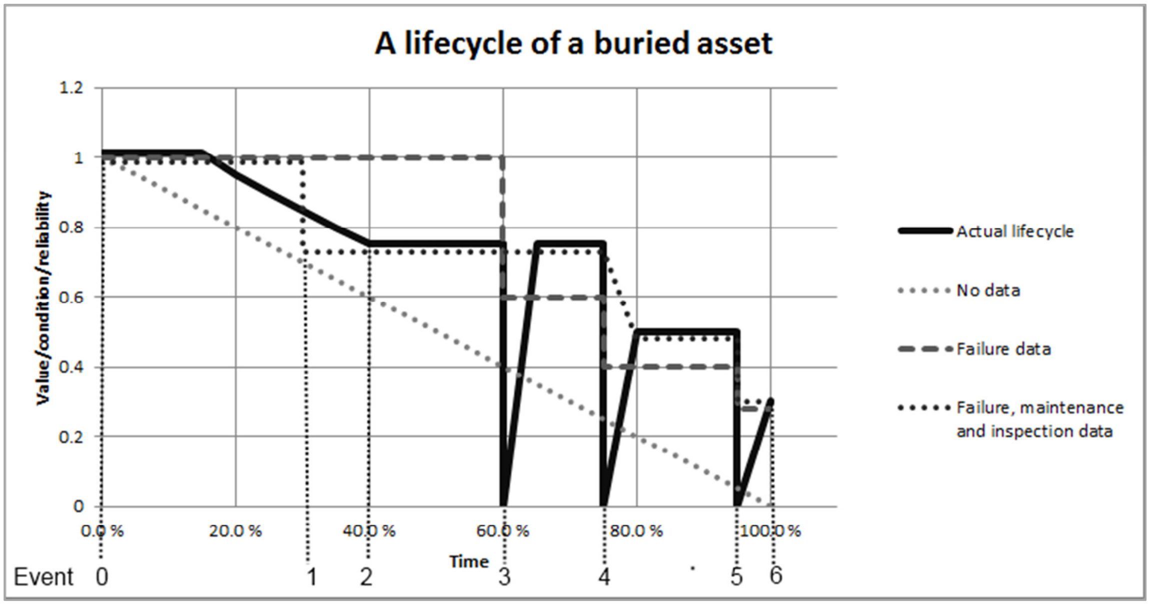


Figure 4: Three ways of keeping track of the condition of a buried asset

The identified interventions (actions performed on the asset by asset owner) are:

- **Commissioning**, where the asset is set to fulfil its functions (**event 0**)
- **Inspection**, where the asset's condition or performance is evaluated (**event 1**). Infrastructure inspections are described in Frangopol et al. (1997).)
- **Maintenance** is a planned procedure where the asset's deterioration is mitigated, either by reducing deterioration rate (as described by van Noortwijk and Frangopol (2004)), or by

increasing the condition or performance (**event 2**). In this example is the deterioration rate reduced. The modelled change in deterioration rate is similar to the reliability or condition changing index principle described by Kong and Frangopol (2004): the maintenance action changes the profile of the deterioration rate. In this chapter will there not be made a distinction between maintenance and larger rehabilitation work – conceptually speaking are these interventions similar, as long as the asset is not replaced.

- **Repair** is an unplanned procedure to mend a detected failure. Repairs can increase, decrease or retain the condition, performance or reliability the asset had before failure (Babykina and Couallier, 2010) (**event 3, 4 and 5**).
- **Decommissioning** occurs when the asset is put out of service, and is not expected to fulfil a function anymore (**event 6**).

**Normal operation** occurs in between these events.

In Figure 4 are three different ways of keeping track of the asset life cycle shown (the figure is partly inspired by Misiūnas (2008) and Kong and Frangopol (2003)). The vertical axis may represent any number of "condition parameters" (economic value, structural condition, reliability, performance) – the figure shows a conceptual progression of the evaluation of the overall "state of the asset". The figure exemplifies that the *objectivity level* of decision-making processes is dependent on the level of available data:

- Having no life cycle data one usually assumes a steady decay of asset condition as it ages, based on a fixed expected service life or based on experienced service life.
- If only failure data is recorded, the condition of the asset will be evaluated by the number of failures (but also based on the age) – an asset owner will lose confidence in an asset's reliability as the failure rate increases.
- If both failure, inspection and maintenance data is recorded, the condition of the asset will be based on all this data. The condition of an asset will be re-evaluated whenever new operational data is registered.

Naturally, the recorded data is possible to use in models for forecasting of future system behaviour or extrapolation to other areas/groups of assets. Although not an intervention event, the *normal operation* procedures may also be a source of information about the buried asset's life cycle.

As an example, on could assume that the asset in Figure 4 were a fire hydrant with a metal-based material, suffering structural and hydraulic deterioration due to corrosion. Table 10 shows an *example* of how an asset can be managed through systematic recording and continuous evaluation of operational data.

Table 10: An example of a recording of an asset life cycle and data-driven decision-making

| Event | Action | Explanation |
|---|---|---|
| **0** | Commissioning | The asset is perceived as functioning 100 %, and having full value. |
| **1** | Inspection | Results from a planned inspection are registered, and the state of the asset is re-evaluated. The test could for instance be a capacity or leakage test. Inspection results for the asset are recorded. |
| **2** | Maintenance | As a consequence of discovering deterioration (event 1), it was decided to invest in a valve flushing, cleaning and replacement of deteriorated gaskets and corroded bolts. Details about the work are recorded. An evaluation of the impact of maintenance is done. |
| **3** | Failure and repair | A failure in the form of a leak is detected and a repair action is performed. Details about the failure and repair are recorded, and an evaluation about the impact of the repair is done. In Figure 4 the first repair has an "as bad as old" (ABAO) effect, whereas the two consequent repair have a "worse than old" (WTO) effect (Babykina and Couallier, 2010). |
| **4** | Failure and repair | |
| **5** | Failure (and repair) | |
| **6** | Decommissioning | After a number of failures and repairs, it is "confirmed" that the state of the asset is severely impaired. This leads to a decision that it should be replaced. This decision can be made based on economic considerations, reliability, level of service etc. The asset can still have a remaining value. |

## 3.2 Analysis of the intervention events in a buried asset's life cycle

As suggested by the FHwA (2001), one should analyse the work processes in an organisation in order to unveil the data input and output from these processes, and hence define the requirements of the data system. In this section are the different intervention types defined in section 3.1 analysed with respect to potential data sources. In section 2.2.1, the "six whats of AM" (Vanier et al., 2006) were reviewed with respect to data needs; now the question to be answered is: How can data recording and data utilisation (model use) throughout the life cycle of a buried asset help to answer the "six whats" of AM? Data sources from each intervention event will be identified with these questions in mind.

The data recording measures suggested in this section both refer to data that describe a specific asset, a group of assets or the whole network of assets, without making an explicit distinction. The results from the discussion and analysis in the following subsections are summarised in Table 11 (page 33).

### 3.2.1 Planning, construction and commissioning

The conception of an asset begins by the planning of the asset. It is fair to assume that an asset is planned because the asset owner wants it to fulfil a certain set of functions; the planning and construction ensures that the asset will comply with regulatory requirements (Murphy et al., 2008) and perform adequately (sufficient flow, pressure etc.) (Kyle et al., 2000). Kyle et al. (2000) state that much of the data, which is relevant for service life oriented AM, is generated in the planning (pre-delivery) phase of a project, but that the asset owner usually retrieves most data after the asset is fully operational. Routines for connecting data from the planning phase of an asset to the life cycle evaluation of the asset will consequently be beneficial. Central data sources in the commissioning phase are:

- Much of the data from the planning and construction phase may be transferred from the contractor, and give a detailed picture of the asset specifications, and thus enabling the assessment of what the utility owns. Data about the environment in which buried assets are situated can be extracted from external GIS sources (soil conditions, traffic etc.).
- The value of newly commissioned assets can be expressed in terms acquiring cost; however, the value of the asset could also be expressed through the value of the service that it is providing, or what one would be willing to pay not to lose it (Lemer, 1998). (Later in the life cycle, the replacement cost is a more relevant value measure.)
- If there are factors with a newly commissioned asset that are deviating from the planned specification, it should also be possible to uncover and register this in the commissioning phase (as-built quality).
- With a new asset, one does not know how fast it will deteriorate, but the utility may have previous experience with similar assets under similar conditions, and may assign this as a "first estimate" of the expected service life. This also goes for the expected time until inspections and maintenance must be carried out.
- The expected time until inspections and maintenance should also be dependent on the criticality of the asset (important assets must be inspected earlier, because consequence of failure is high). Criticality should hence be assessed in the commissioning phase.

### 3.2.2 Normal operation

Normal operation is not a physical intervention on the buried asset. However, it is possible to obtain life cycle data about the asset from the normal day-to-day operation. Important data sources in the operational phase of an asset are:

- Water utilities who have hydraulic models, combined with live measurements from flow and pressure gauges in the network, can extract information about the hydraulic criticality, the utilisation rate and the hydraulic performance of assets. Anomalies in the real-time hydraulic models/measurements may often be the basis for an inspection reaction (section 3.2.3). A rapid reaction to a detected (and unexpected) pressure drop could help mitigating the consequences of a potential burst. Burst (or leak) position detection and condition assessment can be made from hydraulic transient monitoring (Karney et al., 2008, Misiunas, 2003, Misiūnas, 2008, Misiunas et al., 2005). Continuous acoustic monitoring for leak detection has also proved to be an effective measure, allowing for a rapid repair response (Bhagwan, 2009). Also, low hydraulic performance may be an indicator of internal corrosion (Engelhardt et al., 2000) and the overall hydraulic performance may be an important determinant for rehabilitation decisions (Kleiner et al., 2001).
- Complaint occurrence and complaint management may also be a source to asset life cycle data. A complaint leads to a reaction from the water utility, perhaps in the form of an inspection (section 3.2.3) work order. In some cases does the investigation lead to a repair (section 3.2.5), while it in other cases is dismissed. Keeping a registry of complaints, complaint sources, and utility reactions to the complaints may be useful for the utility. Complaints that lead to corrective actions might be registered as repairs; however, all complaints should be filed, even if no corrective action is made – the number of complaints may be used as a performance indicator (PI) (Alegre et al., 2006b), and recurring complaints without corrective actions may be a sign of general customer dissatisfaction or lack of

complaint investigation effectiveness (the real reasons for the complaints are not uncovered).

- Water quality samples can be used to form quality PI’s (Holthe, 2010). A water utility is obliged to collect water quality samples around the network, in order to document that the water is safe; water quality samples can also be used for more than that. Some quality parameters may be indicators of asset deterioration, such as metal corrosion (Broo et al., 1997, Engelhardt et al., 2000). Water quality data is potentially a large source of data, since many samples are taken regularly for regulatory compliance. If water quality results are stored and related to their tapping point (for instance customer address, or connection point), they can prove to be an indirect indicator of the local asset conditions, and also a direct measure of the level of service. Also, the combination of water quality measurements, customer complaints and hydraulic modelling can make it possible to trace back contaminant sources. (It can of course be argued that the extraction of water quality samples indeed is a form of inspection; here it is chosen to classify water quality sampling as a part of the normal operation.)
- It is important to always keep rough estimates on what cost of replacement for a comprehensive cohort of assets are, so that “repair or replace” decisions can be assessed more effectively when a failure or low inspection rating finally occurs. Cost of replacement can be assed based on cost data from previous replacements, material, diameter, soil condition, area use etc.
- Deterioration and failure forecast modelling on asset cohorts may be run continuously under normal operation of an asset, by the aid of explanatory factors from the asset registry and the failure, inspection and maintenance history (Kleiner and Rajani, 2001, Le Gat, 2008, Le Gat, 2009, Poulton et al., 2007, Røstum, 2000, Wirahadikusumah et al., 2001). If operational data is integrated with the asset data, the models can continuously be updated by up-to-date asset history, enabling production of improved estimates on condition and remaining service life.

### 3.2.3 Inspection

An inspection of an asset usually results in an estimate of the asset’s condition or performance, through some sort of condition (in the wide sense) measurement. However, it is possible to achieve more from an inspection than a simple condition estimate:

- When an asset is being inspected, it can be beneficial to control if the asset registry complies with reality. Buried water distribution assets usually have a long history, and asset registries are usually not complete and completely accurate (USEPA, 2008, Vanier and Rahman, 2006), both with respect to *what* it is and *where* it is (Vanier, 2006).
- Modelling of deterioration on asset cohorts has already been discussed in section 3.2.2. Inspection results give the opportunity to update, revise and verify modelling results. An inspection result may be only a condition grade, or an array of measurements (for instance a flow pressure curve for hydrants).
- Asset value and deferred maintenance can be estimated through modelling the cost of condition (and reliability) increase to original level, based on the relationship between previously registered costs of maintenance and maintenance results (Kong and Frangopol, 2004).

- Contextual factors around an inspection action will also prove useful for the optimisation of inspection routines (Frangopol et al., 1997). Whether or not an inspection was planned or a result of a complaint might be relevant for optimizing inspection intervals. Also the cost of an inspection (work hours and materials) is relevant for the LCC optimisation of the asset (Frangopol et al., 1997). Analysis of the cost-effectiveness of inspections is one way of rationalising expenditures.
- The *recommended action* is suggested as a subjective possibility for the inspectors to recommend what course of action that should be taken, based on their expert assessment of the situation. For instance may such evaluations be useful when leakage inspections are being performed, or when the severity of a problem is difficult to quantify objectively.

### 3.2.4 Maintenance

A maintenance action is performed in order to extend the remaining service life of an asset, by improving its condition and/or reducing its condition deterioration rate. *Why* a maintenance action was performed on an asset, what was actually *done*, what the estimated *effect* of it, and recommended *follow-up work* is important data:

- It is interesting to document why the maintenance work order was created (recommended from inspection, time-based maintenance approach etc.), in order to be able to assess what are the precursors for maintenance decisions.
- Registering what was done to the asset will also help to keep track of the asset life cycle. If vital (or valuable) parts were changed, this changes the assessment of “what you own” and “what it is worth”; equally, if characteristics of the asset are significantly changed (for instance if a water main is relined), this also changes “what you own”.
- It is also interesting to register the maintenance expenditures (work hours, material and equipment) in order to monitor the LCC of the asset, and reassess the deferred maintenance.
- In order to reassess the state of the asset, it is necessary to report the effect of the maintenance work. The effect can be assessed through a condition or performance measurement, or be extrapolated from models that model expected increase in condition or performance. It is possible to model the effect of maintenance, using the factors that affect the maintenance decision as model predictors (Prozzi and Hong, 2010).
- As for other interventions, when a maintenance operation is executed, skilled personnel might have an opinion on recommended follow-up work on the asset.

It is important to emphasise that there is a relationship between cost and effect of a maintenance operation (Kong and Frangopol, 2004), and that both cost and effect needs to be recorded (and ideally develop a relationship between them), in order to be able to reassess the value, condition, remaining service life and deferred maintenance.

### 3.2.5 Failure and repair

In this context is a repair defined as a corrective action that is executed after a failure has been detected (and a failure is defined as the inability to perform a required function). Therefore are repairs, unlike maintenance operations, often unplanned and with primary focus on re-establishing the functionality of the asset that has failed, rather than extending the service life. In any case are data about asset failures valuable to establish an assessment of the “health of the assets”. Some important aspects to record from a failure and repair incident are:

- The cause and mode of failure are important parameters for assessment of the condition and reliability of the assets. Cause of failure distinguishes different deterioration *agents* – e.g. the reliability and condition of an asset that has burst due to corrosion may have a very different reliability and condition than an asset that has burst due to third-party causes (e.g. root intrusion, weight from tram etc.). On the other side, the *mode of failure* distinguishes different deterioration *mechanisms* and their consequences. This data can be obtained through operator observation and evaluation. External GIS data sources may be an informational well for the cause and mode of failure.
- The cost of repair is necessary for the assessment of the total LCC. Thus is it necessary to record work hours, material and equipment expenditures, as well as indirect costs as best as possible. The indirect costs of failure are (among other things) related to the *severity* of failure (Frangopol et al., 1997, Kong and Frangopol, 2003, Kong and Frangopol, 2004), which is a factor that should also be estimated in the repair report. Keeping track *of how the failure was detected* (complaint, inspection, monitoring etc.) may help to assess the severity of the failure, and if failure detection measures are lacking
- A failure and consequent repair intervention may have an *effect* on the characteristics of the asset (what do you own), the condition, the reliability, and the remaining service life of the asset. The repair may improve, worsen or conserve the overall condition of the asset (Babykina and Couallier, 2010). Tracking what has actually been done to rectify the failure, accompanied by experienced evaluations (or asset histories), can give an indication of the effect of the repair. As for maintenance, the effect of repairs can also be modelled (Prozzi and Hong, 2010).
- As with the other interventions should evaluations about recommended follow-up work on the asset should be logged.

### 3.2.6 Decommissioning

When decommissioning is planned, an important question to answer is: *Why* is it decommissioned? The answer to this question is the key to assessing the asset's characteristic properties at the end of the life cycle. An asset may be decommissioned because its service is no longer needed, higher function requirements exceed the performance of the asset (for instance higher flow demands), it has reached its maximal technical or economic service life, or because third party activities make it desirable to renew it prematurely (e.g. road works). In practice is decommissioning ordered because the asset owner decides it (Le Gat, 2009, Cemagref, 2010). This decision may be based on formal evaluations (e.g. economic evaluation in a decision support system) or, in contrast, it may only be based on tacit experience about the asset (or asset type). Recording of data about why the asset was decommissioned and how that decision was supported will give the utility the possibility to extract information about genuine service lives, and what factors affect the decommissioning decisions.

## 3.3 Conclusions

Chapter 3 has described an analysis of:

I. How the "health" (condition, reliability, value, risk) of a buried asset can change, due to continuous deterioration and discrete events, such as inspections, maintenance, failures and repairs.
II. How data collection from the different events can aid a utility to obtain the information necessary to implement a predictive AM strategy on its assets.

The events that have been analysed are: commissioning, normal operation, inspection, maintenance, failure and repair, and decommissioning. The data source results are summarised in Table 11 (following page).

All the interventions that have been analysed are potential sources of data that can help to answer the “six whats” of AM, in one way or another. Some data is directly usable whereas other data has to be run through a model in order to produce informational outcomes. It may not be realistic or necessary that all the data that has been suggested in this section will be collected by a water utility. However, the analysis of the asset life cycle that is carried out in this section, has hopefully convinced the reader that data from all types of interventions are necessary to get a just perspective of the asset life cycle. Only collecting failure data, without keeping track of maintenance and inspections will give a skewed perspective, especially with regards to the life cycle costs. All the different sources of information are pillars on which the life cycle evaluation of the asset portfolio relies on – taking one of the information sources away will tilt the this platform away from the balanced evaluation.

The method of analysis may be criticised. Applying the “six whats” of AM as basis questions for each of the analyses carried out in this section may not conform to all aspects of IAM; for instance discussing risk and performance is difficult within that framework – nevertheless, performance can be expressed as condition (in the broad sense), and risk is implicitly accounted for in the “What do you fix first?” question. The “six whats” of AM are an expression of the six facets defined by Vanier et al. (2009) - inventory, performance, service life, life cycle cost, criticality and alternatives – these facets are instruments used to answer the six questions. Further may the reader have sensed that there is a strong relation between the condition, value, deferred maintenance and remaining service life, and that the data sources for determining these parameters in many cases are partially overlapping. The redundancy (overlap) between the questions does however reflect that there are strong correlations between the parameters and that collection of data (and the application of models) can aid to establish quantifications of the relationships between them.

**Table 11: Results from the analysis of data sources for a buried asset's life cycle**

| Phase | Question | Planning data/inventory | External/environmental data | Data audit (control with reality) | Value of service | Operator evaluation | Hydraulic model | Gauging | Complaint management | Water quality samples | Reliability/condition modelling | Cost-effect modelling | Alternative costs | Intervention cost report | Acquiring cost | Service life/LCC deterioration model | Reason for intervention | Failure/deterioration mode | Intervention action report |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Commissioning | What do you own? | ■ | ■ | | | | | | | | | | | | | | | | |
| | What is it worth? | | | | ■ | | ■ | | | | | | | | ■ | | | | |
| | What is the deferred maintenance? | | | | | | | | | | | | | | | | | | |
| | What is the condition? | | | ■ | | | | | | | | | | | | | | | |
| | What is the remaining service life? | | | | | | | | | | | | | | | ■ | | | |
| | What do you fix/do first? | | | | | ■ | | | | | | | | | | | | | |
| Operations | What do you own? | | | | | | | | | | | | | | | | | | |
| | What is it worth? | | | | | | ■ | ■ | | | | | ■ | | | | | | |
| | What is the deferred maintenance? | | | | | | | | | | ■ | ■ | | | | ■ | | | |
| | What is the condition? | | | | | ■ | | | ■ | ■ | ■ | | | | | | | | |
| | What is the remaining service life? | | | | | | | | | | ■ | ■ | ■ | | | ■ | | | ■ |
| | What do you fix/do first? | | | | | | ■ | | ■ | | ■ | | ■ | | | ■ | | | |
| Inspection | What do you own? | | ■ | ■ | | | | | | | | | | | | | | | |
| | What is it worth? | | | | | | | | | | | | | ■ | | | | | ■ |
| | What is the condition? | | | | | ■ | | | | | | | | | | | | | ■ |
| | What is the deferred maintenance? | | | | | ■ | | | | | | ■ | | | | | ■ | | |
| | What is the remaining service life? | | | | | | | | | | ■ | | | | | | | | ■ |
| | What do you fix/do first? | | | | | ■ | | | | | | | | | | | | | ■ |
| Maintenance | What do you own? | | | ■ | | | | | | | | | | | | | | | ■ |
| | What is it worth? | | | | | | | | | | | ■ | | ■ | | | ■ | | ■ |
| | What is the deferred maintenance? | | | | | | | | | | | ■ | | | | | ■ | | ■ |
| | What is the condition? | | | | | ■ | | | | | | | | | | | ■ | | ■ |
| | What is the remaining service life? | | | | | | | | | | | ■ | | | | | | | ■ |
| | What do you fix/do first? | | | | | ■ | | | | | | | | | | | ■ | | ■ |
| Failure | What do you own? | | ■ | ■ | | | | | | | | | | | | | | | ■ |
| | What is it worth? | | | | | | | | | | | | | ■ | | | | | |
| | What is the deferred maintenance? | | | | | ■ | | | | | | | | | | | ■ | ■ | |
| | What is the condition? | | ■ | ■ | | ■ | | | | | | ■ | | | | | ■ | ■ | |
| | What is the remaining service life? | | | | | | | | | | | ■ | | | | | | | |
| | What do you fix/do first? | | | | | ■ | | | | | | | | ■ | | | | | ■ |
| Decommissioning | What do you own? | | | ■ | | | | | | | | | | | | | | | ■ |
| | What is it worth? | | | | | | | | | | | ■ | ■ | | | ■ | ■ | | ■ |
| | What is the deferred maintenance? | | | | | | | | | | ■ | ■ | | | | ■ | | | |
| | What is the condition? | | | | | ■ | | | | | ■ | | | | | ■ | ■ | | ■ |
| | What is the remaining service life? | | | | | | | | | | | ■ | | | | ■ | ■ | ■ | ■ |
| | What do you fix/do first? | | | | | | | | | | | | | | | | | | ■ |

A clarification that has to be made is that not all data that can be relevant for IAM should be linked to a specific asset. For instance has water quality data been mentioned as potential information source – however, linking water quality to a single asset is neither feasible nor completely true; the water quality at a tapping point is affected by a range of assets along the pathway of the water from production and through transportation.

Two generic types of data have been referred to interchangeably in this section: raw data and model data output. The different types of data that were identified in this section have been systemised in Figure 5. One may see that raw data has been classified into nine different (square) types, which are fed to either the GIS database (asset registry) or a set of databases for operational data, which again may feed data models, or be used directly to monitor the state and criticality of the asset. Model results can again be returned to the operations databases.

The findings in this section may be the basis for an evaluation of how well VMW is able to record the life cycle evolution of their buried assets.

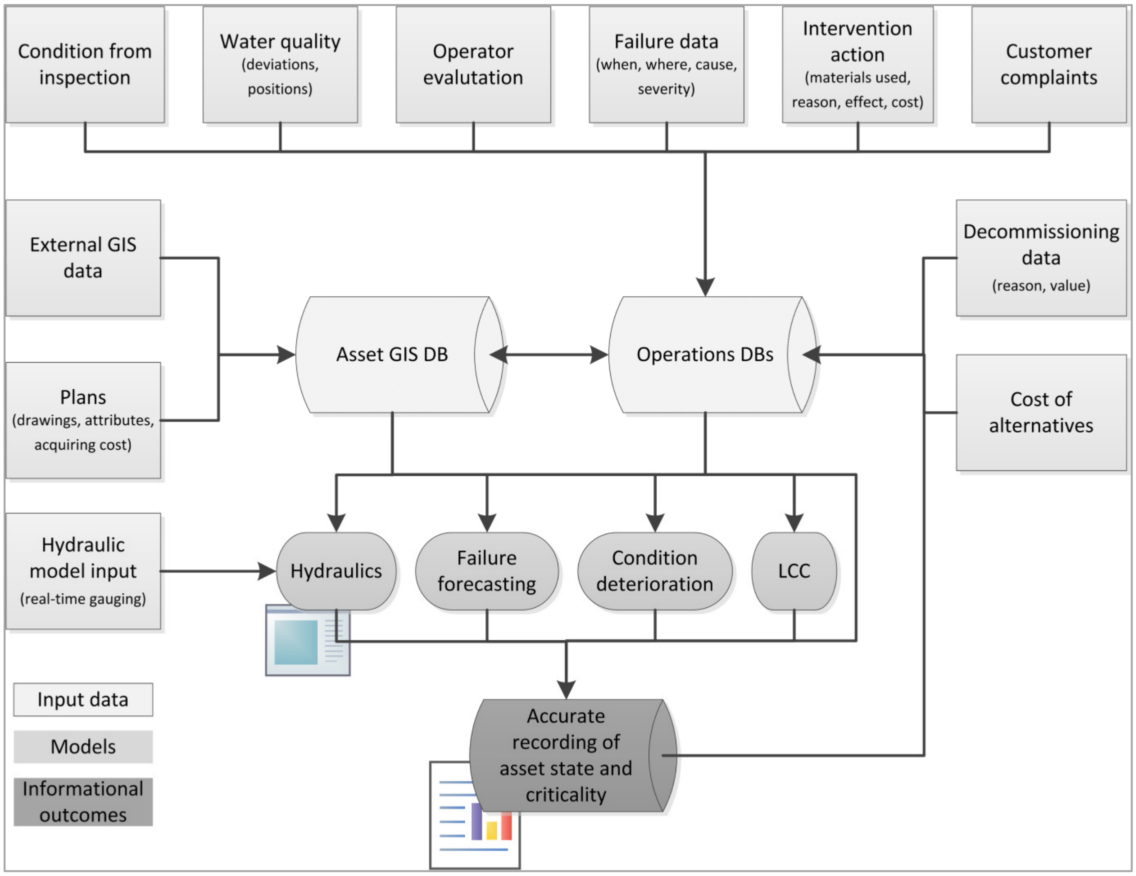


Figure 5: Data collection following the life cycle of a buried asset

# 4 Review of VMW's AM toolbox

Several software tools that aid asset management (AM) are currently being evaluated by VMW, some have been decided to be implemented, and some are already implemented. The most relevant tools for life cycle data will be described here:

- USTORE
- Octopus
- Interactive maps / Mobiele data
- LCC software
- InfoWorks

VMW has a geographic information systems (GIS) inventory database of their buried assets. The GIS database will not be described explicitly, since the database structure is fairly standard and the concept of GIS in water utilities is well known (Halfawy and Figueroa, 2006). However, applications that are dependent on GIS will be described in this section. What also will not be described in detail are the laboratory information management system (LIMS) and the invoicing system (ARCADO) – it is duly noted that they exist.

## 4.1 USTORE

### 4.1.1 Description

USTORE (*uniform storingsregistratiesysteem [Dutch]*) is a failure registry database developed at KWR Water Research Institute and is a shared database of failure incidents gathered from several participating water utilities. KWR has developed a uniform manner of registering failure data, either by standard paper forms or by web applications, which are used for registering failure events in the water utility database; further is the data exchanged (or copied) to KWR's central database, via a web application (USTOREweb), and can be utilised for statistical analysis (both by KWR and the participating utilities), rendering greater possibilities than with the data of each utility alone. (Vloerbergh et al., 2011)

The idea of USTORE is that a uniform registration and sharing of data should result in more valuable knowledge about buried asset degradation much quicker than if the water utilities collected the information by themselves, thus giving *each failure more value* in the form of statistical information. Eight[4] of the KWR member water utilities are participating in the USTORE project with varying degree of implementation level, and the database has over 4 000 incidents registered so far (Vloerbergh and van Thienen, 2011).

The data collected for a failure incident is illustrated in Figure 6. In the centre are the details about the incident, such a spatio-temporal specifications, and the cause and nature of the failure; further out are the characteristics of the asset (component type, material, age etc.); even further out are the specifications about the surroundings, such as soil, groundwater conditions, coverage and traffic. Furthest out are the situational factors.

---

[4] Brabant Water, Dunea, Evides, Waterleidingbedrijf Noord-Holland, Vitens, Waterbedrijf Groningen, Waterleiding Maatschjappij Drenthe, Waterleiding Maatschappij Limburg, Waternet

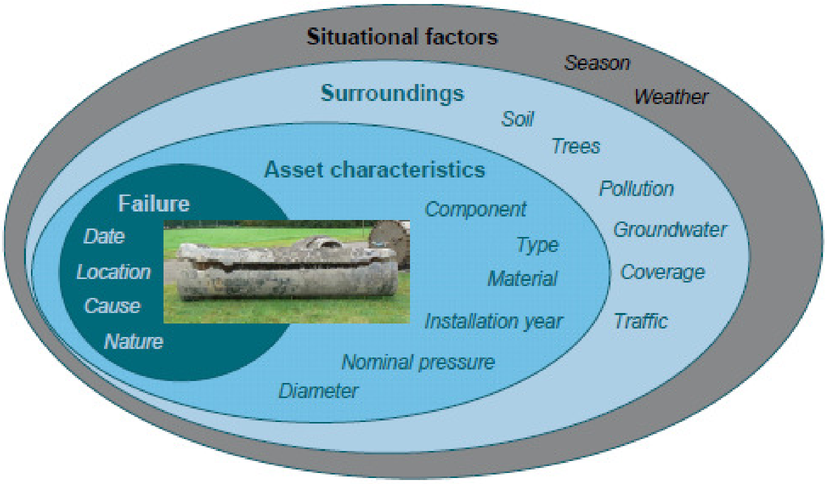


Figure 6: Levels of factors incorporated in USTORE (copied from Vloerbergh et al. (2011))

Table 12: Input fields in USTORE (simplified)

| Area | | Input | | Format |
|---|---|---|---|---|
| 0 | Company | | Company code | selection list |
| 1 | Date | | Registration date | numeric |
| | | | Failure date | numeric |
| 2 | Location | | GPS XY-coordinate | numeric |
| | | | or address | alphanumeric |
| 3 | Failing object/nature of the failure | 3a | Failed object type | selection list |
| | | | Nature of failure | selection list |
| | | 3b | Type of joint | selection list |
| 4 | Characteristics of the failing object | 4a | Material | selection list |
| | | | Material type | selection list |
| | | | External protection | selection list |
| | | | Internal protection | selection list |
| | | 4b | Installation year | numeric |
| | | 4c | Pressure category/wall thickness | numeric |
| | | 4d | Diameter | alphanumeric |
| 5 | Cause of failure | | Cause 1 | selection list |
| | | | Cause 2 | selection list |
| | | | Cause 3 | selection list |
| 6 | Context factors | 6a | Soil type | selection list |
| | | 6b | Pollution | selection list |
| | | 6c | Trees | selection list |
| | | 6d | Groundwater | selection list |
| | | 6e | Ground coverage | selection list |
| | | 6f | Traffic | selection list |
| 7 | Valve functioning | 7a | Need to close valves? | selection list |
| | | 7b | Did the valves function? | selection list |

Table 12 shows the input fields in USTORE (a few additional “special case” input fields have been excluded for simplification). As one can see, most of the input fields that are not based on numeric input (dates and coordinates etc.) are selection lists. That means that the user has a limited list of possible inputs to select from. For instance, for *cause of failure (5)* is it possible to select:

- third party damage
- external degradation
- internal degradation
- installation
- external load, namely:
- other, namely:
- unknown

Usually is the number of selections from three to ten in the selection lists. It is possible to submit *other* as input, where one also can write the cause if no other reason is suitable. This is common for the other input fields in USTORE as well.

Other input fields are being evaluated by KWR, e.g. data describing erosion craters created by fractures, the possibility to include situational photographs etc. (Diericx, 2012).

### 4.1.2 Comments and evaluation

It is evident that the idea of combining efforts to create synergy effects of data from several water utilities is a powerful one. If one sees the USTORE project in the data utilisation maturity framework (section 2.4.3.3) one may say that a rapid production of informational outcomes can be achieved, even though the drivers for data collection are not fully developed. The standardised manner of input in USTORE gives the participating water utilities a solid framework for failure data input, where consideration has been paid to what data should be collected, and in what format it should be stored; the different *levels of input data* in Figure 6 are in compliance with the data needs that have been expressed in the literature (Kleiner and Rajani, 1999, Le Gat, 2009, Wood and Lence, 2006, Wood et al., 2007). Further, the rapid feedback of information from USTORE should help exemplify the importance and value of the data collection, which in turn, if handled rightly, will improve attitude towards the value of manually collected data (section 2.4.3.6). The fact that KWR are already able to produce reports on trends in failures based on USTORE data (Vloerbergh et al., 2011, Vloerbergh and van Thienen, 2011) works as a primer for committing to USTORE.

However, there are certain limitations and pitfalls associated with USTORE:

- *Firstly*, one must assess the similarity between the VMW network and the other participant water utilities. If there is reason to believe that there are significant differences between parameters such as material qualities, common modes of failure or (most of all) maintenance history and maintenance practice, the benefits of USTORE will diminish.
- *Secondly*, not all water utilities participating in the USTORE project have successfully implemented it; according to Vloerbergh and van Thienen (2011) this is due to the way the utilities are organised, especially the way information technology services are implemented in the organisation. The prerequisites for successful implementation should be evaluated before deciding to participate in USTORE.

- *Thirdly*, USTORE lacks control of data quality and reportage percentage, and does not have the possibility to assign levels of uncertainty to the input data (data quality tagging is not included).
- *Fourthly*, USTORE only stores data about the failure, not about the repair. Data about cost of failure, repair methods, assessment of condition after repair, etc. are not stored in USTORE.

In addition to these insufficiencies, KWR has also identified *insufficient help functions and instruction material, lack of operator motivation,* and *lacking interaction between USTORE and systems that are internal in the utilities* as problems with USTORE (Vloerbergh and van Thienen, 2011).

When all these limitations have been mentioned, is it important to emphasise that USTORE should not be perceived as a barrier against the collection of other or additional data. A water utility may very well collect data about inspections and planned maintenance work separately from USTORE, and with adaption is it possible to include additional data about the failure incidents in the local registries. USTORE is a developing concept, and several of the issues discussed here will be resolved as the concept evolves.

Using USTORE without additional data storage solutions will not be sufficient for a predictive AM strategy, since important events and data about the life cycle are not in the scope of USTORE (repair data are excluded). However, USTORE is an important step forward for water utilities that do not yet have a standardised way of collecting failure data (such as VMW), and should not be perceived as "the solution" for life cycle data, but rather be perceived as one component in the array of life cycle data recording tools.

## 4.2 Octopus

### 4.2.1 Description

Octopus is an internal project in VMW that is still in the planning phase. The main objective of Octopus is to organise and systemise the day-to-day operational management of VMW. In essence is Octopus thought to become a set of data collection and retrieval procedures, which are adapted to different work processes related to operational management of the water distribution network. The name *Octopus* refers to the fact that Octopus is supposed to extract data from several data sources in VMW, and connect these sources like the arms of an octopus, in order to make a system that follows operational work processes. Octopus will "stretch its arms" to combine information from the GIS database, customer services, the VMW invoicing system (ARCADO) etc.

It is planned that each work process will have its own report form, that the forms on the long term will be web-based and adapted to portable computers, and that each report form is connected to a GIS viewer application, thus making it possible to connect work processes to a map presentation. (A GIS viewer is an application that displays geographical information and allows the user to perform basic GIS queries (Dempsey, 2010). GIS viewers are available for web browsers (Dempsey, 2011)).

Table 13 shows the different work processes that are planned to be modelled in Octopus. (The setup is similar for work processes related to waste water.) The table shows that Octopus is supposed to aid follow-up of a set of work processes.

Table 13: Work processes in Octopus (waste water is excluded). Translated from Hulpiau (2012)

| Connection | Maintenance | Defects and complaints | Projects |
|---|---|---|---|
| New | Inspection and repair of fire hydrant | In-house leakage | Adjustment |
| Renew | Inspection and repair of valve | Outside leakage | Enlargement |
| Modification | Flushing of pipe | Water quality problem | Renewal |
| Reopening | Control of cathodic protection | Pressure problem | |
| Close | | | |
| Standpipe | Phase 1 | Phase 4 | |
| Gauge replacement | Phase 2 | Phase 5 | |
| Gauge reading | Phase 3 | Phase 6 | |
| Testing | | Phase 7 | |

As an example to illustrate how Octopus is supposed to work, one could consider a situation where a leakage is discovered, reported to VMW, and has to be repaired:

I. Customer services in VMW receive a complaint. If the complaint is by phone, it will be manually registered by customer services with the aid of a GIS viewer. If the complaint is by VMW's web page, the customers can use a GIS viewer themselves. The complaints are stored.
II. Octopus will make a connection to USTORE (if VMW decides to use USTORE), ensuring that the registration form for the incident is produced to the repair personnel.
III. Octopus will make a connection to the calendars of operational personnel in the area of the complaint in order to find who the on-call responsible is, and forward the complaint message (within normal office hours will the message be sent to the local service centre's supervisor). The complaint message medium will typically be a SMS or email.
IV. Operational personnel receive the complaint message, with details about the complaint (address, GIS viewer map etc.), investigate and make a decision about the course of action.
V. If a repair is necessary, the operational personnel will investigate which valves they have to close. A GIS viewer display on a portable device will help them to identify the valves they have to close, and a connection between the invoice system and GIS will help identify the customers that will be affected when the valves are closed. Octopus will then prepare a message with information about the inconvenience to the affected customers.
VI. The repair is executed by the operational personnel, and they are required to report details about the performed operation, such as the time spent, cause of failures, materials used etc. Again, a connection to USTORE must be made by Octopus, to ensure that the failure incident details are stored.

As one may see from this example is Octopus supposed to follow work processes by connecting and integrating different sources of data. Octopus is planned to be operational with basic functionalities within 2013-2014. (itineris, 2011a, itineris, 2011b)

### 4.2.2 Comments and evaluation

It is no doubt that the Octopus concept can fulfil a very important operational function in VMW. Octopus aims to integrate dispersed data sources in order to better follow the information flow in day-to-day work processes, very much according to what is recommended by FHwA (2001) (section 2.4.3.4). The planned use of tablets or other portable computers with GIS viewer functionality, combined with a more streamlined information flow in work processes, is believed to stimulate higher levels of compliance to required standards for manually collected data (as postulated by Murphy (2009)), and might also increase the level of compliance with the USTORE requirements.

However, the Octopus project is still in the planning phase. The usability of data provided through Octopus depends on whether or not Octopus provides relevant data for asset life cycle monitoring, and whether or not it will be accepted and used properly by operational personnel. Some issues that need to be considered are:

- For leakages and other failures should Octopus be planned to follow the USTORE norm for data input[5] (described in section 4.1), and it is reasonable to assume that the USTORE data collection scheme will produce results that are usable for AM in VMW. If the data collection schemes are made similarly for the inspection and maintenance work processes in Octopus, they will also be likely to be usable. For instance, the *test protocol for fire hydrants* developed by Vreeburg et al. (2011) can be an inspiration for the inspection reporting from the fire hydrant inspection work process.
- The work processes that are planned to be modelled in Octopus are shown in Table 13. The number of inspection and maintenance work processes is quite limited, and maintenance that is not related to valves, fire hydrants or flushing, will have to be reported as an *adjustment* or as a *renewal* in the *project* block of Octopus. A consideration should be made on whether or not more inspection and maintenance work processes occur with such high frequency that they deserve to be modelled separately in Octopus, for instance if inspection and maintenance of flow gauges occur sufficiently often, these work process should be modelled separately. The same goes for pump stations, manholes etc.
- Another issue that should be considered when modelling data flow in the work processes is the possibility of embedding input data control and verification routines. When manually collected data is reported back through Octopus, the information flow could go to peers or supervisors who have to review and verify the data before it is permanently, serving the function of increasing the subjective norm of compliance from data collectors (Murphy, 2009) and to give indications of data quality (Wang et al., 1993).

A significant challenge with Octopus is whether or not it is possible to integrate it with all the other tools in VMW. If integration of data is not achieved fully, one will risk the emergence of quality problems such as duplication, and lack of consistency, clarity, completeness, availability and accessibility (FHwA, 2001).

All in all is the potential of Octopus as an integrator of data sources great. Octopus is made for operational purposes, but if designed according to the needs of tactical and strategic AM, the benefits will be much greater.

## 4.3 Interactive maps: Rolsch MapKit and Mobiele data

### 4.3.1 Description

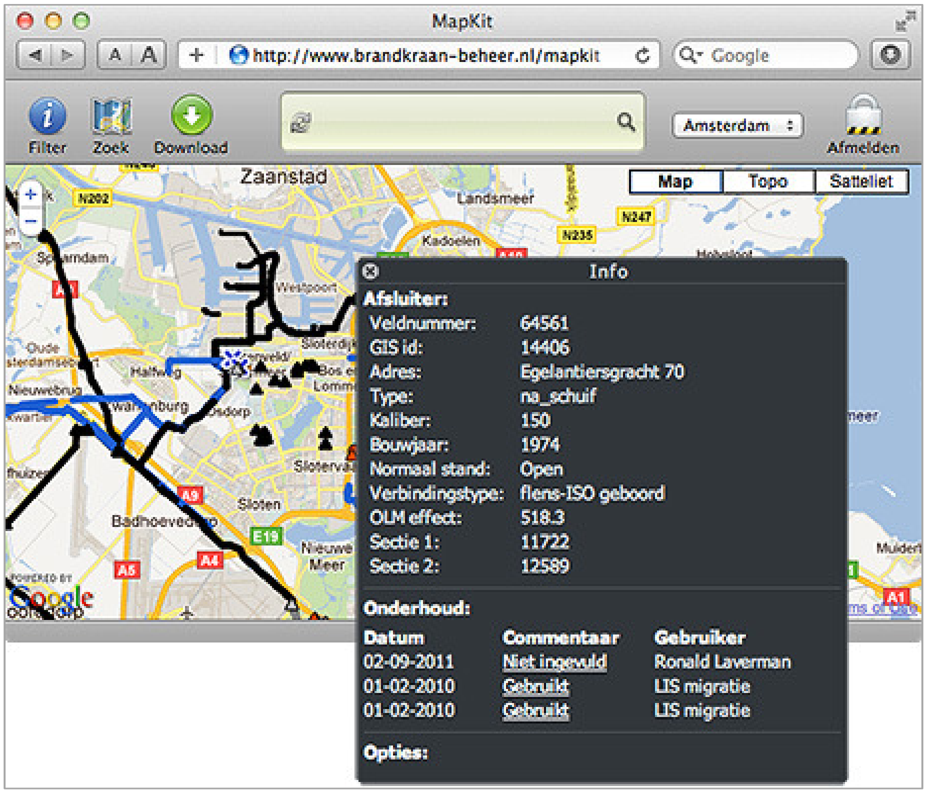
The Rolsch MapKit is used as an example of a browser-based computerised maintenance management system (CMMS) with an interactive map displaying capability; other similar solutions

---

[5] In addition, data about used resources (work hours, materials) will be reported

are also being evaluated by VMW. For reviews of CMMS applications with similar functionality, it is referred to Halfawy et al. (2006a).

Rolsch MapKit is a web-based GIS viewer, where information from the organisation's GIS database can be visualised. The web-based nature of the MapKit service makes it suitable for data exchange to and from portable computers during work in the field; Rolsch visualises the GIS-data on top of the Google Maps service, which makes spatial orientation an easier task. In addition to GIS visualisation, Rolsch MapKit allows the user to display and edit maintenance data through the web-application. (Rolsch AM, 2011b)



**Afsluiter:**

| | |
|---|---|
| Veldnummer: | 64561 |
| GIS id: | 14406 |
| Adres: | Egelantiersgracht 70 |
| Type: | na_schuif |
| Kaliber: | 150 |
| Bouwjaar: | 1974 |
| Normaal stand: | Open |
| Verbindingstype: | flens-ISO geboord |
| OLM effect: | 518.3 |
| Sectie 1: | 11722 |
| Sectie 2: | 12589 |

**Onderhoud:**

| Datum | Commentaar | Gebruiker |
|---|---|---|
| 02-09-2011 | Niet ingevuld | Ronald Laverman |
| 01-02-2010 | Gebruikt | LIS migratie |
| 01-02-2010 | Gebruikt | LIS migratie |

**Opties:**

Figure 7: Rolsch MapKit (copied from Rolsch AM (2011b))

Figure 7 shows an example screenshot from the Rolsch MapKit. By clicking on an item on the map, the user is able to display data about the item; in this example is basic data about a closing valve[6] (ID number, type, address, diameter, age etc.) as well as maintenance[7] data shown. MapKit is interactive and allows the user to display asset groups with (for instance) certain maintenance statuses ("not ok"), and localise critical areas on the map. If maintenance is executed on an asset, the personnel can update the maintenance data through the MapKit application window. (Rolsch AM, 2011a)

The use of interactive maps and GIS viewers can be used not only for maintenance personnel. For instance, Thames Water uses a map application to display current repairs, reported leaks, planned improvements and possible traffic disturbances. Anyone who has Internet access can visit their web-page, and use the Thames Water Live map to display what operational work is going on in their

---

[6] Closing valve is *afsluiter* in Dutch

[7] Maintenance is *onderhoud* in Dutch

neighbourhood – customers also have the possibility to report disturbances through the map interface. (Thames Water Utilities Limited, 2012)

Nevertheless, for a GIS viewer application to work, it needs data systems that support it. In order to follow up pending tasks through MapKit there needs to be structures that store and communicate the pending tasks to the mobile units, and likewise for reporting back results from completed tasks. The project *Mobiele data* is an internal VMW project that aims to develop a system for communicating to and from mobile units. Mobiele data is (in the long term) envisioned to be used for two-way communication of data that is relevant for the operative processes in the day-to-day operation of the water distribution network (inspections, incident reports, gauge reading etc.), through connections with the central data repositories. It is also planned that it will be possible to correct erroneous or missing data in GIS through Mobiele data. However, Mobiele data is a project that is still in the starting phase, and is only planned to have a few example functionalities demonstrated in the second half of 2012 as "proof of concept". (Danckaers and Hammenecker, 2012)

### 4.3.2 Comments and evaluation

It is no doubt that being able to easily visualise the spatial data of the assets on a portable computers makes the MapKit a useful tool for field work. Further, the ability to connect it to other spatial data sources (such as Google Maps) and to asset-specific maintenance data, makes it useful for issuing work orders and keeping track of interventions carried out on the network. If the system is designed suitably, the MapKit interface and the use of portable computers should increase the operators' perceived behavioural control, which in turn may increase the quality of the manually collected data from inspections, maintenance and repair. Also, the direct digitalisation makes it easy to control and verify new input data as it is produced, and quickly cleanse and turn data into useful informational outcomes.

Nevertheless, such a solution will only be as good as the data systems that support it. The success of the use of MapKit and other GIS viewer tools is dependent on the success of the *Mobiele data* project and indeed also the Octopus project. The success of GIS viewers is dependent on effective communication between the data repositories.

## 4.4 LCC-AM/QM – S&G and Partners

### 4.4.1 Description

VMW has decided to purchase an AM software tool from S&G and Partners – the LCC-AM/QM (Life Cycle Costing Asset Management / Quantitative Maintenance). The software is based on life cycle costing (LCC) considerations, which allow the user to produce cost analyses, evaluate economic replacement moments, and compare different maintenance strategies on a one-to-one asset level. The S&G and Partners software is a generic tool and is not tailor-made for water utilities. The intention of the software is to produce a representative long-term investment program for the organisation. (S&G and Partners, 2012a)

S&G and Partners (2009) describes the term *quantitative maintenance* as a combination to reliability centred maintenance and failure analysis. The software allows for analysis of maintenance over past periods, over passed life-times, prognosis of future maintenance cost, and analysis of effect of maintenance (S&G and Partners, 2009).

The input from an asset intervention can be stored as maintenance, failure, or defect, and has a list of input, such as:

- date of intervention
- object (asset) identification
- cost of intervention
- duration
- production loss
- class of failure/defect/maintenance
- cause of failure

The LCC-AM/QM bases itself on an independent database (Nexus); data used by the software is stored in the Nexus database, data can thus not directly be utilised from the GIS database (the same goes for USTORE and other databases). LCC-AM/QM uses a conversion tool for input data called the LCC-ANT (Asset Normalisation Tool), which is partly based on data mining approaches. Output data from LCC-AM/QM is also stored in the Nexus database, but can also be exported as e.g. text files, spread sheet files or graphics files (for graphs). Results can also be exported back to GIS. (S&G and Partners, 2012b, Hasenack, 2012)

### 4.4.2 Comments and evaluation

The LCC-AM/QM software allows the user to consider maintenance and repair scenarios over a long time horizon in the LCC perspective, which is a major step in the predictive AM direction. The successful implementation of the LCC-AM/QM software will hence be a milestone towards better economic control over the assets, and rationalisation of operational activities on the assets.

Nevertheless, there are some issues with software that should be mentioned:

- The fact that LCC-AM/QM works with a separate database counteracts the endeavour towards integrated data repositories. Even though S&G and Partners report that data assimilation for their software solutions can easily be achieved in most cases, the implementation of LCC-AM/QM will in any case result in duplicate storage of data – the data for the LCC-AM/QM in the Nexus database on one hand, and the “raw data repositories” on the other hand. This data duplication will increase the risk of data inconsistency (several versions), doubt about data timeliness and lack of data source credibility, and the inability to use data for multiple purposes (FHwA, 2001, Wang et al., 1993); data will also be less accessible for day-to-day planning, because it will be resource-draining to combine data sources – the data utilisation will hence be impeded especially on an operational level. It will also be more difficult to test and run other AM models “parallel” to LCC-AM/QM. If proprietary data storage solutions are established instead of a standardised central asset data repository, modelling and prediction capabilities will be limited by the proprietary solutions (Halfawy, 2008, Halfawy et al., 2006a, Halfawy et al., 2006b).
- S&G and Partners describe their quantitative maintenance analysis as a combination of reliability centred maintenance and failure analysis. This means that the *effect of maintenance (or repair)* is expressed through the *reliability* of the asset (and the costs associated to this); however, this is just one of the many possible dimensions one can evaluate the effect of maintenance of an asset; the effect can also be measured as hydraulic

functionality, number of complaints etc. LCC-AM/QM does only consider the LCC-reliability relationship, does not explicitly consider risk and condition, and does hence not comply with the common definitions of IAM (section 2.1).

- LCC-AM/QM has the possibility to export data back to GIS, thus enabling for example to combine failure rates with asset data in GIS to calculate expected risk with different maintenance scenarios, however if it is possible to return dynamic risk model results back to LCC-AM/QM is doubtful. If that is not possible one will not be able to balance risk, cost and performance, as in the AM definition of Alegre et al. (2006a). It will only be possible to model risk if risk is accurately expressed in monetary units.

From the limitations of LCC-AM/QM identified above, one may conclude that the LCC-AM/QM software never will be the *complete* solution for AM in VMW. The lack of data integration associated with this solution will then be a challenging issue. LCC-AM/QM is nevertheless a tool that will provide the capability of considering life-cycle cost evaluations based on maintenance scenarios, provided that the necessary data will be collected. The tools is expected to be most useful on tactical level.

## 4.5 InfoWorks

### 4.5.1 Description

VMW uses the InfoWorks WS modelling tool from Innovyze (Innovyze, 2012b) for hydraulic modelling. InfoWorks WS builds models by a direct link to GIS, where GIS data is converted and stored in InfoWorks' relational database (Innovyze, 2012d). InfoWorks WS does not use the GIS database as a data repository for building models, like its sister application InfoWater (Innovyze, 2012a). Instead, InfoWorks has a separate database where several versions and configurations of the model can be stored. Model versions are tagged with timestamps, version ID's etc. (Innovyze, 2012d)

The modelling features in the standard InfoWater software are similar to the ones known from non-commercial models such as EPANET (USEPA, 2012), except for some small adaptions necessary for a GIS link. Modellers can simulate dynamic scenarios, produce tables, graphs and maps with results on water flow, pressure, velocity, water quality etc. Data can also be returned to GIS.

InfoWorks WS also contains tools for automatic model building, such as tools for automatic pipe characteristic lookup, elevation assignation, demand allocation and model merging. Further, InfoWorks provides capability of establishing real-time links with telemetry and logger data systems, enabling the model to be simulated based on real-time calibrations, with the influence of data from data from pressure gauges, flow gauges, pump stations and similar logging devices. (Innovyze, 2012d).

Other features in InfoWorks WS are fire flow modelling, sediment transport modelling, critical link analysis (analysis of failure consequence) and modelling of effect of flushing (Innovyze, 2012c).

VMW uses InfoWorks for day-to-day operational management of their distribution networks, by monitoring the system through gauges and modelling, and controlling the network with live gauges and actuators. InfoWorks is also used for simulating hydraulic impacts of renewal and rehabilitation scenarios.

### 4.5.2 Comments and evaluation

The fact that VMW combines real-time data in their hydraulic models shows that they are capable of combining different data sources (asset GIS data and operational real-time data) to produce information that is useful managing the daily operation of the distribution network.

As mentioned, InfoWorks WS does not use the GIS data repository directly, but uses a link to GIS to construct hydraulic models in a separate relational database. As with LCC-AM/QM, InfoWorks also contributes to a more dispersed data structure in VMW; however, InfoWorks is already operative in VMW, and has proved to be able to exchange data effectively, and data in the InfoWorks database can be tagged with indication of data source and confidence levels (Innovyze, 2012c). One positive aspect about separating GIS and model data is that it emphasises that a model is in fact not real, and lets analysers build model scenarios outside the sphere of the *physical asset* registry.

VMW uses InfoWorks to simulate hydraulic effects of changes when individual projects are being evaluated. However, a hydraulic model can be used much more extensively in the AM context. Combining hydraulic model result with probability (and cost) of failure given certain management scenarios, can form the basis for a risk- or performance-based AM strategy (section 2.1.4). This does however require input from other tools, such as the LCC-AM/QM or other deterioration or failure forecasting tools. Increased value of having a well-established hydraulic model can thus be achieved by utilising it more in a strategic and tactical context. Still, this requires that data gathered through systems such as Octopus and USTORE are easily transformed to usable outcomes, through LCC-AM/QM and other modelling tools. Ensuring effective communication of data between the different data repositories, whilst also ensuring data quality management under the disparate storage structure that is imposed by the selection of tools that is foreseen in VMW, will certainly constitute a challenge.

## 4.6 Conclusions: all tools combined

**Table 14: Main purposes of the asset management tools in VMW**

| Tool | Purpose | Status |
|---|---|---|
| GIS | Storage of spatial data and important attributes about the data | Existing |
| USTORE | Storage of failure data; production of failure statistics | Off the shelf |
| Octopus | Coordination of data for day-to-day operational work processes; data integrator | Under development |
| MapKit and Mobiele data | Assisting retrieval of data that is useful for operational tasks; assisting reporting of results from operational work | MapKit: Off the shelf<br>Mobiele data: Under development |
| LCC-AM/QM | Investment planning | Off the shelf |
| InfoWorks | Hydraulic simulation; day-to-day scenario modelling | Existing |

The purposes of the tool portfolio envisioned in VMW are summarised in Table 14. Even though it is not in the initial intention of Octopus, it will undoubtedly act as a very important integrator of data for all the other tools that are producing or consuming data. The integral role of Octopus raises the issue of timing. Octopus and Mobiele data are internal projects in VMW that are under development, with a time horizon for completion of several years, the GIS database and InfoWorks models are existing, and USTORE, LCC-AM/QM and MapKit are solutions that are commercially available and can be implemented quickly by the support service from the software providers. LCC-AM/QM will hence be implemented before Octopus and Mobiele data are ready – there is thus a gap between the tools that are providing data and the tools that are consuming data, which will inhibit the flow of data to the data consuming tools, which again will inhibit the production of informational outcomes from the data. On the other side, the implementation of data consuming tools will create greater incentives for the development of the tools and structures for data collection needed for optimal usability of the informational outcomes.

The suitability of these tools in a predictive AM context will be evaluated in the following chapter.

# 5 Evaluation of VMW's technical factors for data collection

The previous chapters have been leading up to the evaluation of VMW's position with respect to data collection for asset management (AM):

- Chapter 2 was a theory review, with focus on data and data quality for AM
- In chapter 3 was the life cycle of a buried asset broken down into representative intervention events, and analysed with respect to how data collection from these interventions could help providing the information necessary to answer the "six whats" of AM (Vanier, 2001)
- In chapter 4 VMW's planned "toolbox for AM" was described and commented

In this chapter these findings will be combined: The collective capability of the tools (described in section 4) to oblige the data needs (described in chapter 2 and 3) will be evaluated. This forms the basis for a "diagnosis" of the data situation, and results in suggestions for improvements.

First though, it is suitable to describe the current situation in VMW in general terms. The only tools that are currently functional in VMW are the InfoWorks WS and GIS repository.

## 5.1 Today's level of data collection in VMW

Vanier (2001) state that the experiences show that most water utilities fare well with the two first of the "six whats" of AM, but have not answered the four remaining; meaning that they have control of what they own and what it is worth, but not what the deferred maintenance, condition, remaining service life and priorities on renewal are. This characterisation also applies for the current situation in VMW. The VMW inventory contains a more or less complete dataset of their buried assets, with basic characteristics (materials, dimensions, where they are, how old they are), much data about external factors is however missing (Rokstad, 2011a). As any other company, VMW is obliged by law to keep track of the financial value of their asset portfolio – today the buried assets are valued purely on depreciation of the age of the asset and the expected service life of the asset type and material (VMW, 2010). As shown in the previous project work (Rokstad, 2011a), VMW does not have any central registry of maintenance and condition of their assets; and is hence neither in a position to estimate true conditional value of their assets, remaining service life nor prioritisation of renewal projects.

Remembering Figure 4 (page 25) showed different levels of data collection to monitor the life cycle of an asset – the current situation in VMW is very much similar to the line representing no data collection.

It may seem that the most advanced tool for AM currently used in VMW is the InfoWorks WS software. The fact that the hydraulic model is fed by real-time data from the network, allows VMW to continuously monitor the hydraulic performance of the distribution network. However, the hydraulic model can also be used to a much greater extent.

## 5.2 Evaluation of the planned AM toolbox

In Figure 5 (page 34) the data that could aid the monitoring of an asset's state through its life cycle were categorised into different categories. The tools described in chapter 4 will be put into the context of these data categories here, in order to evaluate how capable the (planned) tools are to collect life cycle data to answer the "six whats" of AM. Figure 8 shows the transition from raw data to informational outcomes. For simplicity, external GIS data sources are included in the "Plans/drawings" box. *Mobiele data* is not displayed, since it is only an agent connecting a GIS-viewer to the central system. Also, the boxes marked *Octopus*, will actually comprise of several sub-units that are connected through Octopus – it is thus less confusing to simplify by writing simply *Octopus*. LIMS is the laboratory information management system, and ARCADO is the invoicing system.

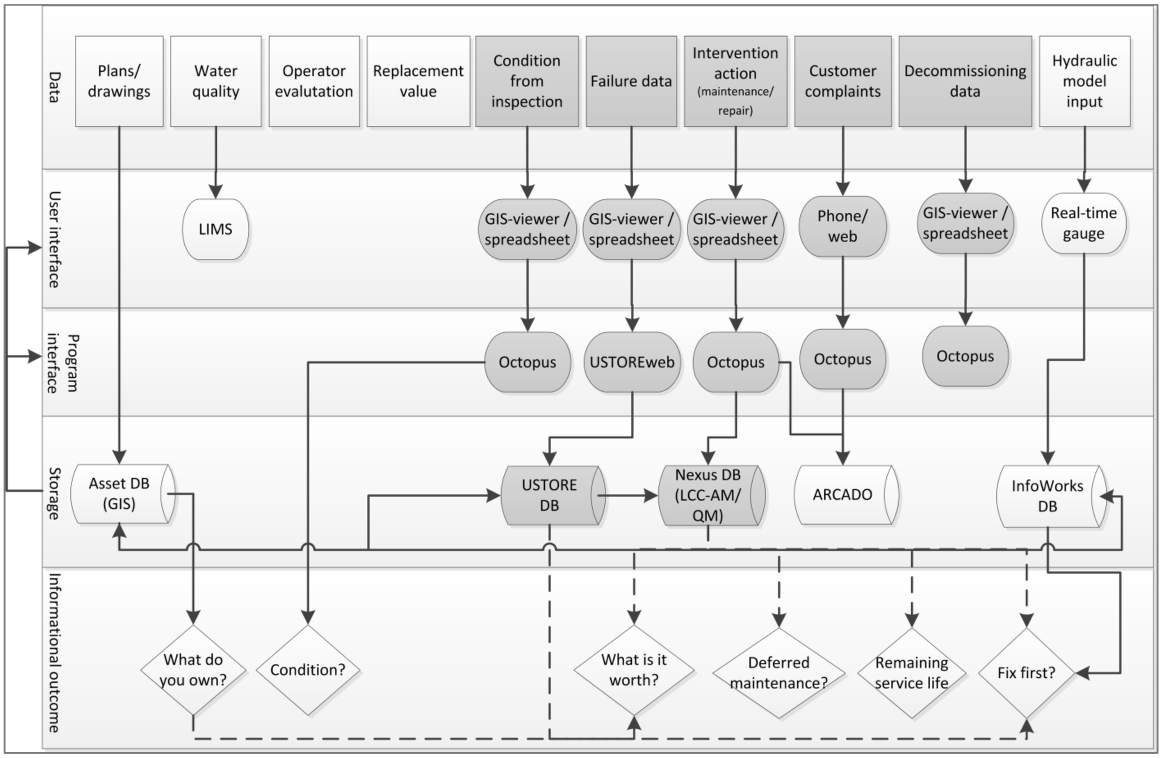


**Figure 8: From raw data to information (white boxes are existing systems, grey are planned)**

### 5.2.1 Performance with respect to the "six whats" of AM

Now, from Figure 8 one may discuss how well the planned AM tools in VMW can aid the utility to answer the "six whats" of AM:

I. **What you own** is a question that is already *well answered* through the existing GIS inventory. The fact that VMW does much of the design of new projects in-house makes it easier to transfer data from a project planning stage to the asset registry after commissioning. Data about external conditions are in many cases scarce.

II. **What it is worth** is a question which is supposed to be addressed through LCC-AM/QM, based on the input from the inventor and maintenance records. However, as previously indicated, LCC-AM/QM only considers the performance of a function of the reliability, and not by any other measure. The question of asset value is only *partly answered.*

III. Calculation of the **deferred maintenance** may also be performed in the LCC-AM/QM. Again, LCC-AM/QM will only consider the interaction between reliability and maintenance, and will not consider the interaction between condition and maintenance (as suggested by Kong and Frangopol (2004)). This question is also only *partly answered.*

IV. **Condition assessment** and inspection result conveyance is something that is suitable for modelling in Octopus; currently inspection of valves and hydrants are planned, however this can be expanded. This question is expected to be *well answered*, however with the configuration displayed in Figure 8, the condition is not utilised to help answer the other questions or to be used in models.

V. **The remaining service life** is addressed in LCC-AM/QM. This is done through the quantitative maintenance calculations in the software, and does hence answer the question of the economic service life *well*. Again, the service life will still only be dependent on the reliability.

VI. **What to fix first**: The quantitative maintenance calculations in LCC-AM/QM will be able to support resource-prioritisation decisions, based on LCC calculations, and identify if maintenance or capital renewal is economically optimal. Yet, LCC-AM/QM does not consider the interdependencies of water infrastructure assets and the variation in criticality (hydraulic importance) – if VMW wants to achieve a risk-based planning, the results from LCC-AM/QM must be combined with output from hydraulic simulations in InfoWorks. Neither here is condition considered. This question is also *partly answered*.

Remembering Figure 4 (page 25) again, one may see that VMW will *collect data* about both maintenance, failures and inspections to LCC-AM/QM, but that the *effect* of these will only be assessed according to the “failure data line” (reliability).

### 5.2.2 Other issues

Further, there are some issues about Figure 8 that need to be addressed:

Firstly, not all the data is converted into informational outcomes. Data about water quality deviances and operator evaluations (perceptions) of intervention situations are not encompassed in the scope of any of the tools described in section 4. Available data will hence not be utilised optimally (Grigg, 2006). It is planned that customer complaints and some[8] asset decommissioning operations are to be modelled in Octopus; however where the data from these interventions will be stored is uncertain – they do not have a clear path to become informational outcomes.

Secondly, the storage structure that appears in Figure 8 is somewhat untidy. Some operational data goes to USTORE, some is stored in the LCC-AM/QM Nexus database, some is stored in the InfoWorks database, and the remaining data (such as condition data) does not have a planned location. The configuration with proprietary databases for each application, makes it difficult to utilise data outside these applications (Halfawy and Figueroa, 2006), which again makes it difficult to apply other tools and models to help mitigate the shortcomings the current tool combinations have. If software is replaced in the future, the transition of data may be challenging with a proprietary data structure. Also, it would be easier to extract desired performance indicators (PI’s) from combined sources and use data in day-to-day operations from an open database – the same goes for documentation of data quality.

---

[8] Decommissioning of customer connections will be modelled in Octopus

Thirdly, from Figure 8 one sees the integral role of Octopus as a program interface – working as a bridge between raw data submission and data storage. The feed of data back to the program interfaces is illustrated on the left side of Figure 8. The success of Octopus (and applications using Octopus) is depending on whether or not the feed of data back from the data repositories is possible – if not, an operator will not be able to get an overview over an asset's attributes and history in the application that he/she is using. With the dispersed nature of the data repositories, the return of data to Octopus may prove to be challenging.

Lastly, timing is an important issue. The integral role of Octopus has been identified – Octopus is the most important step for achieving the data collection procedures that are desired in VMW. However, Octopus is a long-term development project, and it may be years before the first modules are operational. The data consuming tools, on the other hand, need data as soon as possible if they are expected to produce informational outcomes.

### 5.2.3 Conclusions about planned tools

Based on the identification of possible data sources in section 3 and of the tools that are supposed to convey these in section 4, the evaluation of the data collection scheme has been possible. Some conclusions are:

1. The proprietary and dispersed nature of the data repository structure creates a barrier for implementing additional AM tools and using data more effectively. Concern is also raised about the feasibility of returning data to the interfaces that follows work processes under this storage structure. Documentation of data quality may also prove challenging in a dispersed storage structure.
2. The development speed of the Octopus project will be expected to be limiting the production of informational outcomes.
3. Several of the "six whats" of AM are only partly answered, due to the fact that LCC-AM/QM does not model the relationship between maintenance cost and condition or risk, but only model the relationship between maintenance cost and reliability.
4. Not all data that is planned to be collected has a direct link to desirable informational outcomes. A special concern is raised for the lacking integration of condition and inspection data into other informational outcomes. Measures can be employed to ensure optimal use of already existing sources of information (Grigg, 2006).

## 5.3 Suggestions for improvement

Here, some suggestions will be made in order to improve the situation for the first problem discovered in section 5.2. The second problem will be treated in chapter 6, whereas additional tools to solve the two last problems will be discussed in chapter 8.

### 5.3.1 An open data repository structure

The first and largest problem identified in section 5.2 can be solved by designing a more open data structure, by making one non-proprietary central database (parallel to the GIS data repository), where all life cycle data about an asset can be stored, including inspection results, failures, maintenance and repair work details etc. – the database will work like a **"diary"** for each asset. From this database, data can be distributed to LCC-AM/QM, USTORE or any other software or model. Data about pending or planned operations, based on the outcome from up-stream analyses, should also be possible to store in the diary database (through Octopus).

Figure 9 shows how the asset GIS database and the diary database are on equal footing, and how these two main data repositories feed the program- or model specific (or proprietary) sub-repositories, which again produce informational outcomes. Since all operational data is stored in one open database, it may be extracted to any application. It should also be possible to return informational outcomes to the diary database. Informational outcomes do further act as a support for decision-making processes, which on an operational level can result in pending work orders that can be handled through Octopus. (ARCADO is the invoicing system in VMW.)

Figure 9 indeed sums up the concept of data driven AM, from the objective of the water utility (to provide water to its customers), to the recording of day-to-day operational work data, which form the basis for analysis, strategic planning, and tactical decision-making. Tactical decision-making is again realised as projects, and executed at the operational level, and so the circle is closed.

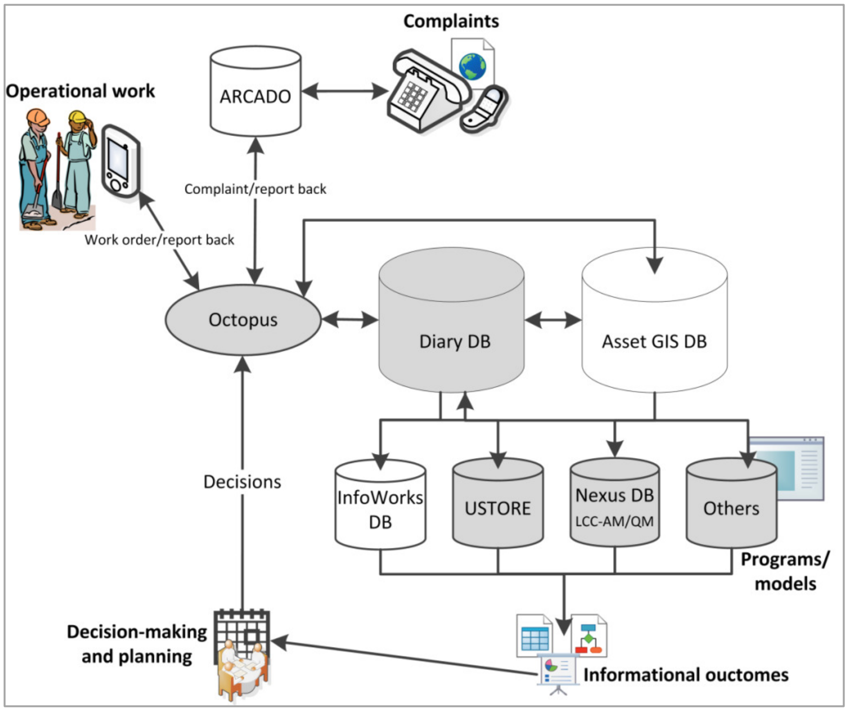


Figure 9: Conceptual diagram of the potential data-information flow

A revised version of the illustration of data flow, from different kinds of raw data to informational outcomes, is shown in Figure 10. The main diary database works as a distributor to the applications that have their own data repositories (USTORE, LCC-AM/QM, InfoWorks). In addition, if VMW is to succeed in distributing real-time information about supply disturbances and repairs through Octopus, the invoicing system (ARCADO) must be connected to the both the asset GIS and the diary database, where each customer (connection) is connected to an asset (pipe) identification number. The use of water quality sample information will require a link to the current laboratory information system (LIMS), so that spatial information (GIS) can be linked to the water quality samples that are registered.

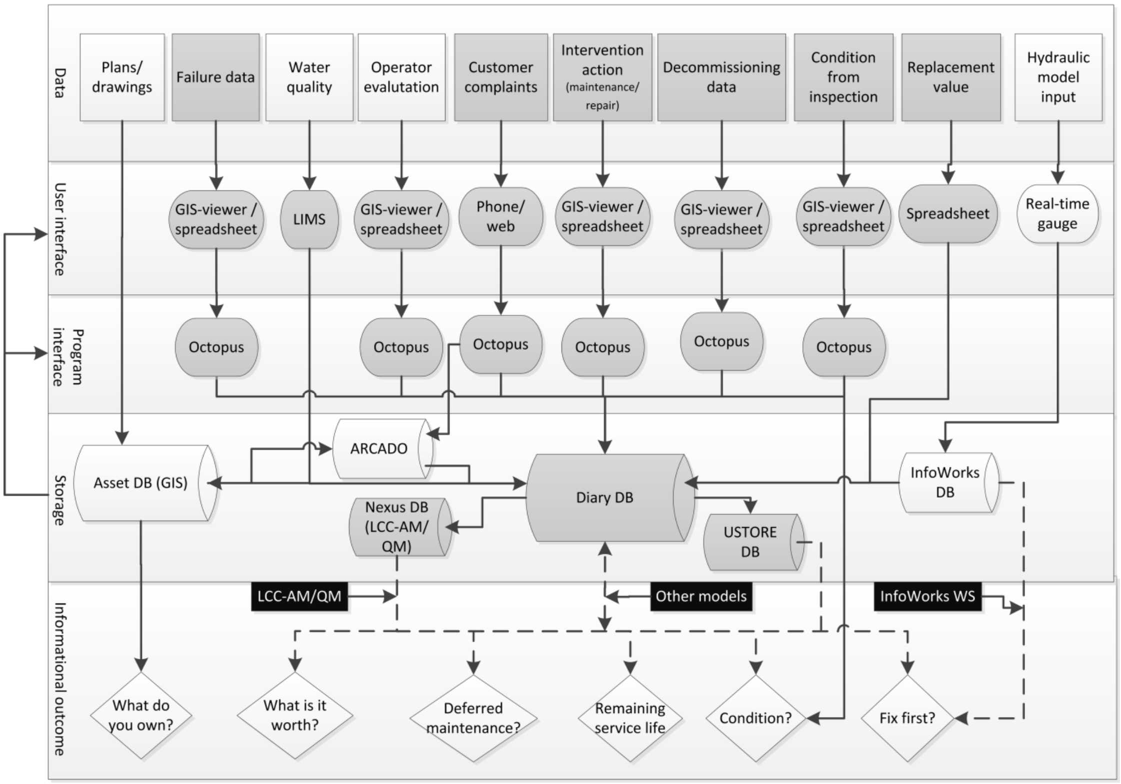


Figure 10: Revised flow scheme of raw data flow to informational outcomes

### 5.3.2 Diary database specifications

The diary database that was suggested in section 5.3.1 has so far only been described roughly according to what functions one could imagine that it should fulfil. Here, the diary database will be specified more concretely. The diary database will be specified according to the findings in the theory review and the evaluation of VMW's AM tools. What will be described is: (1) the purpose, (2) the contents, (3) data quality issues and (4) the development of the database.

#### *5.3.2.1 Database purpose and overall requirements*

The purpose of the database should be to act as a central repository for all non-static data that is relevant for monitoring and describing the asset life cycle, in terms of performance, value, cost of ownership, criticality, risk, condition, remaining service life, level of service etc. The database will act as a complete "diary" of events for each asset, and will be linked to assets in the GIS inventory by the asset identification number. It should also be possible to store pending work orders in the database, in order to aid coordination of day-to-day operational work.

The database should be designed in such a way that:

I. It is able to export data to all AM tools employed in VMW
II. It is able to import and store output from these tools
III. It is possible to enter data from an array of different sources
IV. It is possible to link file attachments to the database (sketches, photographs etc.)
V. Manually collected data can be used for statistical analysis (selection list values)
VI. It is possible to access, use and review the data at the local service centres
VII. It is possible to document and monitor data quality

#### *5.3.2.2 Database contents*

The contents of the database are governed by the purpose of the database: If the database shall act as a diary of events and evaluations of an asset, it must above all be able to store information about all events and evaluations along the life cycle of the asset. This means that it has to be structured in such a way that it can store data about normal operational parameters, inspections, failures and repairs, planned maintenance work, and decommissioning. If a complaint can be linked to a specific asset or group of assets or location this should also be possible to store.

For *normal operation* of a water distribution system, the sources of data that were identified as supporting for AM were water quality samples, customer complaints, and hydraulic modelling and gauging. There should be room for these data in the operational database:

- Water quality samples should be spatio-temporally fixed, either by relating the quality sample to the tapping point address or the connection point, and should contain basic water quality data such as bacteriological quality indicators, turbidity, colour etc. This could give the water quality data added value in the long term.
- Customer complaints that do not lead to any rectifying action (repair, maintenance) should also be filed as complaints, with time of complaint, nature of problem, and identity of the customer.
- Results from InfoWorks that reflect the criticality or importance of an asset should be assignable to all network links.

*Inspection* results should be stored with a classification of inspection method, result and recommended course of action – storage options should be adapted to inspection type. Cost and duration of an inspection event is also relevant, since these parameters are input for LCC-AM-QM.

The main recipient for *maintenance data* will be LCC-AM/QM. Maintenance data must therefore be compatible with, and contain all data necessary for, LCC-AM/QM. Important input for maintenance actions in LCC-AM/QM is date, duration, type, class (inspective, corrective, preventive, adaptive or other) and cost of maintenance.

For *failures and repairs* must the data be compatible with USTORE. Failures stored in the diary database should contain all the data that is delivered to USTORE (section 4.1). However, the input in USTORE is not sufficient for assessment of cost of the failure rectification (repair). Therefore, additional data reflecting the resources invested to repair the failure should also be stored, since this data is required by LCC-AM/QM. It might also be useful to store data describing the severity of the failure (mode of discovery, number of people affected, duration of supply disruption), and whether or not the incident should be followed up by an inspection. Some of these data (such as the cost, production loss etc.) are input data for LCC-AM-QM, and should therefore be stored in a compatible format.

For the final stage of the life cycle, the *decommissioning*, the most important data to record has been identified to be *reason for decommissioning* (can be several) and estimation of the economic conditions of the asset; such data may be dependent on modelling results, and decommissioning entries should therefore be connected to modelling results (see following paragraph). Some decommissioning data are already stored in the inventory, and it is desired by VMW to keep that data structure to ensure completeness of the inventory – additional data should however be stored in the diary.

In addition to recording raw data that emerge from events during the life cycle of an asset, the role of modelled parameters became apparent in the analysis of a buried asset's life cycle (section 3.2). There should also be room for *modelled and estimated parameters* in the diary database, because the modelled parameters will have more value if they are accessible and used in day-to-day operations. InfoWorks results have already been mentioned, but condition, remaining service life, reliability, performance, and economic value are also parameters that can be modelled and used for multiple management purposes. The same goes for estimated costs (of condition/reliability increase, failures, replacement). For modelled results, it is important to classify which parameter is modelled (condition, remaining service life etc.), which method is used to produce the result, and at what time the result applies and in what scenario it is modelled in (relevant for predictive/forecasting models).

Both pending and completed operational work orders should be possible to store in the database.

It is also stressed that the diary database should be able to document the interdependencies between different events; it is suggested that it should be possible to state the reason for an intervention action and, if possible, link an intervention to another event – for instance could a maintenance event be a consequence of an inspection event. If such interdependencies can be analysed on the basis of the data in the diary, it will help the utility to distinguish which elements lead to decisions, and which do not. For instance could inspection techniques that do not lead to a decision, and are hence inefficient, be identified.

The lists of suggested input fields for each event type are enclosed in Appendix A.

### 5.3.2.3 Ensuring data quality

Data quality was an issue that was raised in the theory review (section 2.3), and is something that should also be taken into consideration when defining the data storage structures in an organisation (Wang et al., 1993). It is proposed to implement two structural elements from the data quality modelling theory in the diary database: metadata on *verification* and *quality tagging*.

**Table 15: Verification metadata, as suggested by Wang et al. (1993)**

| **Verification** | |
|---|---|
| **Verification result** | Approved / not approved |
| **Who verified?** | Employee ID |
| **When was it verified?** | Date |
| **How was it verified?** | Method selection list |

**Verification metadata** is a set of metadata that document that an entry in the database has been verified. It is suggested that entries with manually collected data in the diary database should have verification routines, wherein it is possible to document that data has been verified, the time of verification, result of verification, and how and by whom it was verified (see Table 15). VMW should develop procedures for verification responsibilities, i.e. who is responsible for verification of data in different situations (peers or superiors). The verification of data should appear as a pending assignment (work order) in Octopus.

The use of **quality tags** (Wang et al., 1993, Wang et al., 1995a) will help data consumers to evaluate the quality of data. If data is filled in on computers, it is simple to automatically generate quality tags for certain quality attributes; *timeliness of data* and *data source* tags can easily be generated by the computer system when a data producer is using his or her account on a browser-based system to fill in data. Quality attributes can also be generated automatically for modelling results. For certain data, such as inspection results, the quality attributes may be inherent from other input data (e.g. the inspection method). It is suggested that automatic generation of source and time of registration is implemented in all entries where that is possible, and that it should be possible to store model result uncertainty measures, for instance relative uncertainty. It is also possible to give the personnel that are collecting data manually the possibility to declare their perceived level of uncertainty on certain input values. However, this should be limited to only a few critical input fields, to avoid an excessive work load.

Table 16: Suggested quality attributes

| Quality attributes | |
|---|---|
| **Timeliness** | Date of registration |
| **Credibility** | Input source ID<br>Recording method |
| **Accuracy** | Recording method<br>Perceived accuracy |
| **Completeness** | Percentage of fields in entry filled in (schema completeness)<br>Percentage of fields in column filled in[9]<br>Percentage of possible values present (population completeness) |
| **Consistency** | Variation over time<br>Variation over input methods/sources<br>Analytical redundancy control and extreme value detection<br>Number of new entries since last export |

Some quality tags, such as completeness and consistency attributes, will not be feasible to model on an individual data entry level. These quality tags may however be implemented on a meta-level. Consistency of data will be an important issue for VMW, due to the dispersed nature of the VMW organisation and the spread storage structure. There are at least three kinds of consistency to consider:

- *Consistency of raw data.* Variations in data over time, methods and personnel can occur. Monitoring changes over time and variations between methods and the data sources will be instrumental to quantify the consistency of raw data. (If one local service centre differs very much from the others, it is a reason for investigation.)
- *Consistency with expected values.* The comparison of reported values with modelled, estimated, or tolerance band (minimum/maximum) values, can in some cases be used to automatically detect erroneous data (Sun et al., 2011). Automatic control and reporting routines of manual input should be established. For instance, if a manual input showed that

---

[9] This completeness measure describes the completeness of one specific entry field type among all observations, for instance the field "*number of work hours spent*".

500 work hours was spent on an activity that usually only requires 10 hours, this value should immediately be highlighted to the data quality responsible.

- *Consistency between data repositories.* Consistency of data that has been transferred to program-specific repositories can be related to the timeliness of data export and the number of new data entries since the last data export. If many entries have been produced since last data export, the consistency is no longer intact.

Additional technical measures to ensure data quality may include: (1) development of clear definitions of input fields, and (2) the establishment of data quality indicators based on the quality attributes. These measures will be discussed more in chapter 6.

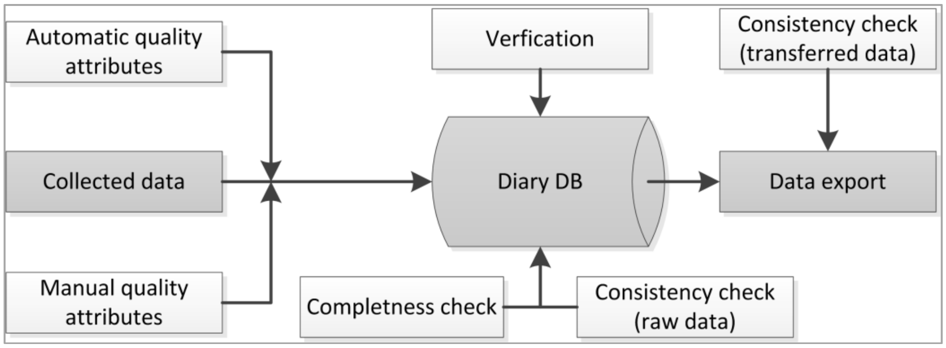


Figure 11: Technical quality assurance system for the diary database

The technical measures in the quality assurance system are summarised in Figure 11. The use of these technical measures in the organisational context will be discussed in chapter 6.

### *5.3.2.4 Development of the database*

Through the analysis carried out through the previous chapters, it is evident that Octopus will play an integral role in almost all data collection procedures in VMW. It is therefore advisable to develop the diary database parallel with the development of the Octopus modules. The need for data storage emerging from Octopus will in that way make the foundation for the diary; the diary will then be developed according to the needs of Octopus, and hence ensure that there is a compatible repository maintaining the data “below the surface” of the user interfaces. The parallel development of a diary will also ensure that it is adapted to these work processes, since Octopus is aimed at modelling work processes. However, the scope of Octopus is only to ensure collection of “data from field work”, and the diary should not be constricted by that scope (water quality samples will for instance not be in the scope of Octopus). The suggestions made here may therefore be perceived as an extension of the Octopus idea.

## 5.4 Conclusions

By the use of the “six whats” of AM, the capability of VMW’s tools to absorb the data needed for AM has been evaluated. Although not perfect, this method casts the light on what data is necessary to produce useful information for AM. It was found that the planned tools in VMW will improve the data situation extensively, but that it will be difficult to ensure that the potential of the data will be fully utilised, unless measures are taken to ensure a central storage of life cycle data. The dispersed storage structure was identified as the most predominant impeding factor. In addition was the lack of criticality- and risk-based management in LCC-AM/QM identified as problematic, and additional tools for assessing hydraulic criticality should be evaluated.

In order to ensure the possibility to use data for several purposes, it was suggested that all data that is instrumental for the monitoring of the life cycle health of an asset should be stored in one open diary database, which other applications could use as a “data well” for whatever purposes they have. A suggestion for the contents of this database is enclosed in Appendix A.

It was also suggested that this diary should include certain metadata for data quality and data verification. Verification of all manually collected data is suggested, and metadata on data source, time of entry, and recording method is suggested for all data. Routines for consistency control are also suggested.

The implementation of a central diary database, instead of distributing data in dispersed repositories, is expected to give VMW greater versatility and comprehensiveness to produce all the perspectives of information that are necessary to make informed decisions about the management of their assets.

# 6 Organisational factors for data collection

In the previous chapter, technical measures to improve data collection were suggested. However, the technical premises are only one foundation for achieving an acceptable data collection situation. Successful collection of data is also dependent on the data collectors' personal circumstances and the organisational context the data is collected in. Important factors that affect the data collectors' level of compliance with the requirements for data collection will be reviewed here and measures to facilitate data collection will be suggested. The question to be answered in this chapter is: How does one ensure that the people who work in VMW will use and collect data for the tools that have been chosen? In chapter 5 it was also found that the implementation timing of the different tools in VMW is unfortunate for the data collection-utilisation relationship – this is something that will be improved by the measures suggested in this chapter.

First, some organisational prerequisites for data collection will be identified, and then the organisational side of VMW will be evaluated. Lastly, suggestions for improvement will be made.

## 6.1 Prerequisites for data quality with manually collected data

Organisational context prerequisites have been classified in Figure 12, based on what was uncovered in the theory review (section 2.4.3). *Data quality planning* is necessary to ensure that goals for data quality are defined, and responsibilities for the tasks necessary to achieve these goals are distributed in the organisation. Then, the conditions for the data collector must be present so that the *data collector* intends to comply with the quality targets (theory of planned behaviour). Finally, after the data has been collected, there must be systems and routines for monitoring and encouraging improvement in the data quality.

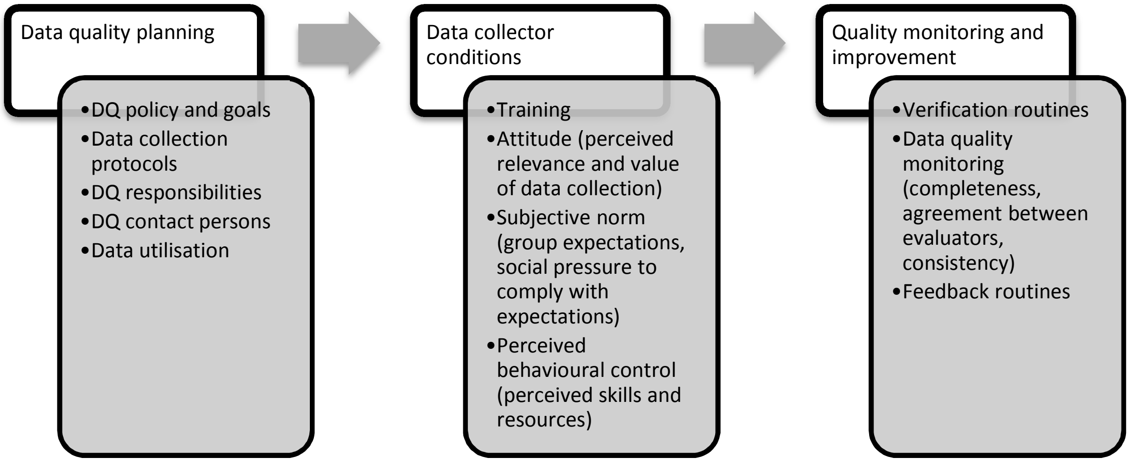


**Figure 12: Prerequisites for data quality**

From Figure 12 one may see that the premises for data quality follow a pattern with *planning, executing and monitoring* – suggestions made in this chapter will be based on these three pillars.

## 6.2 Organisation description and evaluation

### 6.2.1 Organisation description

VMW has a head office in Brussels and one provincial office for each of the four Flemish provinces they are active in – strategic and tactical planning occur both at the head office and in the region offices. In each province there are several sectorial service centres (18 in total); it is from these service centres that the day-to-day operational work like maintenance, repair and customer responses are handled. (VMW, 2012b)

At the head office in Brussels, there are seven different departments (*directies*) (VMW, 2012b), see Figure 13. All these departments will have different functions to fulfil, agendas, expectations, perspectives, and competence that might affect the organisational environment in which data is collected. The technical department will for instance have most interest in (and knowledge about) the technical aspect of asset interventions, and may want to utilise data to investigate which repair and maintenance methods are most effective with respect to technical service life, whereas the complaint services department may want a data collection system that most effectively conveys complaints to the right decision and feed-back to the customer.

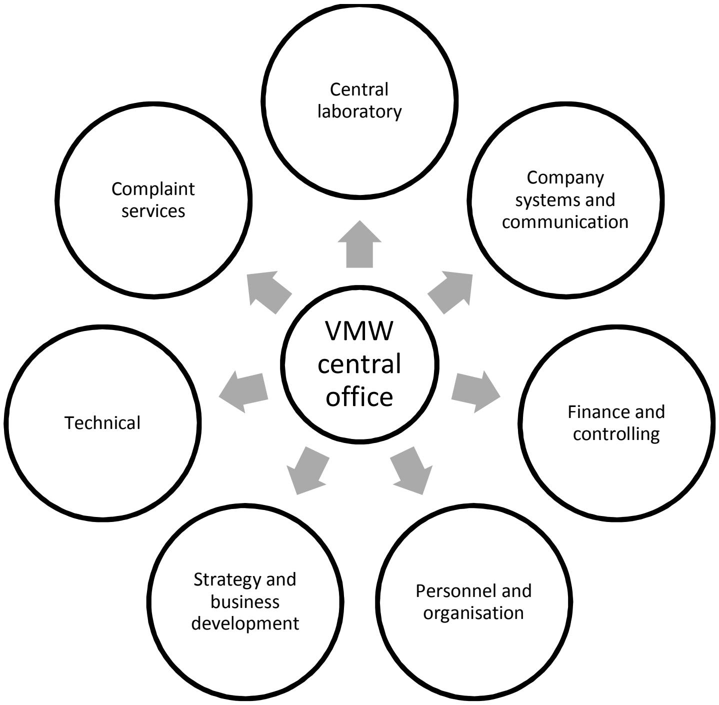


**Figure 13: Departments at central level in VMW**

At the provincial offices and local service centres, the objectives may diverge from the ones defined at the central office. It is fair to assume that operational personnel in VMW (at local and provincial level) desire a predictable working environment, with sufficient behavioural control (i.e. not too complicated and exhaustive tasks), and tasks that are perceived as relevant for them. It is also fair to assume that if new procedures are perceived to threaten such an environment, they will be met with resistance.

Today, data collection about operational work in VMW is mainly done through paper forms, where repairs are most completely reported. Personnel working in the field make use of large paper-based maps of the water main network. Simple repairs that only require a repair sleeve are reported only through a simple form; if a repair is more complicated, the personnel in the field make a sketch where they draw the main features of the repair work. For repairs that are complicated enough for a sketch to be made, are the sketches sent to the drawing office in the GIS department of VMW, where they are later registered as “deviant pipe segments”. These segments are not referenced to the identification number of the asset that has actually been repaired. (Danckaers, 2012)

However, most of the simple repairs never leave the paper format, and are only stored in paper ring binders. The same is true for maintenance and inspection operations. At best, this data is available in electronic databases, but then without any reference to the identification number of the asset they concern. This situation is not optimal with respect to asset management (AM), because it is not possible to easily keep track of what has occurred with each specific asset in the inventory. The transformation of raw data into informational outcomes can thus not be conducted without high levels of uncertainty and interpretative ambiguity. (Rokstad, 2011a)

### 6.2.2 Comments and evaluation

It is clear that systematic collection of *life cycle data* for the purpose of AM is something new in VMW and that the implementation of the new AM tools will create data demands that have not existed before. This will most likely prove to be challenging for VMW, because data collection has not been a central part of operational personnel’s job descriptions – VMW is still working with a reactive AM strategy, where data is not considered as important. Technical solutions can be designed perfectly, but if the personnel that are responsible for collecting or producing data are not convinced of its importance, the organisation will suffer from insufficient and low quality data.

The lack of systematic data collection in the past is a drawback in two mutually dependent ways:

1. The very strong left censoring of data (lack of historical data before the initiation of systematic data collection) will impede the quality of the informational outcomes of the AM tools until sufficient data is collected
2. The lack of credible informational outcomes from the tools that convey the data will reduce the personnel’s motivation for collecting more data.

The emergence of new tools for AM, and tools for data collection, such as GIS viewers and interactive browser-forms for portable computers, will demand a change in the way personnel in VMW works with planning, executing and completing tasks. This is a process which requires learning and establishment of intentions and attitudes about the importance of data collection.

The fact that VMW is a utility with a low level of outsourcing may be used as an advantage in the data quality perspective, because the data that is collected at an operational level is also transformed into informational outcomes inside VMW, which allows for feedback and awareness of data quality within the organisation. This is a fact that will be utilised when making suggestions later in this chapter. In the theory review it has been acknowledged that AM is a multidisciplinary activity – the same is true for data collection for AM – the multidisciplinary considerations that have to be made, in order to achieve a balanced monitoring of the state of buried assets, can be achieved in VMW by involving all relevant departments in the planning process of the data collection schemes, including representatives that inherit strategic, tactical and operational AM functions.

## 6.3 Suggestions for structural and organisational measures to ensure data collection

Here, elements that are mainly from the data quality research framework of Lin et al. (2007), the quality framework for medical registries of Arts et al. (2002), and the theory of planned behaviour (Murphy, 2009) will be used to suggest concrete measures for how the organisation can facilitate compliance with the amount and quality of data that is desired. It will be distinguished between measures at the central level (head office) and the local level (provincial offices and service centres).

### 6.3.1 Measures at central level

At the central level (Brussels office) data quality must be:

- Planned by delegating responsibilities, making a policy, and designing collection training programs
- Executed by providing training, and initiating early utilisation of the data
- Monitored by analysing consistency and completeness measures, and communicating with the local offices

Concrete measures for facilitation of data quality compliance are listed in Table 17.

Table 17: Measures at the central level for data quality assurance

| Measure | Expected effect |
|---|---|
| **Data quality policy**<br>Should contain goals for and dimensions of data quality, and the overall measures and structures planned to reach these goals. Should also emphasise the value VMW places in data quality, to improve collector attitude. It is suggested to define measurable indicators for data quality, which can be used for benchmarking the data collection performance. | Ensuring emphasis on DQ (Lin et al., 2006).<br>Improved collector attitude (Murphy, 2009). |
| **Establishment of data collection protocols**<br>Should describe how data should be collected, and how it is expected to be collected. | Reference point for collection expectations (Arts et al., 2002) |
| **Multidisciplinary and multilevel involvement**<br>By involving several departments at the central office, one should be more able to identify the different objectives associated with data collection, than when only a technical perspective is involved. It is also known that early involvement of the system users (operational personnel) in the decision-making processes will reduce the resistance and improve the attitude to the systems. | Holistic thinking from multidisciplinary teams, positive attitudes from user involvement (Aven and Renn, 2010) |
| **Design of collection interfaces**<br>The design of the data collection forms is important. When Octopus is implemented, the operational personnel are supposed to receive work order forms generated by the aid of the forms, understand and act on the information in these forms, and report back to Octopus through data input forms. If the forms are perceived to be simple, effective, and relevant, are operational personnel more likely to use the system effectively. | Increased self-efficacy (Murphy, 2009) |

| | |
|---|---|
| **Make clear definitions of input semantics – make these available**<br>A clear, concise and available data dictionary should be established. The data dictionary should be easily accessible when reporting data in electronic forms. | Guidance for data collectors and consumers. (Arts et al., 2002) |
| **Give training in data collection**<br>Training should both be focused on the actual data collection process (understanding of input forms), and on the value placed in data quality compliance. It should be emphasised that data collection is an expected part of the job, and a prerequisite for good governance. It is also suggested to include general principles of AM in the training, accompanied by explanation of how the data is used, in order to clarify why data is valuable.<br>All changes in collection systems should be communicated in an understandable way. | Data collector skill improvement (Arts et al., 2002). Improved collector attitude (Murphy, 2009) |
| **Site visits and quality auditing**<br>Open meetings should be performed regularly, where issues from both the data consumer and collection side should be discussed. It is especially important that data collectors are allowed to express issues that are perceived as difficult. Two-way communication is recommended as most effective for achieving the desired (data collection) behaviour (Aven, 2010). | Communication of DQ problems (Arts et al., 2002). Improve data collectors' perceived behavioural control (Murphy, 2009). |
| **Initiate data utilisation**<br>If collected data is used for analysis at an early stage, data quality will be sought after by the analysts; this is in turn expected to increase the strategic priority of data, and thus creating expectations for data collection. Two ways of early data utilisation are suggested:<br>(1) In section 5.2 the timeframe of Octopus was identified as a limiting factor. It is suggested to utilise already existing data, from local service centre records, to create PI's on district metering area level. In this way, one may be able to set a good example of the usefulness of the data, while waiting for Octopus.<br>(2) Further, it is suggested that data is used to communicate with customers, by conveying information about disturbances and activities to the customers and allowing customers submit complaints. | Creating a driver for data quality (Murphy et al., 2008, Pipino et al., 2002). |
| **Testing of evaluator agreement and consistency over time**<br>Tests should be performed periodically to monitor consistency between different service centres and consistency within service centres over time (agreement between evaluators and consistency over time). Excessive variability should be investigated. | DQ monitoring and improvement (Arts et al., 2002, Migliaccio and Cordova-Alvidrez, 2011). |
| **Feedback with quality reports and recommendations**<br>Should be produced periodically and contain key indicators of data quality. Recommendations for improvement should also be given to each service centre.<br>It is suggested to make a standard data quality report form, where the data quality responsible in VMW at all times can view the overall status of the data quality indicators. | Guidance (Arts et al., 2002). Maintenance of collector attitude, improvement of collector self-efficacy (Murphy, 2009) |

In order to evaluate these measures more thoroughly and ensure that they are achieved, it is suggested to form a team of representatives from all the (relevant) departments at central level, and a representative from each of the provincial offices and service centres; it is of great importance that the different expectations and goals are addressed early in the planning stage, in order to plan for data collection that can convey the needs of the multiple disciplines within the organisation.

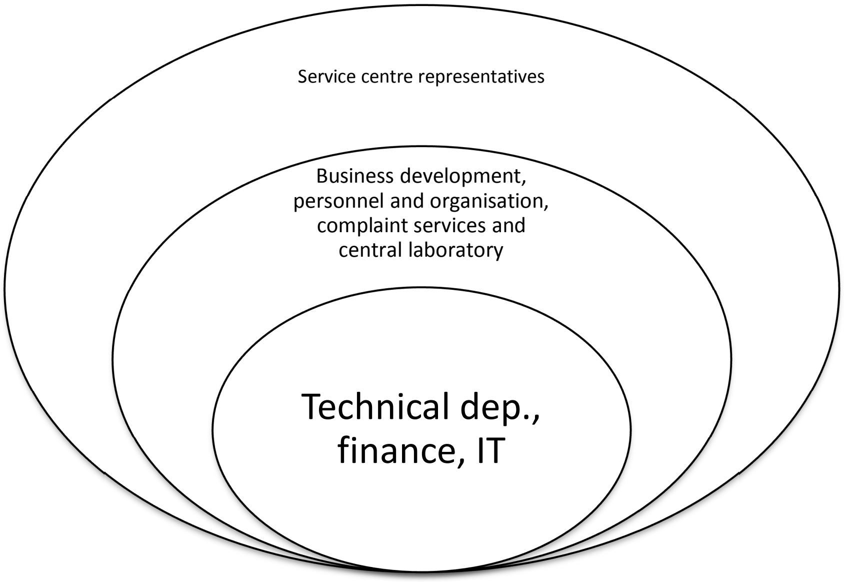


Figure 14: Three layers of planning involvement of the data collection scheme

It is further suggested to give the technical, financial and IT systems departments the most central roles in the data collection planning process, since these departments are most intensively involved in the management of asset life cycle information. The technical, financial and informational perspectives make the core of the planning team, with support from the other departments, and with feedback from representatives of the service centres, as described in Figure 14.

### 6.3.2 Measures at provincial and local level

The most important planning measure at provincial and local level (from now on labelled *local level*) is to assign a data contact person, which also coordinates the execution of the centrally planned training program. At the local level, it is important that data collection emerges gradually, and that the level of data utilisation proportionally follows the data collection level. Concrete measures to facilitate data collection at the local level are enclosed in Table 18.

Table 18: Measures at local level for data quality assurance

| Measure | Expected effect |
|---|---|
| **Assign a contact person**<br>Each local service office should have a person responsible for data quality. The contact person will naturally achieve higher levels of skill and knowledge, and act as an interface between the data consumers and the data collectors. Additionally, a responsible contact person for data collection will act as an agitator for including data collection into the work processes, and will also increase the social pressure to comply with the data requirements. It is suggested to ask someone to volunteer as a | Communication facilitation (Arts et al., 2002). Increased subjective norm (Murphy, 2009). |

| | |
|---|---|
| contact person. | |
| **Voluntary training**<br>It is suggested to make training programs voluntary for maintenance personnel. Experience has shown that voluntary training for staff, combined with prospects of promotions or other rewards resulting from the competence acquired in the training, successfully sets an example that the company management regards data collection as a measure of professionalism (Bhagwan, 2009). Other staff will become aware of this, and be more inclined to follow a good example. | Linking data collection to professionalism (Murphy, 2009, Bhagwan, 2009). |
| **Use current data collection procedures as basis**<br>Today, data is reported through paper forms. If the local service centres are guided to change the manner in which these forms are made, so that the reports immediately become more useful, the resistance to collect more data will be less when the new systems are introduced (Octopus and portable computers). Also, if there is a gradual change from what has previously been done, the change will not be perceived as excessive when the planned technical systems are ready. | Improve attitude (Murphy, 2009).<br>Minimising perceived change. |
| **Collective data collection**<br>It is a common concern that new data collection requirements will not be complied with, because the personnel responsible for it perceives it as "even more bureaucracy" or additional work in an already strained work environment. It is therefore suggested that reviewing of missing data becomes a routine at operational planning meetings, where time is allocated so that the data responsible can enquire about missing data. The service centre supervisor is often more inclined to comply with requirements from the management, and asking questions about data in team meetings will increase data awareness. Further, when personnel are being asked about data in an informal way, it is likely to be accepted more easily than with unfamiliar paper forms etc. | Improved attitude (Murphy, 2009).<br>Resolving problems and collecting missing data. |
| **Data quality indicators**<br>Establishing data quality indicators will help to monitor the data quality. If these indicators are made openly available, they will act as a statement of the value the organisation places in data quality compliance. | DQ monitoring (Arts et al., 2002). Improved attitude and subjective norm (Murphy, 2009). |
| **Use registry for local purposes**<br>If data is used for local purposes, it will prove to the personnel in the local service centres that data collection is useful. If the data is used and relied on at a local level, the expectations to comply with the collection standards within the work teams will be higher.<br>• It is suggested that the local service centres get the possibility to view PI's for the distribution network, which have been generated on the basis of their entries – graphical comparisons over time and between district areas will illustrate the usefulness of the data.<br>• As suggested in section 5.3, the data repository should be used to schedule planned activities, which can be followed up through Octopus. If work orders are conveyed through the same system as data is entered into, it forces personnel to interact with this system. | Increased subjective norm (Murphy, 2009).<br>Harmonisation of data drivers and informational outcomes (Murphy et al., 2008) |
| **Work process routines**<br>It is suggested to collect as much data as possible before and after an intervention. All data that can be registered in the comfort of the office should be registered there; if data is always collected at the end of a demanding assignment, will the fatigue of the data collector act | Improved self-efficacy (Murphy, 2009). |

| | |
|---|---|
| counterproductively. Data collection should therefore be stressed as an important element in both the planning and the completion of an assignment. | |
| **Data collection close to source – use portable devices.**<br>Data that has to be collected close to the source, should be collected as close to the source as possible. The use of portable computers will be instrumental to be able to carry out effective data collection at the data source. Data that can be easily obtained close to source, such as work hours invested, materials used etc., should be registered as close to the source as possible.<br>It is also suggested to allow attachment of photographs and/or sketches to registered data. Photographs may prove useful for reviewing data, and help to conjure memories and tacit information that can be transformed to input data at a later stage. | Higher level of compliance (Arts et al., 2002). Improved controllability (Murphy, 2009) |
| **Control and review data locally**<br>If data is controlled and verified locally it is more likely to be resolved (by direct communication with the personnel that possess the hands-on knowledge). It is also expected to increase the group expectations to comply with the quality requirements.<br>The central level data control should be more holistic, and focus on issues related to consistency. | Problem resolving (Arts et al., 2002). Improved subjective norm (Murphy, 2009). |

### 6.3.3 Data quality and performance indicators

The most important measures that have been proposed are the establishment of performance and data quality indicators:

- PI's are used to monitor the network characteristics (such as failure frequencies, complaint frequencies, average cost of repair, etc.). Establishment of PI's will show the benefit of collecting data at an early stage, without having to rely on extensive analysis. As time goes and more data is collected, the data will become less influenced by left-censoring, allowing for more advanced data utilisation. VMW could select 5-10 PI's, depending on what data that already are available.
- Data quality indicators measure the quality of the collected data. It is suggested to establish data quality indicators that measure completeness (percentage of filled in fields), amount of verified data, timeliness of data, and consistency. Suggested data quality indicators are shown in Table 19 (based on the attributes defined in Table 16, page 55).

**Table 19: Suggested data quality indicators**

| Data quality indicator | Explanation |
|---|---|
| **Verification completeness** | Percentage of total number of entries that are verified |
| **Schema completeness** | Average percentage of input fields completed in the entries |
| **Column completeness** | Percentage of one specific entry field type completed (e.g. completeness of *work hours* input field) |
| **Entry age** | Average age of entries |

Both types of indicators must be made available at the local service centres, and should be easily displayable for different subset categories of data.

## 6.4 Discussion and conclusions

Measures that can be implemented in the organisation to facilitate quality data have been made for the planning, collecting and monitoring of the data collection process. It is suggested to implement these measures at different phases during the implementation of the data collection routines; it is proposed to implement data utilisation, PI's, data quality indicators and informal collective data collection at an early stage. More formal measures, such as feedback reports, formal quality auditing etc., are best saved for later, when data collection has been a more accepted part of the work routine in the organisation.

Table 20: Three steps of data collection implementation

| Planning phase | Initiation phase | Established phase |
|---|---|---|
| Data quality responsible (central) | (Voluntary) training (local) | Work process review |
| Early user involvement | Influence existing routines | Data collection close to source |
| Data contact persons (local) | Collective data collection | Site visits and quality auditing |
| Define policy and protocols | PI's | Consistency control |
| Plan training | Data quality indicators | Local review and control of data |
| | Data utilisation (local) | Feedback reports |

There is an important principle behind implementing the data collection in different phases; Kelman (1958) suggest that behavioural change (towards data collection compliance) occurs in three steps: compliance, identification and internalisation. The idea behind the implementation process proposed in Table 20 is that some of the personnel *comply* by collecting some data in the initiation phase (through voluntary training and collective data collection) and early involvement in the planning phase, whereby certain (organisational and personal) benefits from data collection are *identified* by the personnel. After a short period of data collection will the establishment of the PI's and local data utiliste facilitate the personnel's *internalisation* of the data collection procedure – meaning that the value of the data collection actions have been accepted as rewarding for the group.

So, first the utility collects some data collectively (even if this may be of low quality) in order to achieve initial compliance and identification of the benefits. Then, when some informational outcomes have been produced, and internalisation has been achieved, the emphasis on stricter requirements for data quality can be introduced. In this way one will be able to utilise the fact that personnel's inclination to comply with data collection requirements are dependent on the perceived value of data; this is illustrated in Figure 15, where the data collection level is low until PI's are introduced, and the perceived value of data works as a driver for the collection of data. Feedback on quality of data can be given when the process is well-established.

While waiting for Octopus to be operative, VMW should focus on using and transforming existing data collection routines (the paper forms), so that these can help to produce informational outcomes on short terms. If these collection routines can be improved, and one can show the results to the local service centres, the personnel will be more inclined to accept to collect more (and more precise) data when Octopus is ready. One temporary measure to achieve better data quality is to require the reporting of an asset reference on all incidents that are repaired, from the maps that personnel in the field use, and to use the GIS department to register these in the GIS repository – in this way, a much higher amount of data will be more easily available for analysis, and it will also guide the data collectors into the data quality requirements that are set by the implementation of systems such as USTORE, Octopus and LCC-AM/QM.

Following this implementation process, where the data collection compliance level is expected to grow over time, by voluntariness, and collective adaptation and learning, one may be able to avoid the common objections that often arise from sudden "top-down" requirements to change working routines. Hopefully, the early onset of informational outcomes from the PI's will curtail the organisational inertia that is often observed when initiating AM programmes (Cromwell and Speranza, 2007), and boost positive attitudes towards data collection value.

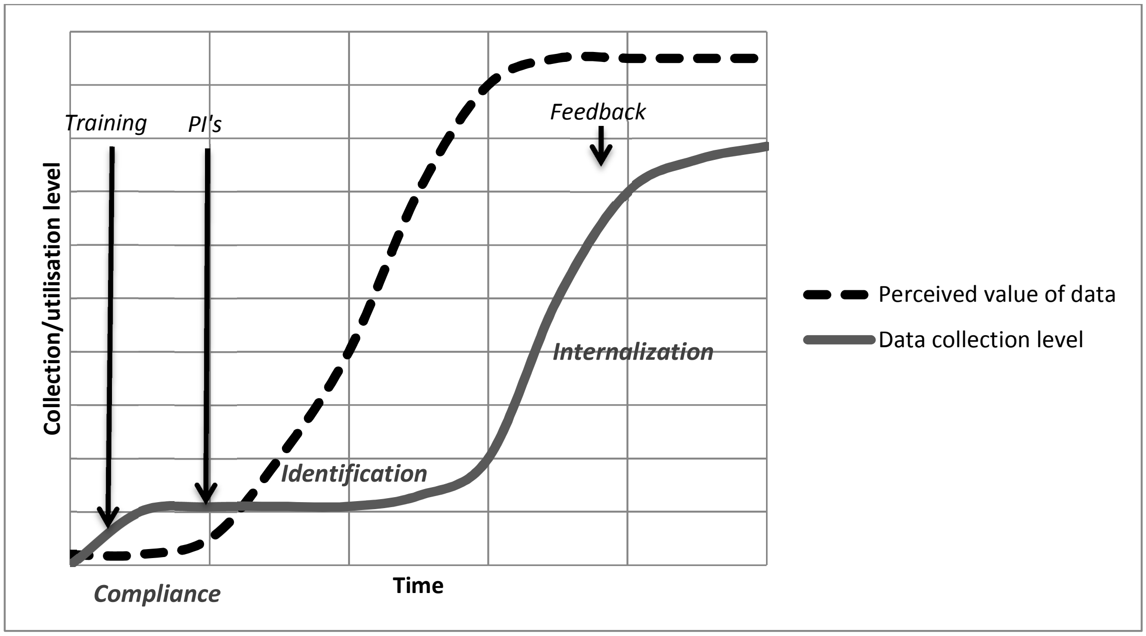


Figure 15: The idea of utilising perceived value of data as driver for data collection

One should be able to show the personnel, through all the measures that have been suggested here, that data collection has a positive effect on a personal, group and organisational level.

# 7 Case study: Trondheim Vann og avløp

Technical and organisational measures to improve the current data collection situation in VMW were suggested in chapter 5 and 6. The case study in this chapter aims at gathering more practical experiences about data collection and utilisation, and to identify threats and opportunities related to data in a water utility. The scope and questions asked in the case study is based on the theory review and the analyses that have been made in VMW in the previous chapters.

*Trondheim Vann og avløp* has been selected as a case study area because has had a longer experience period with respect to data collection and utilisation (by being involved in several scientific projects in cooperation with NTNU and SINTEF). There are two central questions to be answered in the case study:

1. What experiences has *Trondheim Vann og avløp* made with the collection and quality control of the data that is needed for asset management (AM)?
2. How has *Trondheim Vann og avløp* worked to utilise the data, and what has been the benefits (in informational outcomes)?

The methodologies have been a literature study of project reports from NTNU and SINTEF and direct inquiry with the staff in *Trondheim Vann og avløp*. Examples from both the water distribution and the drainage system will be used, since the principles of planning are similar for the examples that have been selected here. *Trondheim Vann og avløp* will be compared with *VMW East Flanders*, since *VMW East Flanders* has been used as a case study area for statistical analysis of failures previously (Rokstad, 2011a).

## 7.1 *Trondheim Vann og avløp* compared *to VMW East Flanders*

Trondheim is a city of about 175 000 inhabitants (Eiksund and Relling, 2012), situated in Sør-Trøndelag, Norway. The city is served by a municipal water and sewerage department, which extract water from the surface water source *Jonsvatnet*, treats the water by disinfection and corrosion control (Trondheim kommune, 2011), and distribute the water through 750 km of water distribution pipes (Trondheim kommune, 2007).

Table 21 contains some key figures from Trondheim and East Flanders. On average is the water network in Trondheim twelve years older than the VMW network in East Flanders. Due to the topography is the water distribution system in Trondheim much more reliant on pumping stations than in East Flanders. The average failure rate (of 2010) is not very different, but the failure rate in Trondheim is slightly higher than that of East Flanders, as one would expect from a twelve year older network.

Figure 16 shows the material distribution on the distribution pipes in Trondheim and in East Flanders. As one may see are the characteristics of the two networks totally different – while Trondheim is dominated by cast iron (approximately 80% ductile or grey cast iron), is East Flanders dominated (79 %) by PVC and asbestos cement.

Table 21: Key figures for Trondheim Vann og avløp and VMW East Flanders

| | Trondheim Vann og avløp | VMW East Flanders |
|---|---|---|
| Number of connections | N/A | 211 000 |
| Number of clients | 175 000 | 492 000 |
| Total network length [km] | 750 | 4800 |
| Average network age [years] | 40 | 28 |
| Number of pumping stations | 22 | 5 |
| Number of production centres | 1 | 4 |
| Average failure rate [#/km/year] | 0.29 | 0.25 |
| Sources | (Høseggen, 2012, Trondheim kommune, 2011, Trondheim byteknikk, 2005) | (Rokstad, 2011a, VMW, 2010, VMW Oost-Vlaanderen, 2010) |

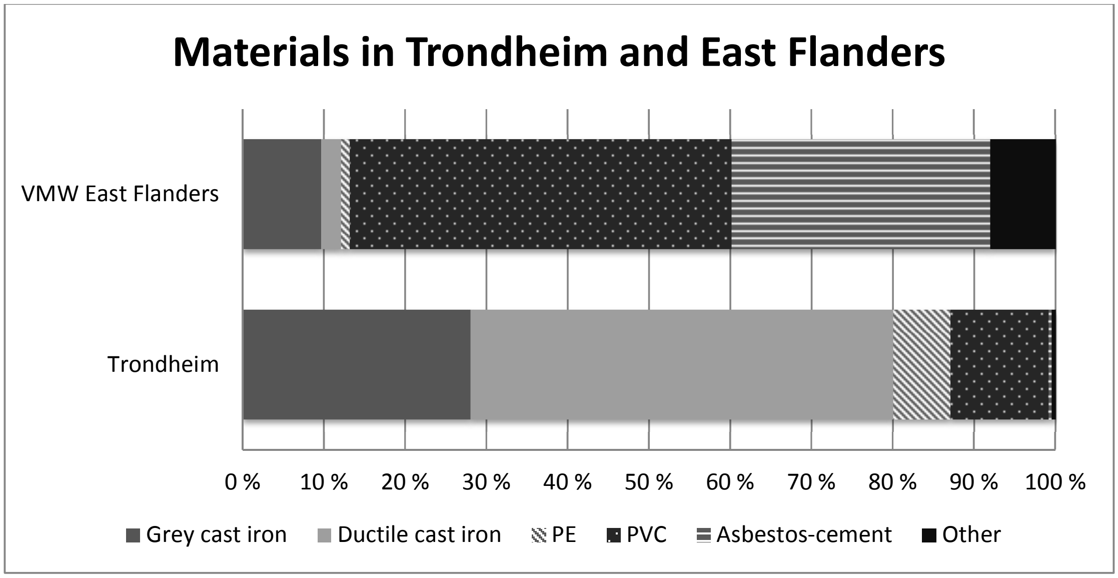


Figure 16: Material distributions in Trondheim and East Flanders (Høseggen, 2012, Rokstad, 2011a)

## 7.2 AM experiences in Trondheim – selected issues and initiatives

Section 7.2.1 describes the experiences with data collection in Trondheim, whereas sections 7.2.2 to 7.2.5 describe selected data utilisation methods.

### 7.2.1 Experiences with data collection and the use of Gemini VA

**Data consuming tools:** *Trondheim vann og avløp* uses *Gemini VA*, a Norwegian inventory and operating data software solution (Powel, 2011). *Gemini VA* is in essence an asset inventory, where spatial and non-spatial data about water utility assets are stored in a relational database, and can be mapped in the application's main window. In addition contains the software a diary (*Gemini Dagbok*), wherein a limited set of data about repairs, leakages, inspection and (planned and executed) maintenance works can be registered on assets. The diary entries can also be spatially referenced. *Gemini Melding*, an add-on for complaint management, is also utilised in Trondheim.

For logging and storing data from on-line pressure and flow gauges is the *EA Driftskontroll og overvåking* used (www.ipj.no, 2012). A separate software for inspection and maintenance management of reduction- and zone-separating valves is used, the *DASH FDV* (DASH Software AS, 2011); it is being evaluated if this is also to be registered in *Gemini VA*.

Today, a *Mike Urban* hydraulic model is used in Trondheim (DHI, 2011). On a day-to-day basis is the model mainly used to calculate fire flow capacity and effects on pressures the closing valves have (for maintenance work). The model has also been used sporadically to trace water travelling paths. (Høseggen, 2012)

**Data collection history:** Registration of operational log books started in Trondheim in the early 1950's. An electronic database for failures was already established in 1975 for Trondheim's water distribution system. However, the collection of failure data in the period 1975-1987 is scarce, and the records may not be considered to be "complete" until 1988 (Røstum, 2000). *Trondheim Vann og avløp* does hence have 24 years of more or less complete data on water mains failures.

Since 1975 have approximately 60 000 diary events been stored in Trondheim (between 4 000-6 000 per year the past ten years) for water and waste water systems. The most frequent events are flushing works (ca. 22 200), pipe inspections (ca. 16 300), and breaks/leakages (ca. 5 800). In addition are approximately 8 000 pictures and sketches stored and linked to assets in the database. Portable computers are not yet used for registering data about repair and maintenance incidents. (Høseggen, 2012)

Since registration of operational activities started already some 60 years ago, does data collection come from "a long tradition" in Trondheim – data collection has been a part of the job since long before the careers of the current staff started. In Trondheim is operational work both executed by personnel employed at the municipality, and by external entrepreneurs. Experiences show that external entrepreneurs are less inclined to comply with the data collection requirements than the personnel employed by the municipality; the external entrepreneurs do not know the value the data has for the municipality. The managers of the utility wants the data collectors to "see the whole picture", i.e. all personnel who collect data learn what the data is for – naturally, that is easier to accomplish internally, than for external stakeholders. (Ellefsen, 2012)

**Quality assurance:** *Gemini VA* has limited capabilities of documenting data quality. Data quality tagging is not inherent in the database structure. Data quality control in Trondheim is ensured only through data collection standards, graded access (some users can only view data, some can edit a limited amount of fields, while some can edit all fields), and automatic compatibility control routines[10]. According to *Trondheim Vann og avløp* is the most pressing challenge with regard to data quality that the way data is collected – the definitions and semantics of input – has changed gradually over time; i.e. the meaning behind a reported incident in 1988 is different than one from 2012. (Ellefsen, 2012)

**Benefits and drawbacks:** In the day-to-day operation is the data in the diary perceived as important, especially since the data is always accessible for operational personnel in the *Gemini VA* user interface. The history of failure repairs is easily represented spatially in the map user interface that operational personnel use. Further can the planning (scheduling) of activities also be administered in

---

[10] A compatibility control is a control wherein illogical or impossible values are detected, mostly applied to inventory data compatibility, for example the compatibility between material and diameter

*Gemini VA;* the main data repository does thus automatically become an integrated part of the day-to-day operation and maintenance of the network. (Ellefsen, 2012)

*Gemini VA* also has a set of automatic statistical reports that can be produced, but does not allow report production outside a standard framework of data combinations; functionality related to linking basic inventory characteristics to failure history is especially missed (Høseggen, 2012, Røstum, 2000).

From a planning perspective is one of the largest problems for Trondheim today that it is difficult to combine all the necessary data to form informational outcomes (such as performance indicators (PI's)). *Gemini VA* is a proprietary system, and does not have the possibility to store all sources of data. Combining data sources does then become more challenging, and is done only sporadically (for instance in a scientific case study, where a certain data combination is desired). When it is difficult to combine data sources, data will not be used optimally, especially not in day-to-day planning (Ellefsen, 2012).

In the next sections a selection of **examples of data utilisation** for AM are described.

### 7.2.2 Statistical failure modelling of pipe failures (Røstum, 2000)

The thesis of Røstum (2000) demonstrates the use of statistical models for pipe failures (the non-homogenous birth-process, the Weibull proportional hazard method (PHM) and the Cox PHM). The models were calibrated with failure observations from 1988 to 1996, amounting to 1897 failure observations. Observations from the period 1997-1999 were used as verification.

Covariates that were used in the analysis were material, length, diameter, age, presence of clay, presence of artificial masses, and number of previous failures. Calibrations were made both on cohort level and individual pipe level. Calibrations for different materials were executed separately. The calibrations on cohort level (where pipes were grouped according to similar properties) yielded precise results with good prediction capabilities. At individual pipe level were the prediction powers not as strong, but still sufficient in the situations where the number of failure observations was high enough for the specific statistical stratum.

The study by Røstum (2000) showed *Trondheim Vann og avløp* that readily available data about asset characteristics could explain the distribution of failure rates for different pipe cohorts. It also showed that nine years of failure data collection was sufficient to obtain predicted estimates on failure rates at network level. The model results also gave insight in the relative importance and effect of the different factors, providing useful knowledge about pipe deterioration-affecting factors; for instance was the effect of sulphate reducing bacteria (SRB) in clay environment on unprotected ductile iron pipes "confirmed" and could be assessed quantitatively by applying the models. On the other hand, the study also showed that predictions of failures on individual component level require a much higher level of data availability, and is not always possible.

### 7.2.3 Renewal planning with KANEW LTP

SINTEF made a long-term renewal need prognosis for both the water distribution (Selseth and Sægrov, 2001) and the waste water (Selseth and Røstum, 2002) system pipes in Trondheim, with the KANEW software in the beginning of this millennium. KANEW is a cohort survival model (Liu et al., 2012) developed in the CARE-W project (Sægrov, 2005), and makes an analysis based on cohorts of pipes with similar characteristics and expected service lives, where survival functions are assigned to these cohorts. The tool is thus able to produce prognoses for the expected amount of pipe length needing to be replaced for a given year and cohort. Further, if assumptions about different renewal alternatives are entered into KANEW, will the software be able to produce the expected costs associated with the renewal scenario. (Baur and Herz, 1999, Herz, 1994, Herz, 1996)

Pipes with unknown material and/or year of installation must be discarded from a KANEW prognosis. The asset data from Gemini VA was "cleansed"; missing, impossible or illogical data (e.g. material and diameter are not compatible) were discarded, before it was inserted into KANEW,

The prognoses in Trondheim were made by grouping the pipes in the Gemini VA database into eight different cohorts for the water distribution, and nine different cohorts for the waste water pipes, selected through applying knowledge about material production history (material generation), production standards, installation standards, jointing techniques etc. Each cohort was then assigned with a survival function, based on the applied knowledge; KANEW requires three input values in order to be able to produce the survival function – a pessimistic, an average and an optimistic service life estimate, representing 100 %, 50 % and 10 % survival probability, respectively.

After the cohorts were assigned survival curves could a pessimistic, an average and an optimistic rehabilitation need scenario be produced for Trondheim. For the water distribution pipes was a prognosis made for 20 years (2000-2020) – this predicted that the renewal rate should be 4-10 km/year in 2000 and 5.5-7.5 km/year in 2020. The prognosis for the waste water pipes was produced for 50 years (2002-2052), and showed that the renewal rate would have to increase from 2-5 km/year in 2002, up to 6-10 km/year in 2052.

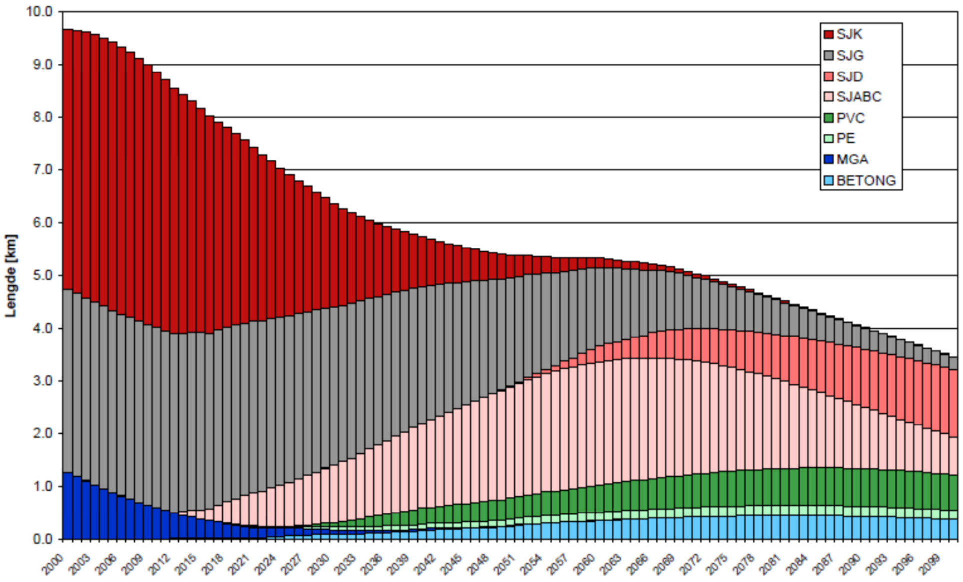


**Figure 17: Prognosis for water main renewal needs [km/year] in 2000-2100 (copied from Selseth and Sægrov (2001))**

For the water distribution network was a prognosis for 2020-2050 and one for 2050-2100 also made, and for the waste water network was a prognosis for 2050-2100 also made; all these with the same methodology as previously described. A graphical representation of a KANEW result is shown in Figure 17.

The expected rehabilitation costs through these periods were calculated by assuming different renewal techniques (rehabilitation or replacement), each with an estimated unit cost. Each of the materials in the analyses (for water distribution and waste water) was then assumed to be renewed with a certain proportion of each renewal technique. This made it possible to estimate the expected annual expenditure.

The KANEW approach is by no means a data-intensive analysis. It is an example of a methodology where knowledge and experience about material behaviour is used in the strategic planning process, where only basic data is utilised (materials, lengths, year of installation, nominal diameter). KANEW rather relies on knowledge (about materials, local conditions and practices) and assumptions (about renewal) than specific data. The accuracy of the methodology is limited by the accuracy of the survival functions.

The KANEW prognosis is a good example of a product that can be used for general network-level, long-term resource planning for the utility. The prognosis on expected rehabilitation costs have been known to be an effective communication measure to non-technical decision-makers, such as politicians.

### 7.2.4 Using PI's for identification of critical DMA's (Sjøvold and Selseth, 2009)

When *Trondheim Vann og Avløp* started the work on composing a new renewal plan (*saneringsplan*) for the water distribution network, it was based on the ideas of the Care-W project (Sægrov, 2005): Methodical planning based on objective and clear criteria, and utilisation of the available data. The planning was divided into three parts:

1. *Preliminary study: Setting the premises for the plan.* The goals for the plan were defined in this step, and PI's and other evaluation criteria were selected. This step was a network level consideration.
2. *Preliminary study: Selection of focus areas.* In this step PI's were used to evaluate the condition of different district meter areas (DMA's) in the city (Sjøvold and Selseth, 2009). Further was criticality and failure modelling conducted on asset and cohort level (Hafskjold and Selseth, 2008). (A KANEW analysis could also be part of this step.)
3. *Renewal plan for each district meter areas.* Based on the priorities in step 2, detailed asset-level renewal plans were made, using the annual resource planning (ARP) tool developed in Care-W.

Here, the work with PI's will be explained. The criticality and failure modelling will be explained in the following section (7.2.5).

There are two important rationales behind the decision that DMA level PI's were to be used. The first is that *Trondheim Vann og avløp* wanted to obtain an overview of the situation in the city, and find focus areas where more advanced and precise considerations could be made on asset level. The second is that one wanted to include water quality measures and complaints in the consideration, which are in most cases not linked to specific assets. *Trondheim Vann og avløp's* goal for the PI study was to identify the ten (of 32) DMA's that were in the worst state, which then would be subject for more detailed considerations in step 3 of the renewal plan.

PI's were selected based on (1) their ability to describe several aspects of the state of the DMAs, and (2) whether or not the benefit of the information was in proportion with the data necessary to construct the PI. *Trondheim Vann og avløp* also acknowledged that the PI's should be usable both in day-to-day operations, short-term and long-term planning, as well as benchmarking. The PI's that were selected were:

- **Leakage**, measured in percentage of total water production [%], cubic metres lost per year [m$^3$/year], and cubic meters lost per year and meter water main [m$^3$/year/m]. The data source for this PI was the water balance calculations based on water gauge readings in the boundaries of the DMA's. The three metrics were consolidated into one ranking, where each metric was weighted 0.1, 0.2 and 0.7, respectively.
- **Amount of non-planned disruptions**, measured in customer hours per year [hours/year]. The data source here was the *Gemini Dagbok* entries, and required that data about disruption durations and number of clients between two closing valves were available.
- **Number of repairs per length of water main** [#/km]. The data source here was also *Gemini Dagbok.*
- **Amount of water pipe sensitive to SRB corrosion** [%], measured as percentage of unprotected ductile cast iron mains. This PI reflects a problem distinctive for Trondheim, where areas with oxygen-poor clay ground have made conditions ideal for SRB corrosion of unprotected iron (Røstum, 2000).
- **Amount of galvanised pipes** [%]. The galvanised pipes in Trondheim are generally in poor condition, and are expected to need a high renewal rate.
- **Amount of asbestos-cement pipes** [%]. The asbestos-cement pipes in Trondheim are also generally in poor condition, and are expected to need a high renewal rate.

The data source for the three last PI's is the *Gemini VA* inventory. All PI's were calculated based on data for year 2007. All the PI's each gave a numeric value for each of the DMA's, and the PI numeric values were transformed into rankings from 1 to 6 (best to worst). All the six PI rankings were then consolidated into one joint PI, where the two first PI's were weighted 0.3, and the four last were weighted 0.1. The joint PI was used to identify the ten DMA's that were in the worst state, and should be prioritised in the detailed renewal planning (step 3). Radar diagrams were also produced, to visually show the difference between the DMA's – an example is shown in Figure 18.

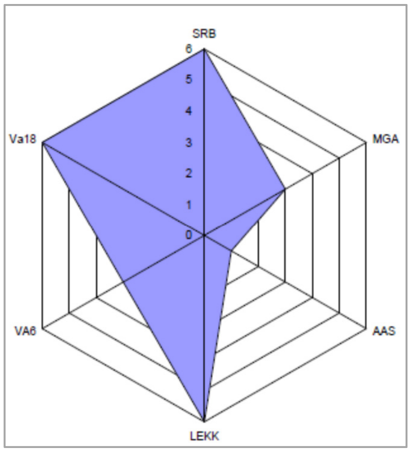


**Figure 18: Example PI radar diagram (copied from Sjøvold and Selseth (2009))**

The idea of PI's (and performance indices) also appear *in Trondheim Vann og avløp's* master plan (Trondheim byteknikk, 2005), where measurable goals for the level of performance and service are set. Examples of the goals described in the current water distribution master plan are *"customers should not be disconnected from the water services for more than three hours"* or *"in an average water consumption situation should the water pressure at a customer connection be between 30 and 80 metres"* (Trondheim byteknikk, 2005). These goals can be measured through PI's.

The use of PI's as a rationale for identification of focus areas for renewal shows how data on network level can be used to establish an overview of the state of the assets in different areas of the distribution system, and how this may serve as an instrument for comparison and prioritisation. It also shows that PI's can be selected depending on the data availability and that they can be constructed from small amounts of data (only data from 2007 were used), and that the data does not always have to be connected to a specific asset to be useful. However, it is also acknowledged that PI's like these only give a rough estimation of the state, and are therefore only precursors for more detailed planning in the prioritised areas: More specific planning with failure and criticality models is described in the next section.

### 7.2.5 Criticality and failure models as basis for renewal planning (Hafskjold and Selseth, 2008)

The previous section showed a DMA level application of PI's in order to establish which areas in Trondheim that should be prioritised. The next step in the renewal planning for *Trondheim Vann og avløp* was to analyse their distribution system on asset (pipe) level in order to establish the **probability of failure**, and the hydraulic **consequence of failure** for each water main.

**The probability of failure**, or failure rate, was assessed using the Casses software (Cemagref.fr, 2011), which is based on the LEYP model. LEYP predicts the failure rate of water mains based on three factors:

- The water mains' history of failures
- Ageing (time)
- A vector of explanatory covariates

LEYP is also used in the Aware-P toolbox (section 2.3.2, page 12). The failure history from 1988 to 2005, amounting to 3 665 failures, was used for calibration. Material, length, diameter and year of installation were extracted from the *Gemini VA* inventory, and used as explanatory covariates. The water pressure was estimated and extracted from an EPANET model, and used as an explanatory covariate. The analysts also wanted to include laying depth of the water main and the type of material surrounding the main as explanatory covariates; however, these factors proved to be too time-consuming to obtain.

The result from the analysis was a list with the probability of failure for each water main segment in Trondheim (in ranked order). The results showed that 14 % of the expected breakages in Trondheim's water mains could have been avoided by renewal of only 2 % of the water mains. Failure rates did not seem to be affected by the pressure, but the analysts doubted that the accuracy of the pressure calculations were sufficient.

It may be worth noting that a new analysis with Casses was executed in 2009-2010 (Bruaset et al., 2010), this time also with soil type, joint type, and corrosion protection added as covariates. The calibration was done with grey cast iron, ductile cast iron and other materials as separate calibration strata. This time did pressure indeed prove to be a significant contributing factor to the failure rate. The number of previous breaks proved to be the most determining factor, while ageing in general does not affect the failure rate. Clay soil was again confirmed as an important contributing failure factor for ductile cast iron mains. The analysis showed that the different factors included in the model apply to different degrees, depending on the material. In the list of the hundred mains with highest break probability, 85 % of them had two or more previous failures, average pressure of 7.5 bars (quite high), and were mostly of grey cast iron from the period 1964-1975. (Bruaset et al., 2010)

**The hydraulic consequence of failure** was assessed using the RELNET[11] application, developed in Care-W. RELNET gives each link in the distribution network a hydraulic criticality index (HCI) between 0 and 1. The calculation is done by closing network links individually, one at the time, and comparing the demand the system can satisfy with the link out of service, compared to the original capability of the system (Andralanov, 2012). The input needed for such a consideration is a hydraulic model file, with fairly accurate water consumption distributions.

The result from the RELNET analysis showed that the supply safety in Trondheim overall was quite good. About 500 (of ca. 10 000) water mains were identified as having considerably high hydraulic criticality, and these were recommended to be considered in detail in step 3 of the renewal plan. Maps showing critical mains were produced for each DMA in Trondheim.

**Comments:** The use of the two analytical tools Casses and RELNET shows how two aspects of supply safety risk (probability and hydraulic consequence) can be assessed, using available data about network characteristics, failure history and water consumption estimates. The results from these two tools were prioritised lists and maps showing the water mains that are most likely to fail or most hydraulically critical, and thus being worthy of more attention in the detailed planning. The analysis in Casses provides valuable insights into which factors contribute to risk of failure. The results from RELNET are especially useful for identifying pipes with medium criticality, since these are not so obvious for experienced operators.

The data needed for failure rate and criticality modelling tools are much higher than for the PI analysis at DMA level, or the long-term cohort prognosis in KANEW, but the result is also much more detailed, accurate, and less dependent on assumptions. The fact that the results can be connected to individual water mains makes the results applicable on a tactical and operational planning level.

### 7.2.6 On-going efforts in Trondheim

Currently there are several on-going efforts in *Trondheim Vann og avløp*, where the accumulated knowledge described in section 7.2.1 to 7.2.5 is developed further. Currently there are two master thesis projects related to AM in *Trondheim Vann og avløp*. One master thesis candidate is investigating the possibility of implementing the *Aware-P toolbox* in Trondheim, both for the water distribution and the sewerage system. The other candidate deals with improving the input data for KANEW. (Ugarelli, 2012)

---

[11] RELNET is a predecessor of CIMP (section 2.3.2, page 11)

## 7.3 Lessons learned from Trondheim case study

The previous sections have presented information about data collection and utilisation in Trondheim. Now, it is time to discuss this information in light of the theory and observations made previously in this thesis.

### 7.3.1 Data collection

*Trondheim Vann og avløp* can report a long history of data collection, with digitalised and (more or less) complete records already since 1988. The implementation of digitalised records and stricter collection requirements has run smoothly, because data collection has in some way always been part of the operational work processes. The management in *Trondheim Vann og avløp* has also focused on communicating why data needs to be collected ("the whole picture").

Systematic data collection over several years has made it possible to use advanced statistical models on asset level. It is interesting to see the evolution of the accuracy and trustworthiness of the results from the failure models; the results from 2000 (Røstum, 2000) were in some cases not precise on asset level predictions; in 2009 a new model was used (LEYP) with a twice as long observation window, this time the results on asset level were easier to obtain – this shows that more advanced tools can be implemented as the observation windows grows, and that data collection is a long-term investment for the utility. New explanatory factors have also been possible to add (e.g. pressure).

However, even though data has been collected for many years, there are some data that analysts miss, and this is reflected in the reports that have been reviewed in this case study. For instance, for one of the analyses with LEYP it was foreseen to use soil type as a covariate (because operational experiences have shown that soil type may be important), but the data was not stored in *Gemini VA*, and extracting the data from spatial analysis of external data sources did not succeed. Gathering missing data for the whole asset portfolio may be resource-draining.

### 7.3.2 Data utilisation

Through the case study it has been uncovered how *Trondheim Vann og avløp* utilise data in several analyses, wherein raw data about network characteristics and operations have been used to produce useful rationales for the resource planning in the utility – modelling tools have become a natural part of the planning process, which is proved by the fact that measurable goals and targets for the water distribution system are inherent in the master planning.

Four quite different ways of utilising data have been described in this case study, ranging on a wide scale, from tools that require only basic inventory data to tools that are very data-demanding – the way data can be utilised is dependent on the level of complexity of the model and the input data:

- On one side of this scale tools such as the KANEW LTP (section 7.2.3) and PI's (section 7.2.4) are situated, which are not very data-demanding. These tools are used to find overall investment needs and identify focus areas – the results are hence very superficial, and dependent on assumptions, but indeed very useful on a strategic planning level, e.g. "this issue is more problematic than that issue".
- Situated on the other side of this scale are tools such as the statistical failure models, and hydraulic criticality models (section 7.2.2 and 7.2.5). Such tools are much more data-demanding, less dependent on expert assumptions, and deliver more specific results. These tools are used to identify asset cohorts or specific assets that are in bad condition (high risk,

low reliability); i.e. results that are more useful on a tactical planning level. An example of an informational outcome in this type of model is the identification and the quantification of the effect of SRB on ductile cast iron pipes.

The experiences from Trondheim show that the planning tools can be adapted to the data availability, and that it is possible to produce useful information from a short observation window. Especially interesting are the PI's that were used to identify prioritised DMA's – PI's were chosen so that there should not be disproportionality between the effort needed to obtain the data and the usefulness of the result.

It is also interesting to discuss the synergy between the strategic and tactical level information. As an example, one was able to quantify the effect SRB corrosion had on the failure rate of ductile cast iron pipes in a complicated data-demanding model; later, this information was used to form a PI which shows the percentage of ductile cast iron pipes in each DMA – the specific knowledge obtained on an asset level model, can later be used as an generalised indicator of the condition of a larger subsystem.

The variation of data-intensity in the tools used in Trondheim is also reflected in the way their planning is organised. When renewal plans have been made in Trondheim, the approach has been to start at a strategic level, by defining goals and criteria, then move down to a tactical level where criticality and reliability for zones, cohorts or individual assets have been evaluated. Last, when each asset has been evaluated, the assets are prioritised based on multi-criteria analysis (Hafskjold and Selseth, 2008) – questions about project economics and timing are not introduced before this stage. One does hence start with processes that only require small amounts of data, and move towards more data-intensive evaluations when the candidate renewal projects are identified. In this way, one does not spend time on comparing project details (which require much data) before the most important candidate projects are selected.

### 7.3.3 Problematic data issues

The most important problem identified in *Trondheim Vann og avløp* is on data quality, and not data quantity. Even though the operational data diary is believed to be fairly complete since 1988, are the data quality problems such as inconsistencies over time, wrong and lacking data identified in Trondheim. *Gemini VA* does not have any quality tagging system, which makes it virtually impossible to track data quality (except for compatibility checks).

The problems with data quality in Trondheim have three main causes: (1) data definitions (dictionary) are not existing, easily available or known to the personnel who provide the data, (2) the systems for quality control are not sufficient, and (3) the dispersed and proprietary storage structure makes it difficult to combine data sources, so that data will not be used optimally. The difficulty with combining data sources makes it especially problematic to use data optimally in a day-to-day operational context.

### 7.3.4 Implications for VMW – some "take home messages"

Aspects regarding data collection and utilisation have been studied in *Trondheim Vann og avløp*. But what can VMW learn from this, and how do the findings in Trondheim confirm or conflict the previous findings in this thesis? Some important points from Trondheim, to "take home" to VMW are:

- *Trondheim Vann og avløp* has taken advantage of existing data collection processes when digital repositories were introduced. Since personnel in VMW also fill in paper forms and make sketches, this is also something that could be done in VMW – the data collection requirements will meet less resistance in VMW if the data collection routines that exist today are gradually changed.
- The problems related to data quality in Trondheim have three main reasons: (1) the data storage structure is dispersed and proprietary, which makes it difficult to combine data sources and optimise the value of the collected data, especially on day-to-day basis, (2) the data definitions (dictionary) are not known and understood uniformly by the personnel who collect the data, which results in inconsistencies, and (3) the main data repository, *Gemini VA*, does not have sufficient capabilities to document data quality, which again forces analysts to be more careful with their conclusions. These three matters have all been discussed previously, and the experiences from Trondheim confirm that these issues should be taken seriously in VMW also.
- The study in *Trondheim Vann og avløp* has shown that it is possible to produce useful informational outcomes from small observation windows with limited data. The results obtained for strategic (long-term) planning in Trondheim, in KANEW LTP and with PI's, have been obtained with a very limited amount of data. Quantifiable predictions (e.g. failure rates and criticality) on asset level require much more data of higher quality, but this data is more useful on a tactical and operational level. VMW should consider focusing on simple tools on strategic level the first years after data collection has started, and await the implementation of models on a tactical and operational level which they do not yet have sufficient data for. In other words, general long-term resource planning and identification of focus areas (by PI analysis) should be prioritised. Some measurable PI targets should also be set in the long-term plans for VMW.

# 8 Cost-benefit analysis of data collection

Collection of data is resource-consuming, and in chapter 5 it was found that "data is not used optimally in VMW". Collection of data should be balanced with the consumption of data (to produce informational outcomes). In this balancing process are data providers often tempted to simplify models and tools so that the data requirements will be reduced, while analysts are often tempted to incorporate an excess of factors in their models, bearing no regard of the high data requirements it imposes (Alegre et al., 2006a). The challenge is to find a suitable level of data collection, where the data collection efforts stand in proportion with the value of the informational outcomes they can provide – a cost-benefit analysis of data collection.

Some of the questions one could answer with such an analysis are: What will be the cost of data for a given combination of asset management (AM) tools? What combination of tools is optimal for the utility, with respect to data cost and information output? What data are most valuable? What data should one prioritise to collect when a new collection system is introduced?

This chapter proposes a methodology for assessing the cost-benefit relationship between data cost and management information benefits for buried assets. The methodology has been calibrated for VMW's buried assets in East Flanders.

## 8.1 Defining the model

The objective of the cost-benefit analysis is to establish a relationship between:

1. The cost associated with collection of different classes of data, which are related to the AM of buried infrastructure
2. … And the benefit these data yield, measured as achieved informational outcomes.

An example of a data cost may be the work hours required to fill in and process data after a repair has been completed. An informational outcome may for instance be the reliability of a specific type of asset. Ultimately may the benefits be expressed as the avoidance of losses, due to unforeseen failures or unexpected and premature investments.

### 8.1.1 Model assumptions and structure

The analysis is carried out with *AM tools* as the common denominator. In this context is an AM tool defined as an *instrument that transforms raw data into useful informational outcomes*. The model used to establish the cost-benefit relationship is illustrated in Figure 19. The assumption is that a certain set of input data classes, treated through a certain set of tools will produce a certain set of informational outcomes. The combination of data classes and tools may be synergetic, i.e. the production of informational outcomes may be greater when tools and data sources are combined, than when separated. For each combination of tools, it is possible to calculate a cost and a benefit measure.

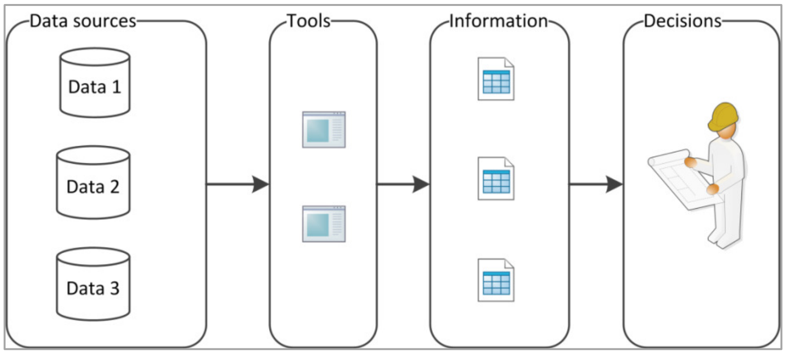


**Figure 19: Cost-benefit data production model assumption**

It is assumed that VMW develops or acquires a GIS-viewer based set of data collection forms in the Octopus project, where linking operational activity and events to physical assets can be done relatively effectively – considerations where data has to be replicated without connection to the asset inventory will not be made.

The cost-benefit model is built in a spread sheet. The model has three sheets:

1. **The data collection cost sheet** (DataCost). Here, all the different data items are described and assigned costs. The data requirement for different tool combinations are also defined in this sheet (see section 8.1.2)
2. **The base cost sheet** (ToolCost). In this sheet are the base costs of the different tools calculated.
3. **Results sheet** (Results). Costs and benefits of the different tools are summarised in this sheet, based on the two cost sheets. For each informational outcome have certain rules and criteria been defined to determine whether or not they emerge from the tool combination in question. The sheet displays the preliminary results, which are to be used in the further evaluation.

In addition to the three model sheets, there are two sheets for evaluation the results:

4. **The tool comparison sheet** (ToolComparison), where the different tool combinations can be compared graphically with respect to the benefits as information items, informational benefits or achieved planning strategies. Weighting factors can be assigned in order to differentiate the importance of the different outcomes within each group. The different tool combinations are ranked[12] based on the outcomes and the weighting factors.
5. **The data comparison sheet** (DataComparison), where the cost-benefit ratio of different classes of data can be compared graphically, based on the weighting factors from the previous sheet. A number of data dependencies also been defined in this sheet; i.e. if the utility collect one class of data, it must also collect another – for instance, it makes no sense for a utility to collect specific data about the cost of an inspection, if the basic data about the inspection is not accounted for.

In sheet 4 is the objective to evaluate and select a suitable combination of tools for the utility. In sheet 5 is the objective to evaluate the cost-benefit ratio for different data classes directly, in order

---

[12] The ranking is primarily used to assess the sensitivity of the model (section 8.3)

to assess what data are most valuable, regardless of tool combinations. The structure of the spread sheet model is illustrated in Figure 20. In section 8.1.2 to 8.1.5 will the meaning behind the elements in this model be explained further.

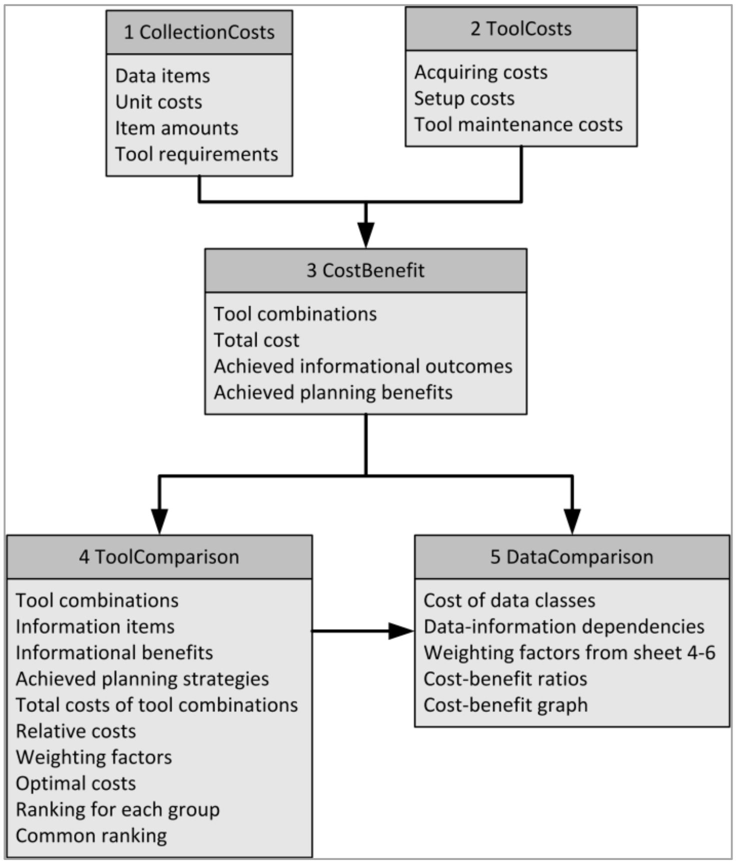


**Figure 20: Structure of spread sheet model**

The user instructions for the spread sheet model are enclosed in Appendix C.

### 8.1.2 Data costs

There are two types of costs associated with the production of informational benefits: (1) base costs and (2) data collection costs.

**The base costs** are costs associated with purchasing tools, treating data for analysis, and maintaining the tools. The base costs are calculated according to equation (8.1). For the acquiring cost are both one-time purchase and annual license costs for software included, and distributed over the expected lifetime of the software (usually 10 years). Setup costs are also distributed over the lifetime of the tool. Interest rates and inflation are not included in the calculation in equation (8.1).

$$C_{\text{base}} = \frac{C_{\text{tool}}}{C_{\text{workhour}}} + \frac{WH_{\text{setup}}}{UL_{\text{tool}}} + WH_{\text{maintenance}}$$

where:

$C_{\text{base}}$ = the base cost $\left[{}^{\text{hours}}\!/_{\text{year}}\right]$

$C_{\text{tool}}$ = the acquiring cost of the tool $\left[{}^{€}\!/_{\text{year}}\right]$ (8.1)

$C_{\text{workhour}}$ = the average cost of a work hour $\left[{}^{€}\!/_{\text{hours}}\right]$

$WH_{\text{setup}}$ = the work hours necessary to set up the tool $[\text{hours}]$

$UL_{\text{tool}}$ = the expected usable lifespan of the tool $[\text{years}]$

$WH_{\text{maintenance}}$ = the work hours necessary to maintain and use the tool $\left[{}^{\text{hours}}\!/_{\text{year}}\right]$

**The data collection costs** are costs associated with the time invested in collecting the data. In order to calculate the costs associated with data collection for different scenarios, different types of data have been organised according to Table 22. Data are organised into classes and sub-classes. A data class is defined as *a set of data items that are internally similar with respect to what kind of information they are describing, their aggregation level, and how they are collected.*

Table 22: Data item attributes

| Attribute | Explanation | Example |
|---|---|---|
| Item # | Identification number of the item | 3 |
| Class name | Name of the data class the entry item belongs to | Failure |
| Sub-class name | Identifies a sub-branch of the data class | Basic |
| Data aggregation level | The level at which the entry it linked. District meter area, cohort, address or asset levels are possible. | Asset |
| Item name | The name of the data item | Date of failure |
| Number of entries per year | The number of events occurring each year, that need to be continuously recorded (new assets, repairs, inspections, complaints, decommissioned assets etc.) | 200 |
| Data backlog | Backlog of missing or unrecorded data (missing diameters, materials etc.), and the assumed time horizon for collecting these. If the utility does not want to invest resources in data backlogs, this field can be excluded from the model. | 1600 |
| Backlog horizon | The time horizon under which the utility will spend resources to fill in data backlog in order to make records complete. | 15 years |
| Unit work hour cost | The estimated number of work hours associated with collecting one item of the data class. The unit cost is different (higher) for collection of backlogged data, than for collection of current/new data. | 1 min |
| Tools using the data item | Each data point has a list of tools that use the data field. Data can be indicated as absolutely required (1) or optional (2). | LLIFE, FAIL, USTORE… |

The data classes and items have been defined in Appendix B. The cost to collect one data item per year can then be calculated according to equation (8.2):

$$c_{\text{data item } i} = n_{\text{cont.},i} \cdot t_{\text{cont.},i} + \frac{n_{\text{backlog},i}}{BH_i} \cdot t_{\text{backlog},i}$$

where:

$c_{\text{data item } i}$ = the cost of data item $i$ $\left[{}^{\text{min}}\!/_{\text{year}}\right]$

$n_{\text{cont.},i}$ = the number of item $i$ that needs to be registered each year (8.2)

$t_{\text{cont.},i}$ = the unit cost to collect item $i$ [minutes]

$n_{\text{backlog},i}$ = the number of backlogged item $i$'s

$t_{\text{backlog},i}$ = the unit cost to collect a backlogged item $i$ [minutes]

$BH_i$ = the backlog correction horizon (time until all data should be registered) [years]

The cost of collecting data required by a certain combination of tools can be calculated according to equation (8.3):

$$C_{min} = \sum_{i}^{\text{all data required}} \left(c_{\text{data item } i}\right)$$

$$C_{max} = \sum_{i}^{\text{all data required}} \left(c_{\text{data item } i}\right) + \sum_{i}^{\text{all optional data}} \left(c_{\text{data item } i}\right)$$

where: (8.3)

$C_{min}$ = the cost of collecting all data items that are required by the tool combination $\left[{}^{\text{min}}\!/_{\text{year}}\right]$

$C_{max}$ = the cost of collecting data that are both required or optional in the tool combination $\left[{}^{\text{min}}\!/_{\text{year}}\right]$

The data requirements have been defined for each tool in the spread sheet. Each data item is indicated as *1* if it is absolutely required by the tool, *2* if it is optional, and *3* if it is not relevant for the tool. When there is more than one tool in a tool combination, the minimum of the indicators are selected for each data item when the costs are calculated. Data items indicated as *1* are included in the calculation of $C_{\text{min}}$, whereas data items indicated as *1* or *2* are included in the calculation of $C_{\text{max}}$.

**The total cost of data** is hence dependent on the cost of setting up and maintaining the tools that treat the data, the number of asset events occurring each year, the backlog of unregistered items and the unit cost to record these items.

### 8.1.3 Included tools

The tools act as a bridge between raw data and informational outcomes. Different tool combinations use different raw data, and enable the utility to produce different informational outcomes. In this analysis, the tools that will be included and combined are:

- VMW's GIS inventory
- The analytical tools that have been evaluated by VMW (section 4), i.e. LCC-AM/QM and USTORE.
- The analytical tools described in the Trondheim case study (section 7.2), i.e. KANEW LTP.
- Octopus / diary functions (section 5.3.2)
- The tools in the Aware-P toolbox (section 2.3.2)

The *Octopus / diary functions* are functions that are related to storing all life cycle events (as described in 5.3.2), including inspection results, water quality samples, work order management and customer (complaint) communication. The tools in the *Aware-P toolbox* are introduced, because each module in the toolbox nicely represents an AM informational outcome "idea" (e.g. FAIL represents the idea of modelling failure rates based on failure history and asset-specific covariates, CIMP represents the idea of assessing asset criticality, etc.). Only KANEW LTP is explicitly included from the Trondheim case study, because all the other concepts described in the case study are already included in the Aware-P toolbox in some way. Note that the *Aware-P PI tool* will not be included, since the PI's in the Aware-P library are so diverse, and the data demand is very dependent on which PI's the utility selects. The benefit of the PI tool is more or less constant, and could be assessed in a separate analysis.

### 8.1.4 Benefits

The benefits in this model are expressed as informational outcomes. In this context is an *informational outcome* defined as an *independent set of information at a certain aggregation level, which can be used to assess a certain aspect of the assets' characteristics*. The informational outcomes are listed in Appendix B, Table 36. The outcomes will also emerge in the results from the analysis.

The outcomes will be expressed in three different ways:

1. As information items (such as *inventory, performance* or *reliability*)
2. As informational benefits of the information items (which of "the six whats" (Vanier, 2001) it answers, which of the planning levels (strategic, tactical and/or operational) it acts on, whether or not it helps to control goal or regulatory compliance, and whether or not the information is instrumental for resource planning).
3. As the planning strategies the information items allow the utility to plan by.

It is when information is used as rationales for planning, that the information receives a true decision incentive value. It is therefore the planning strategies (point 3) have been included as benefits. The strategies that have been identified are:

- Reactive
- Time-based
- Performance-based (decisions are made to avoid that assets are performing below standard)
- Condition-based
- Criticality-based (decisions are made to avoid assets with excessive criticality)
- Proactive reliability-based
- Proactive risk-based
- Predictive or LCC-based
- Predictive risk- and LCC-based (combination of LCC and risk considerations)

(Some of these strategies were reviewed in section 2.1.4; the ones that were not described there have descriptions in brackets)

Within each of the three benefit groups, the benefit will be evaluated as the completeness of all the possible items, as expressed in equation (8.4):

$$\text{Completeness}\left[\%\right] = \frac{\sum_{i=1}^{n}\left(w_i \cdot x_i\right)}{\sum_{i=1}^{n} w_i}$$

where:

$$w_i = \text{ the weighting factor for benefit } i \left[0{,}100\%\right] \qquad (8.4)$$

$$x_i = \begin{cases} 1 & \text{if benefit } i \text{ is present} \\ 0 & \text{if benefit } i \text{ is not present} \end{cases}$$

$n =$ the number of possible benefits in the group

In the end, the three benefit types will be calculated as an aggregated measure, the **Total informational completeness**, which is calculated in the same way as in equation (8.4).

### 8.1.5 Model input data

The input for this model is based on the GIS inventory and operational data from VMW East Flanders:

#### 8.1.5.1 Data amounts

The estimated data amounts from each data class are based on the numbers shown in Table 23.

Table 23: Data amounts generated in VMW

| Entry | Number | Sources |
|---|---|---|
| New assets in inventory [#/year] | 2054 | (Rokstad, 2011b) [13] |
| Decommissioned assets [#/year] | 184 | |
| Repairs / unplanned interventions [#/year] | 1200 | (VMW Oost-Vlaanderen, 2010), (VMW Oost-Vlaanderen, 2009) |
| Complaints [#/year] | 940 | |
| New connections [#/year] | 2067 | |
| Maintenance on existing connections [#/year] | 3276 | |
| Flushing [#/year] | 6771 | |
| Leak detections [#/year] | 350 | |
| Repair history backlog [#] | 1600 | (Rokstad, 2011b) |

The numbers extracted from the activity reports (VMW Oost-Vlaanderen (2009) and (2010)) are believed to be fairly consistent over time. Backlogged data will not be considered in the analysis.

#### 8.1.5.2 Data unit costs

The timeframe for the cost-benefit analysis was too stringent to allow observational assessments of the unit cost for data collection; i.e. it was not feasible to test how much time a data collector would spend to collect and fill in the data. The unit costs are also believed to be dependent on which data collection technology is used. Therefore have qualified estimations on the unit costs been made instead. The estimated unit costs for each data class are shown in Table 38, Appendix D2.

#### 8.1.5.3 Base costs

The base costs, or the costs for purchasing, setting up and maintaining the tools have been obtained based on purchasing cost data from the information and communication department of VMW, and qualified estimations. Some rules for calculation have been:

---

[13] Based on average number of new assets in the period 2005-2010

- VMW East Flanders supply 17.7 % of the customers and has approximately 16 % of the pipe length in VMW. For tools that apply in the whole of VMW, the purchase cost has been set to 17 % in order to estimate the cost for VMW East Flanders alone.
- The different tools are assumed to be used for 10-15 years.
- The total cost of one work-hour is set to 50 €. One work week is set to 40 hours.
- It is assumed that one third (1/3) of the costs invested in Octopus can be related to data collection benefits, since Octopus also has other applications within VMW.

The tool costs are summarised in Table 24.

Table 24: Tool costs summary

| Tool | Acquiring cost [€/year] | Setup cost [weeks/year] | Maintenance cost [weeks/year] | Base cost [weeks/year] |
|---|---|---|---|---|
| GIS inventory | € 11 798.56 | - | 40.00 | **45.90** |
| Hydraulic model | € 1 179.86 | - | 40.00 | **40.59** |
| USTORE | € 13 273.38 | 4.00 | 4.00 | **14.64** |
| LCC-AM/QM | € 4 601.44 | 4.00 | 22.00 | **28.30** |
| KANEW LTP | € - |  | 3.00 | **3.00** |
| Aware PLAN | € - | 10.00 | 10.00 | **20.00** |
| Aware-P PX | € - | 2.00 | 2.00 | **4.00** |
| FAIL | € - | 4.00 | 2.00 | **6.00** |
| LLIFE | € - | 4.00 | 2.00 | **6.00** |
| CIMP | € - | 2.00 | 1.00 | **3.00** |
| UNMET | € - | 4.00 | 1.00 | **5.00** |
| IVI | € - | 2.00 | 1.00 | **3.00** |
| Diary/Octopus | € 12 585.13 | 20.00 | 10.00 | **36.29** |

#### *8.1.5.4 Weighting factors*

Weighting factors have been selected according to subjective judgements. The following rationales have been used to weight the information items:

- All “strategic” information items (the inventory and all DMA level information) are given a weight of 100 %
- Information items on a cohort level are given a weight of 50 %
- Information items on an asset level are given a weight of 30 %. An exception is the two risk dimensions (hydraulic and economic), which are actually a composite of cohort and asset level information – these items are given a weight of 50%.

The same rules apply for informational benefits. By applying these rationales one ensures that information that applies at a higher, more general level, are given more emphasis than information that applies at a low level, thus reflecting that information that apply for the whole portfolio of assets is more useful (in a decision-making process) than information that only apply to individual assets.

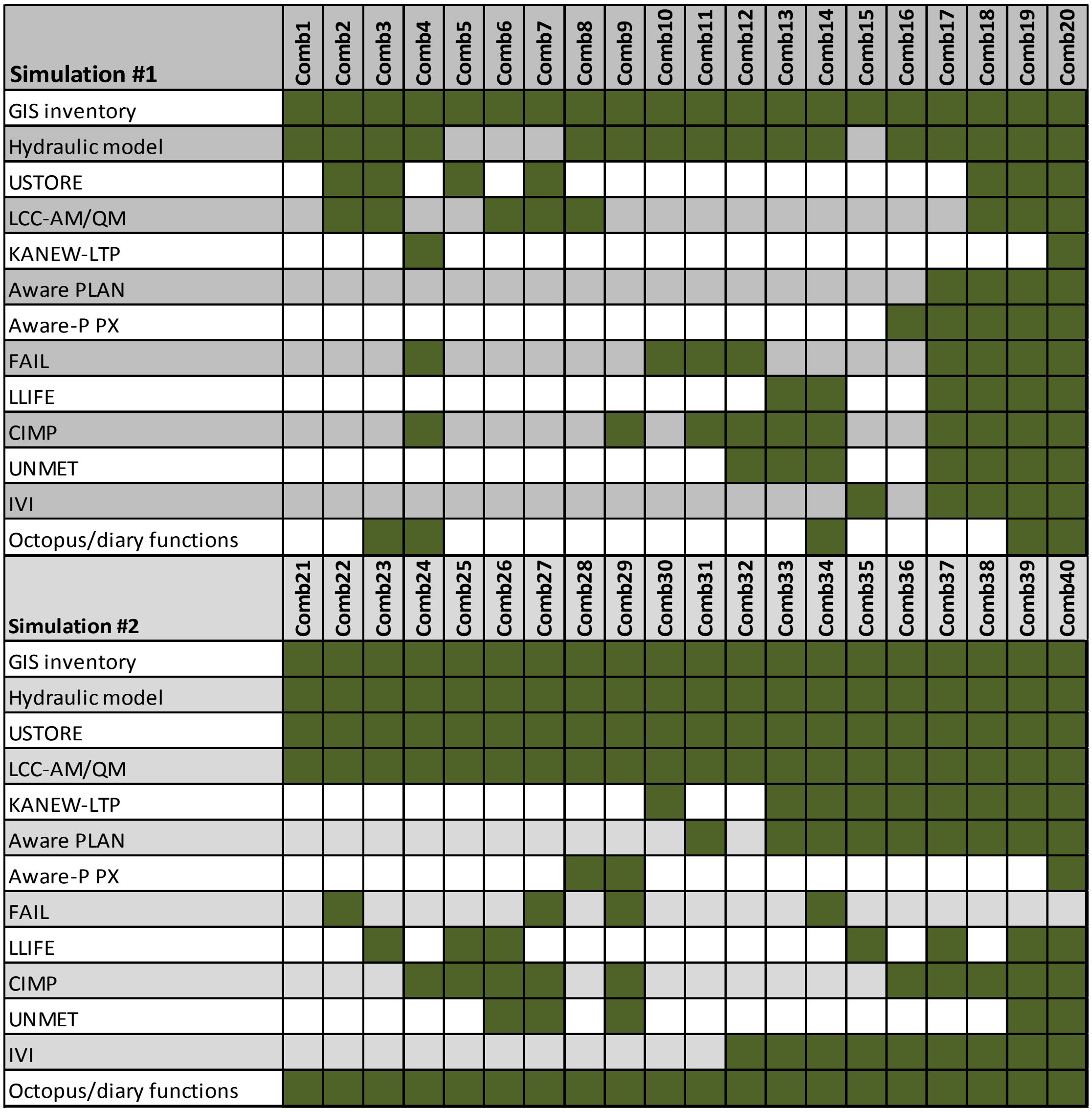
For the calculation of the **Total information completeness**, the information items are weighted 100 %, and the informational benefits and the planning strategies are weighted 50 %.

The weighting factors used are shown in Table 37, Appendix D1.

## 8.2 Model results

### 8.2.1 Tool consideration simulations, results and comments

**Table 25: Tool combinations selected for analysis in ToolComparison (dark colour indicates *yes*)**



| **Simulation #1** | **Comb1** | **Comb2** | **Comb3** | **Comb4** | **Comb5** | **Comb6** | **Comb7** | **Comb8** | **Comb9** | **Comb10** | **Comb11** | **Comb12** | **Comb13** | **Comb14** | **Comb15** | **Comb16** | **Comb17** | **Comb18** | **Comb19** | **Comb20** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GIS inventory | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Hydraulic model | ■ | ■ | ■ | ■ |  |  |  | ■ | ■ | ■ | ■ | ■ | ■ | ■ |  | ■ | ■ | ■ | ■ | ■ |
| USTORE |  | ■ | ■ |  | ■ |  | ■ |  |  |  |  |  |  |  |  |  |  | ■ | ■ | ■ |
| LCC-AM/QM |  | ■ | ■ |  |  | ■ | ■ | ■ |  |  |  |  |  |  |  |  |  | ■ | ■ | ■ |
| KANEW-LTP |  |  |  | ■ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ■ |
| Aware PLAN |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ■ | ■ | ■ | ■ |
| Aware-P PX |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ■ | ■ | ■ | ■ | ■ |
| FAIL |  |  |  | ■ |  |  |  |  |  | ■ | ■ | ■ |  |  |  |  | ■ | ■ | ■ | ■ |
| LLIFE |  |  |  |  |  |  |  |  |  |  |  |  | ■ | ■ |  |  | ■ | ■ | ■ | ■ |
| CIMP |  |  |  | ■ |  |  |  |  | ■ |  | ■ | ■ | ■ | ■ |  |  | ■ | ■ | ■ | ■ |
| UNMET |  |  |  |  |  |  |  |  |  |  |  | ■ | ■ | ■ |  |  | ■ | ■ | ■ | ■ |
| IVI |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ■ |  | ■ | ■ | ■ | ■ |
| Octopus/diary functions |  |  | ■ | ■ |  |  |  |  |  |  |  |  |  | ■ |  |  |  |  | ■ | ■ |
| **Simulation #2** | **Comb21** | **Comb22** | **Comb23** | **Comb24** | **Comb25** | **Comb26** | **Comb27** | **Comb28** | **Comb29** | **Comb30** | **Comb31** | **Comb32** | **Comb33** | **Comb34** | **Comb35** | **Comb36** | **Comb37** | **Comb38** | **Comb39** | **Comb40** |
| GIS inventory | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Hydraulic model | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| USTORE | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| LCC-AM/QM | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| KANEW-LTP |  |  |  |  |  |  |  |  |  | ■ |  |  | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Aware PLAN |  |  |  |  |  |  |  |  |  |  | ■ |  | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Aware-P PX |  |  |  |  |  |  |  | ■ | ■ |  |  |  |  |  |  |  |  |  |  | ■ |
| FAIL |  | ■ |  |  |  |  | ■ |  | ■ |  |  |  |  | ■ |  |  |  |  |  |  |
| LLIFE |  |  | ■ |  | ■ | ■ |  |  |  |  |  |  |  |  | ■ |  | ■ |  | ■ | ■ |
| CIMP |  |  |  | ■ | ■ | ■ | ■ |  | ■ |  |  |  |  |  |  | ■ | ■ | ■ | ■ | ■ |
| UNMET |  |  |  |  |  | ■ | ■ |  | ■ |  |  |  |  |  |  |  |  |  | ■ | ■ |
| IVI |  |  |  |  |  |  |  |  |  |  |  | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Octopus/diary functions | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |

The ToolComparison sheet analysis was separated into two simulations, each with 20 tool combinations, as shown in Table 25. The combinations of tools in the first simulation were selected in such a way that the analysis would yield cost-benefit results from a wide range of different data collection and utilisation strategies. In the second simulation the objective was to evaluate different combinations of tool additions, added to the long-term tool combination selected in VMW (Comb3 and 21); in order to show what additions to the current plans in VMW would be most beneficial.

From Table 25 one may see that Comb1 is the current situation in VMW, Comb2 is the near-future situation in VMW (within one year), and Comb3 is the far-future situation in VMW (implementation of Octopus). Comb4 is a situation similar to the one of *Trondheim Vann og Avløp*.

#### *8.2.1.1 Simulation #1*

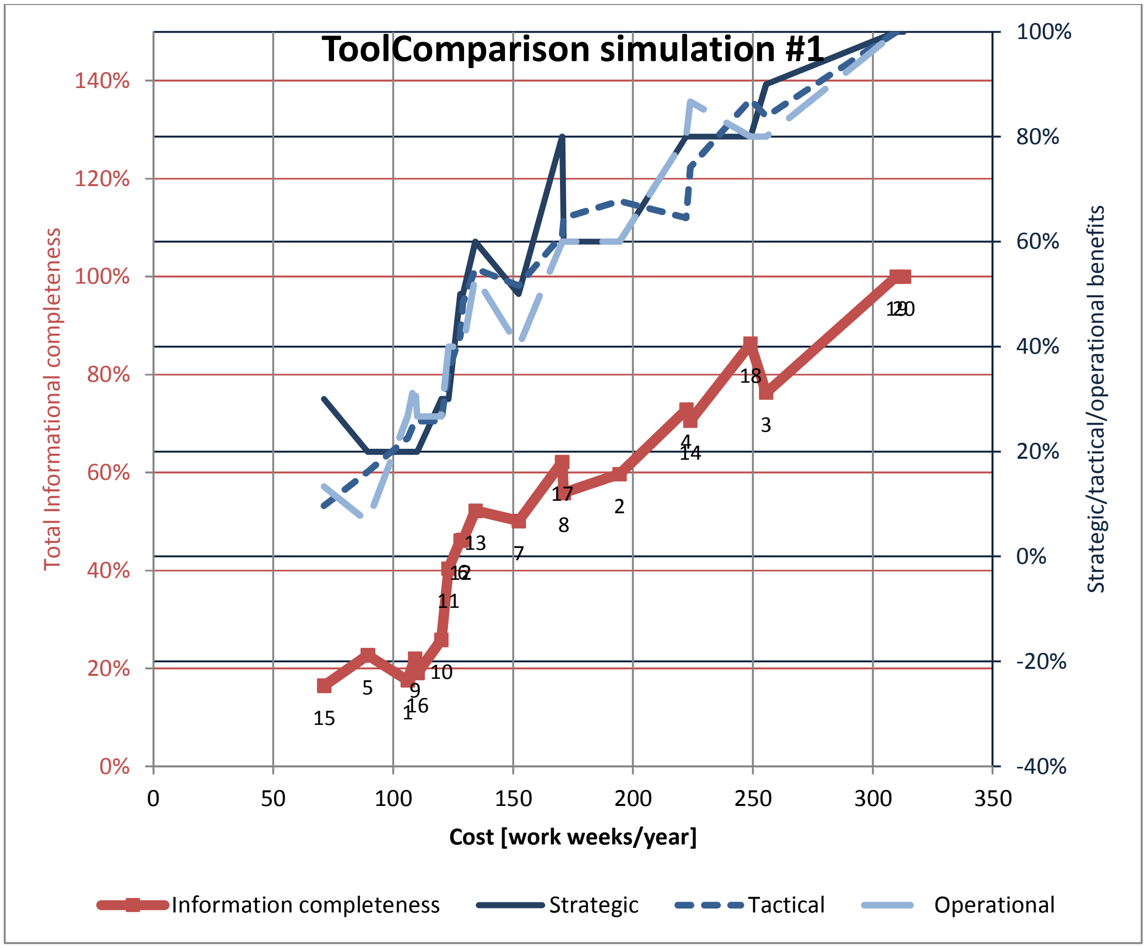


Figure 21: Results from simulation #1 (the labels correspond to the numbers in Table 25)

**Comments:** In Figure 21 the results from the first cost-benefit consideration are shown. The current situation in VMW (Comb1: GIS inventory and hydraulic model) is a low benefit situation, but still with a quite significant cost (106 weeks/year yields 18 % benefits). This is due to the fact that the hydraulic model and the GIS inventory both have quite high base costs compared to the other tools, while yielding virtually no information about the “health” of the assets. In this situation, there is an underrepresentation of information that is useful on a strategic level. When looking at Figure 21 it appears that VMW is currently standing at “the foot of the mountain” – substantial efforts need to be made to get a high level of informational outcomes.

Adding USTORE and LCC-AM/QM to the tool portfolio (Comb2) requires almost double the investment compared to the current situation (194 weeks/year), but yields an over three times higher benefit (60 % total informational completeness). With the LCC-AM/QM software implemented, are a lot of the cost- and investment-related information items rendered achievable. The information is most useful on a tactical level in this situation, since LCC-AM/QM and USTORE introduce reliability-related information items.

Adding Octopus to the tool portfolio (Comb3) will represent an increased investment from Comb2, at the same time increasing the total informational completeness from 60 % to 76 %. Information on a strategic level is better represented than on a tactical level, which again is better represented than information on an operational level in this situation. This is mostly due to the fact that Octopus and diary functions will ensure that strategic issues such as *complaints* and *level of service* can be properly monitored. With Comb3, VMW is not at "the top of the mountain", but yet in a much better position than the current.

As for the situation of the case study area, *Trondheim Vann og avløp* (Comb4), one may see that Trondheim has a quite high level of both cost and benefits compared to status quo in VMW. The most prominent impeding factor for the informational completeness in Trondheim seems to be the cost-related information that can be obtained through software like the LCC-AM/QM (which ultimately allows the utility to plan based on LCC-considerations).

The steep increase in benefits registered from Comb1 to Comb13 can be explained by the introduction of the different combinations of the FAIL, LLIFE, CIMP and UNMET tools – these tools have small data costs when a hydraulic model already is implemented, but each contribute a lot to the benefits. Implementing these tools, or similar tools, gives a "boost" to the amount of information one achieves, without requiring much extra data collection.

Some important **conclusions** from simulation #1 are:

1. The implementation of USTORE and LCC-AM/QM will double the costs, but will at the same time triple the benefits, compared to status quo.
2. The implementation of Octopus also represents a major increase in investments and benefits, but the range of information will still not be complete: criticality- and risk-based decision-rationales cannot be achieved with this tool combination.
3. It is suggested to investigate which combination of added tools that will give the best benefit increase, compared to the situation with a GIS inventory, hydraulic model, LCC-AM/QM, USTORE and Octopus.

It is point 3 that is the basis for simulation #2.

#### *8.2.1.2 Simulation #2*

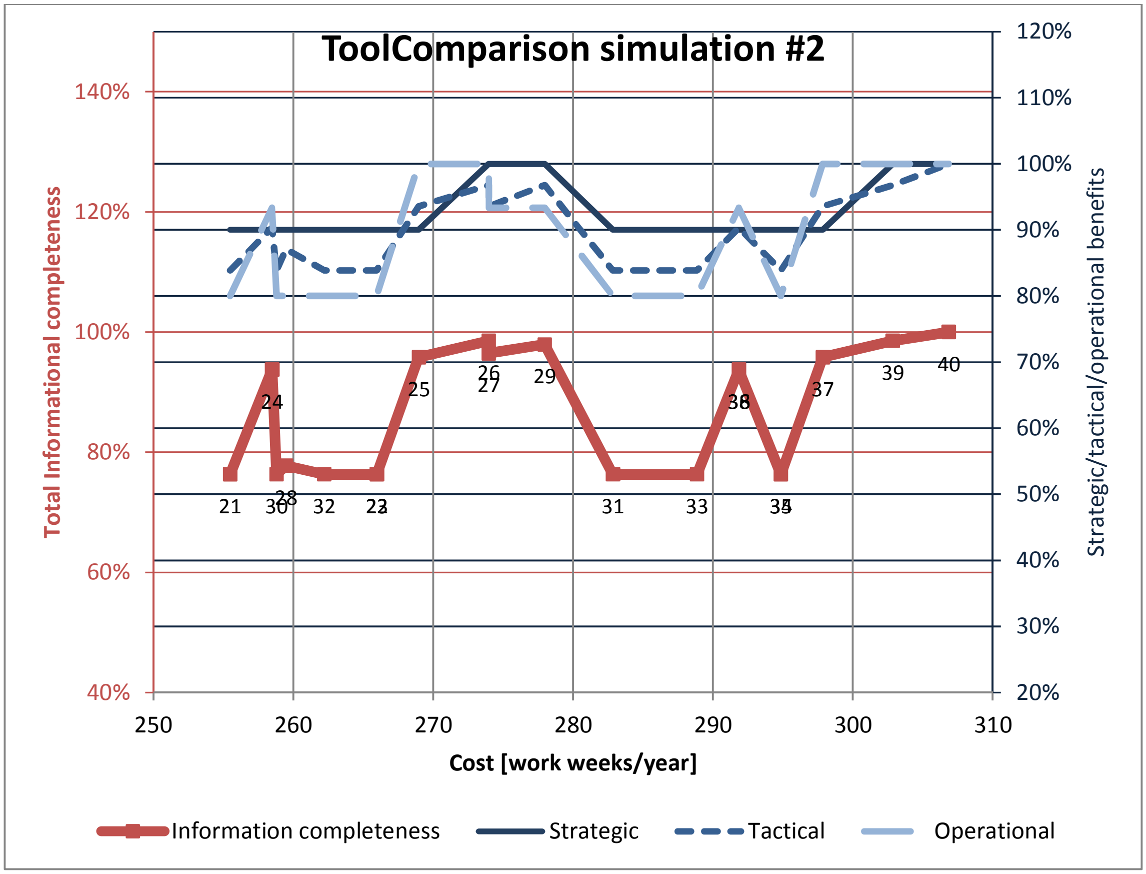


**Figure 22: Results from simulation #2 (the labels correspond to the numbers in Table 25)**

**Comments:** The results from simulation #2 are shown in Figure 22. This simulation has a much narrower span of results than simulation #1, both along the cost and benefit axes. The analysis shows that extra investments in data collection, compared to the long-term plan in VMW (Comb21), does not necessarily yield extra benefits.

Compared to the situation with Comb21 (long term plan in VMW) is it still possible to achieve higher payback, without excessive increase in the invested time to collect data. Some of the good "peak alternatives" identified are:

- Comb24, which is realised by adding the CIMP (component importance) considerations to the tool combination. This requires a 1.2 % cost increase and gives 18 % points benefit increase. This combination has most value on an operational level.
- Comb25, which is realised by adding CIMP and LLIFE. This requires a 5.5 % cost increase and gives a 20 % point benefit increase. Also this combination has most value on operational level.
- Comb26, which is realised by adding CIMP, LLIFE and UNMET. Requires a 7.5 % cost increase and gives a 23 % points benefit increase. This alternative has 100 % score on the strategic and operational level, and a 97 % score on the tactical level.

It seems like a very sound option for VMW to keep the plans of implementing USTORE, LCC-AM/QM and Octopus, but at the same time also implement tools for calculating the hydraulic importance of each component (CIMP), models for calculating economic risk of failure (LLIFE) and the impact this has on service continuity (UNMET). The extra investments (7.5 %) will improve the information situation drastically, and allow decision-making based on both criticality and hydraulic risk considerations, which is not possible with LCC-AM/QM as the main analytical tool.

The conclusion from these two simulations is consistent with the recommendations made in section 5, where concerns were raised about lacking measures for hydraulic criticality and hydraulic risk measures. Criticality measures can be produced in a short time span, and can hence be used as an early decision rationale, which in turn is beneficial for the organisational perspective of data collection.

### 8.2.2 Data class consideration results

In section 8.2.1, the question of which tool combination is most beneficial was discussed. Now the question remains: What data classes are the most useful (has the highest cost-benefit ratio)? The results from the DataComparison sheet are illustrated in Figure 23, where the benefit-cost ratio for each data class is graphed for each planning level, and sorted descending. This graph is based on data costs only, and does hence not include costs of tools – the idea here is to make considerations that are more independent of the tool combination costs.

As one may see, it is in general the *basic* data that yields the highest benefit-cost ratio. Even though the basic data classes have quite high data costs, they are still the data classes that are most widely used. The *complaint basic* class ranges highest because it is strategic information that, if organised right, will have a low cost and a high impact on the strategic level.

It is interesting to see that the *inspection basic* class has a quite low benefit-cost ratio. This is due to the fact that very few current tools actually can make use of performance and condition evaluations – it is only LCC-AM/QM that takes inspection data into account. It seems that there is a mismatch between high resources invested in inspection data collection, and low benefits in terms of decision incentives from the inspection data. Condition-related information is also the “most expensive information”, according to the model. In VMW are most inspection data leakage detection results – leakage data is usually not utilised to extrapolate information from an asset level to a cohort level. The development and implementation of deterioration models that take inspection data into account could increase the benefits from such data.

Data about disturbances (duration and number of clients affected) generally score low. However, repair disturbance data score higher, because this data can be used in a wider context (unexpected unmet demand calculations).

Further, it is noted that extended data classes in general score low. Although extended data classes can be used, they are seldom a requirement for modelling tools. For instance *can* extended external factors about a failure incident (e.g. traffic density) prove to be very useful in a failure model, yet it not said that it *is* useful. The usefulness can only be proved by testing.

There seems to be a balance between the strategic, tactical and operational usefulness of the different data classes. The exceptions are the *complaint basic* and the *decommissioning extended* classes. The former has a high strategic impact because it reflects customer satisfaction; the latter because the extended decommissioning class contains the reasons for decommissioning, which reflects the strategic issue of *why* assets are replaced.

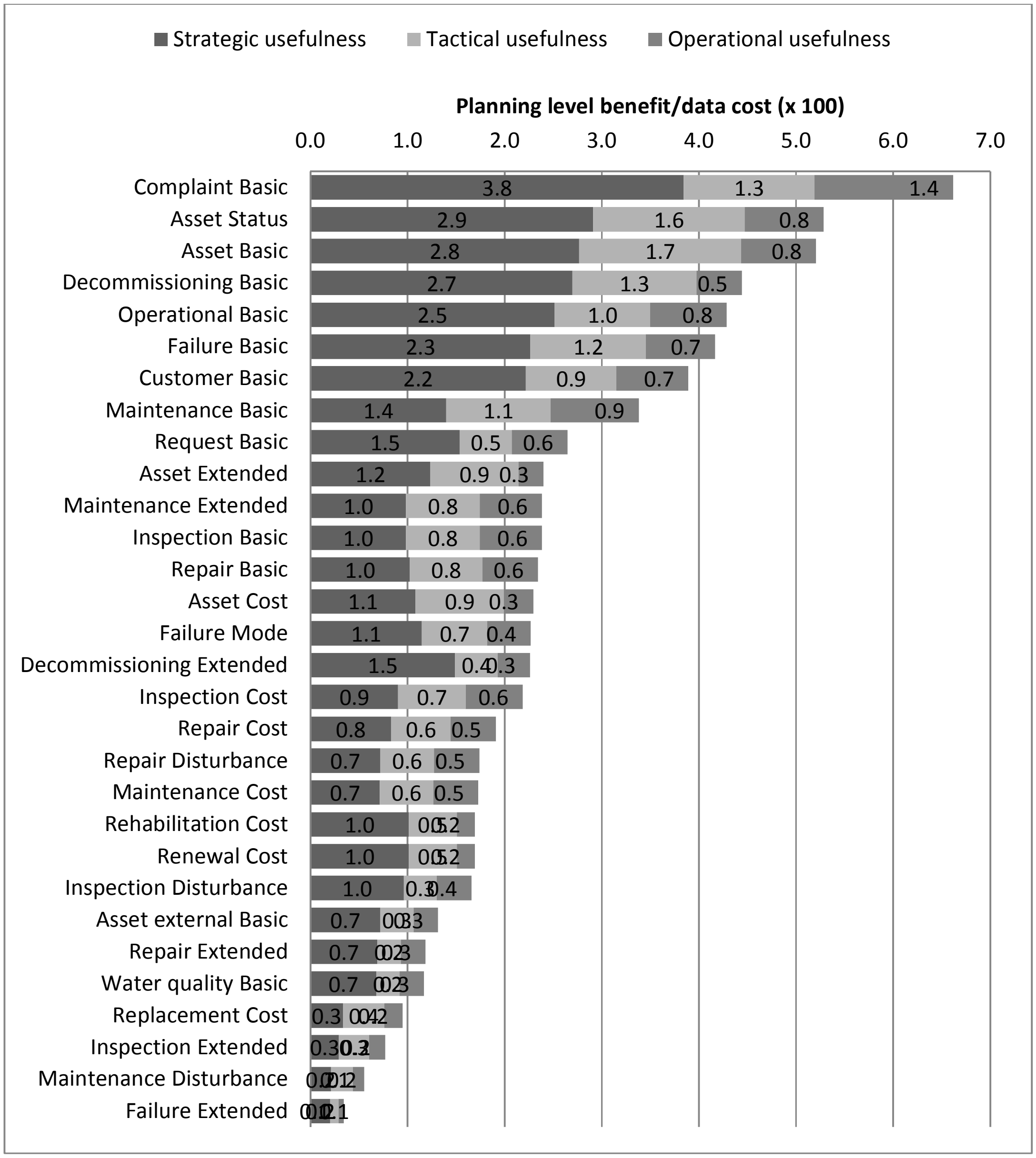


Figure 23: Results from DataComparison sheet

**Recommendations:**

- The efforts invested in collecting and storing inspection data should be considered carefully. It is suggested to search for or develop tools that can increase the usefulness of inspection data.
- Disturbance data is most useful for repair incidents, and should be prioritised before inspection and maintenance disturbance data.

- The usefulness of extended data classes should be tested before being collected in full scale. It is suggested that different types of extended asset data should be collected in different service centres in VMW in a test period; the most useful data can then be identified by comparing model results from the service centres, and form the recommendation for which data should be collected at full scale.

## 8.3 Model sensitivity and uncertainty

### 8.3.1 Sensitivity

The sensitivity of the model was assessed with respect to the following input parameters:

- Weighting factors
- Data unit costs
- Tool costs

Table 26: Sensitivity analysis results (simulation #1)

| Input parameter | Method | Output parameter | Maximum change | Average change | Most sensitive input |
|---|---|---|---|---|---|
| **Weighting factors** | Weighting factors versus no weighting factors | Rank | 4 | 1.45 | Relative rank weights |
| | | Total information completeness | 7.0 % | 3.9 % | |
| | | Benefit/cost-ratio | 27.3 % | 10.8 % | |
| **Data unit costs** | 25 % increase on each data class | Rank | 1 | 0.15 | Data classes with many entries per year |
| | | Total cost | 3.2 % | 1.6 % | |
| | | Benefit/cost-ratio | 20 % | 0.8 % | |
| **Tool costs** | 25 % increase in base cost for each tool | Total cost | 18 % | 9 % | High acquiring cost tools |
| | | Rank | 6 | 1 | |

The results from the sensitivity analysis for simulation #1 are summarised in Table 26. Sensitivity results for simulation #2 are less severe than for simulation #1, due to a more uniform tool selection (see Appendix D4). The weighting factors have a high impact on the different output parameters. This is a positive finding, because it shows that the model is able to convey a cost-benefit consideration based on an array of different priorities that the utility may have; the model user can express these priorities through the weighting factors.

The model is not alarmingly sensitive for a 25 % increase for each data class; the change in each individual data class will maximally change the ranking of a tool combination with *1* place. However, for the base tool costs, can the rank of a tool combination change as much as *6* places, which is a large difference when considering only 20 combinations. It is therefore considered as more important to achieve an accurate assessment of what are the base costs (acquiring, set-up and maintenance) of each tool, than to assess unit costs for the data classes accurately. The high impact on the sensitivity in base costs is mostly due to the high base costs of the inventory and the hydraulic model.

Detailed figures from the sensitivity analysis are shown in Appendix D.

### 8.3.2 Uncertainty

The uncertainty of the model results is difficult to assess, since the model is not yet based on observed unit cost data. However, the uncertainty of the output can most certainly be assessed if observed unit costs (with variances) are included in the model at a later stage. For now can the uncertainty measures be based on the minimum and maximum data costs (see section 8.1.2): The minimum and maximum data costs make out the uncertainty limits. The relative uncertainty measures are calculated according to equation (8.5):

$$\text{Relative uncertainty } [\%] = \frac{C_{max} - C_{min}}{C_{average}} \tag{8.5}$$

The uncertainty is 4.0 % on average. The tools that contribute most to both the relative and the absolute uncertainty are the Diary/Octopus functions, and the failure models, due to the fact that these tools have many optional data that are quite costly.

The whole list of results is enclosed in Appendix D5.

## 8.4 Conclusions, evaluations and suggestions for further development

### 8.4.1 Conclusions

The main findings of the tool considerations are summarised in Figure 24.

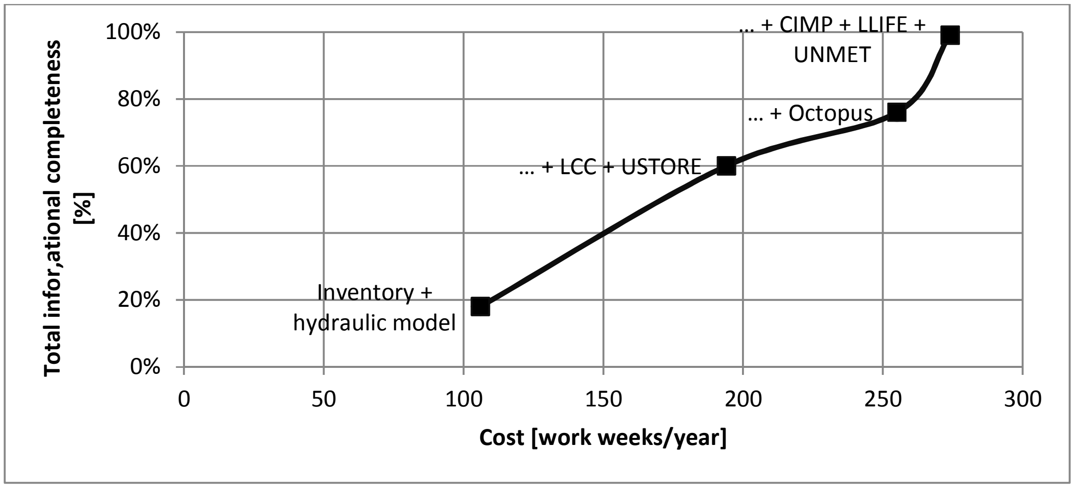


Figure 24: Four alternatives for data utilisation in VMW

The cost-benefit considerations carried out in this section has given some useful insights for the data collection and utilisation:

- By investing 7.5 % more on data collection and treatment could VMW increase the accumulated amount of information drastically, compared to their current long-term plans for data utilisation, by introducing tools that consider component importance (CIMP), economical risk of failure (LLIFE) and calculate expected unmet demands (UNMET). These tools allow the utility to assess hydraulic and economic risk, as well as keeping track of the level of service through measuring the average *unexpected unmet demand.*
- VMW should search for innovative solutions that could increase the utilisation degree of inspection (leakage) data.
- Collection of *extended* data classes should be considered carefully. It is suggested that different service centres collect different extended incident data in a test period, which can

be compared in order to make an informed decision about what extended data is most useful.

### 8.4.2 Evaluation and suggestions for further developments

The methodology shown for the cost-benefit assessment suggested in this section represents a step in the direction of formalisation and rationalisation of the evaluation process that should always be present before selecting AM tools in a utility. The spread sheet model that has been developed is a suitable tool for evaluating and comparing a limited set of tools in a limited set of combinations.

Through the example study of VMW it is shown how such a model can be used to interpret the costs and benefits associated with different data collection and utilisation scenarios:

- Firstly, a utility can use the model to assess its current position with respect to costs and benefits.
- Secondly, a utility can use the model to evaluate which tool combinations that will be most suitable for the utility to invest in (optimal cost-benefit points).
- Thirdly, the model can be used to identify data classes that yield low benefits compared with the investments associated with them; these results can again be used to assess where efforts should be made to utilise data better.

The results show that it is indeed possible to assess costs and benefits of data collection within a systematic framework, and further that such assessments may be useful for the utilities.

The model already has basic capabilities, but the unit data costs still need to be assessed through measurements. However, the sensitivity analysis shows that the most prominent factor for costs are the base cost of the tools, thus is it more critical to determine the base costs then the data unit costs. If the utility has good estimates for the base costs and fairly good estimates for the unit costs, will the cost-benefit considerations still be useful.

The model is unfortunately not “all-knowing”. For instance, the model does not describe the added benefit of combining the utility’s data with data from other utilities (the USTORE principle). Further, the model says nothing about which data utilisation configuration will be best with respect to the personal and organisational side of data collection (chapter 6); the model should be used as one element of several in the data collection planning process – organisational factors must also be considered. Another issue that has not been considered in the model is temporal changes. The model works well to predict data costs for a situation similar to status quo, but fails to account for changing conditions in the utility – i.e. changes in failure rates, maintenance intensity etc. If the way of managing the utility changes the costs and benefits of data will also change.

In the beginning of this chapter it was stated that the benefit of data ultimately could be expressed as the avoided losses from unforeseen failures and premature replacements. Developing the model further, so that it expresses the benefit as the reduced economic risk of failures and gained capital from extended service lives, would be a significant step forward for assessing the true monetary value of data collection. A first step in this development could be to assess the reduced economic risk of failure. This could be done by a case-study in a utility, where the effect of realised renewal projects, expressed as reduced economic risk of failure, could be calculated under different decision rationales. The difference in the reduced risk between the different rationales would reflect the economic value of the data that is needed to support the different decision rationales.

Suggestions for other cost-benefit model developments are:

- Better assessment (measurements) of the unit costs for the different data classes
- Including a wider array of different AM tools
- Including temporal change considerations
- Including automatic sensitivity testing in the model

For the improvement of data utilisation is it suggested that the value of inspection data should be increased by developing and implementing models that utilise inspection data in the same way failure data is utilised today: by extrapolating results to the future (forecasting) and by extrapolating from asset to cohort level.

# 9 References

ALEGRE, H. & COVAS, D. 2010. Quick reference guide for infrastructure asset management of water supply and wastewater system. *In:* AWAREP (ed.). LNEC.

ALEGRE, H., COVAS, D., MONTEIRO, A. J. & DUARTE, P. Water Infrastructure Asset Management: A Methodology to Define Investment Prioritization. *In:* BUCHBERGER, S. G., ed., 2006a Cincinnati, Ohio, USA. ASCE, 29-29.

ALEGRE, H., HIRNER, W., BAPTISTA, J. M., PARENA, R., CUBILLO, F., CABRERA, E. J., MERKEL, W. & DUARTE, P. 2006b. *Performance indicators for water supply services,* Alliance House 12, Caxton Street, London, IWA Publishing.

ALEGRE, H. & MATOS, R. 2009. Paving the way for a sustainable asset management of urban water infrastructures: outcomes of the 5th World Water Forum. *Water asset management International,* 5**,** 3.

ALEGRE, H. M., RAFAELA 2009. Paving the way for a sustainable asset management of urban water infrastructures: outcomes of the 5th World Water Forum. *Water asset management International,* 5**,** 3.

AMADI-ECHENDU, J. E., WILLETT, R., BROWN, K., HOPE, T., LEE, J., MATHEW, J., VYAS, N. & YANG, B.-S. 2010. What Is Engineering Asset Management? Definitions, Concepts and Scope of Engineering Asset Management. *In:* AMADI-ECHENDU, J. E., BROWN, K., WILLETT, R. & MATHEW, J. (eds.). Springer London.

ANDRALANOV, A. 2012. *MIKE NET and RELNet: which approach to reliability analysis is better?* [Online]. Water and Environmental Engineering, Department of Chemical Engineering, Lund University. Available: http://www.vateknik.lth.se/exjobb/E315.pdf [Accessed March 26th 2012].

ARTS, D. G. T., DE KEIZER, N. F. & SCHEFFER, G.-J. 2002. Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. *Journal of the American Medical Informatics Association,* 9**,** 600-611.

AVEN, T. & RENN, O. 2010. *Risk Management and Governance*, Springer.

AVEN, T. R., O. 2010. *Risk Management and Governance*, Springer.

AWARE-P.ORG. 2011. *AWARE-P: About the project* [Online]. aware-p.org: AWARE-P. Available: http://www.aware-p.org/np4/home [Accessed March 21st 2012].

BABYKINA, G. & COUALLIER, V. 2010. Modelling Recurrent Events for Repairable Systems Under Worse Than Old Assumption - Advances in Degradation Modeling. *In:* NIKULIN, M. S., LIMNIOS, N., BALAKRISHNAN, N., KAHLE, W. & HUBER-CAROL, C. (eds.). Birkhäuser Boston.

BASEFORM.ORG. 2012a. *Aware-P asset management: Details* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/apps/awareAppDetails.html [Accessed March 21st 2012].

BASEFORM.ORG. 2012b. *baseform Core* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/tools/core.html [Accessed March 25th 2012].

BASEFORM.ORG. 2012c. *baseform tools* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/tools/ [Accessed March 25th 2012].

BASEFORM.ORG. 2012d. *EPANET - Hydaulic and water quality network simulation* [Online]. baseform.org. Available: http://www.baseform.org/np4/tools/epanetTool.html [Accessed March 26th 2012].

BASEFORM.ORG. 2012e. *IVI - Infrastructure Value Index* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/iviTool.html [Accessed March 26th 2012].

BASEFORM.ORG. 2012f. *PI - performance indicators* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/tools/piTool.html [Accessed March 26th 2012].

BASEFORM.ORG. 2012g. *PLAN - AWARE-P infrastructure asset management planning* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/tools/planTool.html [Accessed March 26th 2012].

BASEFORM.ORG. 2012h. *PX - performance indices* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/tools/pxTool.html [Accessed March 26th 2012].

BASEFORM.ORG. 2012i. *UNMET - Pipe service interruption risk expressed as expected volume of unmet demand* [Online]. baseform.org: baseform.org. Available: http://www.baseform.org/np4/unmetTool.html [Accessed March 26th 2012].

BAUR, R. & HERZ, R. K. Service Life Management of Water Mains and Sewers. 13th European Junior Scientist Workshop, 1999 Dresden.

BHAGWAN, J. N. 2009. *Compendium of best practices in water infrastructure asset management*, Glogal Water Research Coalition.

BROO, A. E., BERGHULT, B. & HEDBERG, T. 1997. Copper corrosion in drinking water distribution systems — the influence of water quality. *Corrosion Science,* 39**,** 1119-1132.

BRUASET, S., SELSETH, I. & HAFSKJOLD, L. S. 2010. Statistisk analyse av brudd på vannledningsnettet i Trondheim og feilprediksjon med programmet Casses. *In:* MILJØ, S. B. V. O. (ed.). Trondheim: Sintef Byggforsk Vann og miljø.

BURN, S., MARLOW, D., TRAN, D. & MASAHIRO, F. 2010. Modelling asset lifetimes and their role in asset management. *Journal of water supply : research and technology. AQUA (Print),* 59**,** 362-377.

CEMAGREF 2010. LEYP model for recurrent failures of water mains. 50 avenue de Verdun, 33612 Cestas Cedex: Cemagref.

CEMAGREF.FR. 2011. *Presentation of the software* [Online]. Available: https://casses.cemagref.fr/presentation-of-the-software [Accessed November 2nd 2011].

COELHO, S. T. & VITORINO, D. 2011. AWARE-P: a collaborative, system-based IAM planning software. *LESAM - Leading Edge Strategic Asset Management 2011.* Mülheim, Germany: International Water Association.

CONSTANTINE, A. G. D., J. N. 1993. Pipeline Relibability: Stochastic models in engineering technology and management. *World Scientific*.

CROMWELL, J. E., III & SPERANZA, E. 2007. Asset management too complicated? Just think about your car. *American Water Works Association. Journal,* 99**,** 46-51,10.

D'ÁGUA, M. R., COVAS, D. & ALEGRE, H. 2007. *The Importance of Rehabilitation Planning for the Waste Management in Water Supply Systems* [Online]. iswa.org. Available: http://www.iswa.org/uploads/tx_iswaknowledgebase/6-372paper_long.pdf [Accessed November 9th 2011].

DANCKAERS, G. April 18th 2012. *RE: FW: Waiting for Octopus.* Type to ROKSTAD, M. M.

DANCKAERS, G. & HAMMENECKER, J. 2012. Studie mobiele data-uitwisseling. Conclusies eindrapport en voorstel verdere aanpak. *In:* VMW (ed.). VMW: Vlaamse Maatschappij voor Watervoorziening.

DASH SOFTWARE AS. 2011. *Programvare for Vedlikehold/FDV og Service samt Logistikk og Produksjon (MPS/ERP)* [Online]. dashsoftware.no: dashsoftware.no. Available: http://www.dashsoftware.no/no/forside/ [Accessed April 12th 2011].

DAVIS, P., BURN, S., MOGLIA, M. & GOULD, S. 2007. A physical probabilistic model to predict failure rates in buried PVC pipelines. *Reliability Engineering &amp; System Safety,* 92**,** 1258-1266.

DEMPSEY, C. 2010. *Free data viewers* [Online]. gislounge.com: GIS Lounge. Available: http://gislounge.com/free-data-viewers/ [Accessed February 11th 2012].

DEMPSEY, C. 2011. *Web-based GIS* [Online]. gislounge.com: GIS Lounge. Available: http://gislounge.com/web-based-gis/ [Accessed February 15th 2012].

DHI. 2011. *MIKE URBAN Water Distribution* [Online]. mikebydhi.com,: mikebydhi.com,. Available: http://www.mikebydhi.com/Products/Cities/MIKEURBAN/WaterDistribution.aspx [Accessed April 16th 2012].

DIERICX, E. 2012. Inventaris informatie velden OCTOPUS -> AMT. VMW: VMW.

EIKSUND, S. & RELLING, S. Å. 2012. Befolkningsendringer 2011 - Tabell- og figursamling. *In:* KOMMUNE, B. I. T. (ed.). www.trondheim.kommune.no: Trondheim kommune.

ELLEFSEN, T. April 10th 2012. *RE: Svar fra Trond Ellefsen: Spørsmål til Trondheim kommune –case study for VMW.* Type to ROKSTAD, M. M.

ENGELHARDT, M. O., SKIPWORTH, P. J., SAVIC, D. A., SAUL, A. J. & WALTERS, G. A. 2000. Rehabilitation strategies for water distribution networks: a literature review with a UK perspective. *Urban Water,* 2**,** 153-170.

FHWA. 2001. *Data integration primer* [Online]. isddc.dot.gov: U.S. Department of Transportation, Federal Highway Administration, Office of Asset Management. Available: http://isddc.dot.gov/OLPFiles/FHWA/010393.pdf [Accessed January 19th 2012].

FRANGOPOL, D. M., LIN, K.-Y. & ESTES, A. C. 1997. Life-Cycle Cost Design of Deteriorating Structure. *Journal of Structural Engineering* 123**,** 12.

GOULTER, I. C. & KAZEMI, A. 1988. Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Canadian Journal of Civil Engineering,* 15**,** 91-97.

GRIGG, N. S. 2006. Condition Assessment of Water Distribution Pipes. *Journal of infrastructure systems,* 12**,** 147-153.

GUSTAFSON, J.-M. C., D. V. Modelling the occurence of breaks in cast iron water mains using methods of survival analysis. AWWA Annual Conference, 1999 Chicago.

HADZILACOS, T., KALLES, D., PRESTON, N., MELBOURNE, P., CAMARINOPOULOS, L., EIMERMACHER, M., KALLIDROMITIS, V., FRONDISTOU-YANNAS, S. & SAEGROV, S. 2000. UtilNets: a water mains rehabilitation decision-support system. *Computers, Environment and Urban Systems,* 24**,** 215-232.

HAFSKJOLD, L. S. & SELSETH, I. 2008. Saneringsplan Vann Trondheim, kjøring av modulene FAIL og REL i CARE-W. Trondheim: Sintef Byggforsk.

HALFAWY, M. R. 2008. Integration of Municipal Infrastructure Asset Management Processes: Challenges and Solutions. *Journal of Computing in Civil Engineering,* 22**,** 216-229.

HALFAWY, M. R. & FIGUEROA, R. Developing enterprise GIS-based data repositories for municipal infrastructure asset management. 2006-06-14 2006 Montreal. 10.

HALFAWY, M. R., NEWTON, L. A. & VANIER, D. J. 2006a. Review of commercial municipal infrastructure asset management systems. *Electronic Journal of Information Technology in Construction,* 11**,** 211-224.

HALFAWY, M. R., VANIER, D. & FROESE, T. M. 2006b. Standard data models for interoperability of municipal infrastructure asset management systems. *Canadian Journal of Civil Engineering,* 33**,** 1459-1469.

HASENACK, H. February 6th 2012. *RE: LCC-AMQM questions about data input/output.* Type to ROKSTAD, M. M.

HERZ, R. K. 1994. Alterung und Erneuerung von Infrastrukturbeständen – ein Kohortenüberlebensmodell. *Jahrbuch für Regionalwissenschaft,* 14**,** 14.

HERZ, R. K. 1996. Ageing process and rehabilitation needs of drinking water distribution networks. *Journal of Water SRT - Aqua,* 45**,** 11.

HOLTHE, K. 2010. *Use and Collection of data in Gemini VA in Asset Management.* Master of Technology, Norwegian University of Science and Technology.

HULPIAU, H. 2012. *RE: Fasering Octopus.* Type to ROKSTAD, M. M.

HØSEGGEN, R. S. April 10th 2012. *RE: SV: Spørsmål om data for strategisk planlegging i Trondheim vann og avløp.* Type to ROKSTAD, M. M.

IBM 1993. *IBM Dictionary of Computing*, McGraw-Hill, Inc.

INNOVYZE. 2012a. *InfoWater - Overview* [Online]. innowyze.com: Innovyze. Available: http://www.innovyze.com/products/infowater/ [Accessed February 16th 2012].

INNOVYZE. 2012b. *InfoWorks WS* [Online]. innovyze.com: Innovyze. Available: http://www.innovyze.com/products/infoworks_ws/ [Accessed February 16th 2012].

INNOVYZE. 2012c. *InfoWorks WS - Features and Benefits* [Online]. innovyze.com: Innovyze. Available: http://www.innovyze.com/products/infoworks_ws/features.aspx [Accessed February 16th 2012].

INNOVYZE. 2012d. *InfoWorks WS - Technical review* [Online]. innovyze.com: Innovyze. Available: http://www.innovyze.com/products/infoworks_ws/technical_review.aspx [Accessed February 16th 2012].

IRGC 2008. *An introduction to the IRGC Risk governance framework,* Geneva, International Risk Governance Council.

ITINERIS 2011a. Vergadering notule: Project Octopus. itineris.

ITINERIS 2011b. VMW - Octopus. itineris.

KARNEY, B., KHANI, D., HALFAWY, M. R. & HUNAIDI, O. A simulation study on using inverse transient analysis for leak detection in water distribution networks. 2008-02-21 2008 Toronto, Ontario. 1-21.

KELMAN, H. C. 1958. Compliance, Identification, and Internalization: Three Processes of Attitude Change. *The Journal of Conflict Resolution,* 2, 51-60.

KLEINER, Y., ADAMS, B. J. & ROGERS, J. S. 2001. Water distribution network renewal planning. *Journal of Computing in Civil Engineering,* 15, 15-26.

KLEINER, Y. & RAJANI, B. 2001. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water,* 3, 131-150.

KLEINER, Y. & RAJANI, B. B. 1999. Using limited data to assess future needs. *Journal (American Water Works Association),* 91, 47-61.

KONG, J. S. & FRANGOPOL, D. M. 2003. Life-Cycle Reliability-Based Maintenance Cost Optimization of Deteriorating Structures with Emphasis on Bridges. *Journal of Structural Engineering,* 129, 818-828.

KONG, J. S. & FRANGOPOL, D. M. 2004. Cost--Reliability Interaction in Life-Cycle Cost Optimization of Deteriorating Structures. *Journal of Structural Engineering,* 130, 1704-1712.

KYLE, B. R., VANIER, D. J., KOSOVAC, B. & FROESE, T. M. Information needs towards service life asset management. 2000-10-15 2000 Baveno, Italy. 1-18.

LE GAT, Y. 2008. Modelling the deterioration process of drainage pipelines. *Urban Water Journal,* 5, 10.

LE GAT, Y. 2009. *Une extension du processus de Yule pour la modélisation stochastique des événements récurrents.* PhD, Agro Paris Tech.

LEMER, A. C. Progress toward integrated infrastructure-assets-management systems: GIS and beyond. APWA International Public Works Congress: NRCC/CPWA Seminar Series “Innovations in Urban Infrastructure”, 1998 Las Vegas, Nv. 7-24.

LIN, S., GAO, J. & KORONIOS, A. The need for a data quality framework in asset management. 17th Australasian conference on information systems, December 6-8 2006 Adelaide, South Australia.

LIN, S., GAO, J., KORONIOS, A. & CHANANA, V. 2007. Developing a data quality framework for asset management in engineering organisations. *International Journal of Information Quality,* 1, 100-126.

LIU, Z., KLEINER, Y., RAJANI, B., WANG, L. & CONDIT, W. 2012. *Condition Assessment Technologies for Water Transmission and Distribution Systems,* Cincinnati, OH 45268, USEPA.

MIGLIACCIO, G. C. B., SUSAN M. & CORDOVA-ALVIDREZ, A. A. 2011. Continuous Quality Improvement Techniques for Data Collection in Asset Management Systems *Journal of Construction Engineering and Management*, 27.

MISIUNAS, D. 2003. *Burst Detection and Location in Pipelines and Pipe Networks - With Application in Water Distribution [Elektronisk resurs]*, Dept. of Industrial Electrical Engineering and Automation, Lund Institute of Technology.

MISIŪNAS, D. 2008. Failure Monitoring and Asset Condition Asssessment in Water Supply Systems. *The 7th International Conference “Environmental engineering”.* Vilniaus Gedimino technikos universitetas.

MISIUNAS, D., VITKOVSKY, J., OLSSON, G., SIMPSON, A. & LAMBERT, M. 2005. Pipeline Break Detection Using Pressure Transient Monitoring. *Journal of Water Resources Planning and Management,* 131, 316-325.

MORRISON, J. 2004. Managing leakage by District Metered Areas: a practical approach. *Water21,* 6, 3.

MURPHY, G. D. 2009. Improving the quality of manually acquired data: Applying the theory of planned behaviour to data quality. *Reliability Engineering &amp; System Safety,* 94, 1881-1886.

MURPHY, G. D., CHANG, A. & BARLOW, M. 2008. Leveraging engineering asset data : strategic priorities, data types and informational outcomes. *In:* NI, G. J. J. L. J., MA, L. & MATHEW, J. (eds.) *3rd World Congress on Engineering Asset Management and Intelligent Maintenance Systems Conference (WCEAM-IMS 2008): Engineering Asset Management – A Foundation for Sustainable Development.* Springer.

MÜLLER, K. & FISCHER, B. Objective Condition Assessment of Sewer Systems. LESAM 2007 - 2nd Leading Edge Conference on Strategic Asset Management, October 17-19 2007 Lisbon, Portugal. IWA Publishing.

PIPINO, L. L., LEE, Y. W. & WANG, R. Y. 2002. Data quality assessment. *Commun. ACM,* 45, 211-218.

POULTON, M., GAT, Y. L. & BRÉMOND, B. 2007. The impact of pipe segment length on break predictions in water distribution systems. *LESAM 2007 - 3nd leadning edge conforence on strategic asset management.* Lisbon, Portugal: IWA.

POWEL. 2011. *Gemini VA* [Online]. www.powelgemini.no: powelgemini.no. Available: http://www.powelgemini.no/no/Produkter/Gemini/gemini-va/ [Accessed December 5th 2011].

PROZZI, J. A. & HONG, F. 2010. Average Treatment Effect for Modeling Maintenance Work. *Journal of infrastructure systems,* 16, 122-128.

RAJANI, B. & KLEINER, Y. 2001. Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water,* 3, 151-164.

REDMAN, T. C. 1998. The impact of poor data quality on the typical enterprise. *Commun. ACM,* 41, 79-82.

ROGERS, P. D. & GRIGG, N. S. Failure Assessment Model to Prioritize Pipe Replacement in Water Utility Asset Management. *In:* BUCHBERGER, S. G., ed., 2006 Cincinnati, Ohio, USA. ASCE, 19-19.

ROKSTAD, M. M. 2011a. Statistical failure analysis of the VMW water distribution network in East Flanders. Trondheim: Norwegian University of Science and Technology.

ROKSTAD, M. M. 2011b. VMW GIS inventory database East Flanders. *In:* VMW (ed.) 2011 ed. VMW,: Danckaers, Gert.

ROLSCH AM. 2011a. *Demo-fil brandkraan applicatie Vitens NV* [Online]. www.rolsch.nl: Rolsch Asset Management. Available: http://www.rolsch.nl/video/401_01_Brandkraan_app_Vitens.mp4 [Accessed February 3rd 2012].

ROLSCH AM. 2011b. *MapKit voor onderhoudsmonteurs* [Online]. www.rolsch.nl: Rolsch Asset Management. Available: http://www.rolsch.nl/waterleiding5.html [Accessed 3rd February].

ROSSMAN, L. 2000. Epanet 2.0 User's Manual. Water Supply and Water Resources Division. *In:* NATIONAL RISK MANAGEMENT RESEARCH LABORATORY (ed.). U.S. Environmental Protection Agency.

RØSTUM, J. 2000. *Statistical modelling of pipe failures in water networks.* 2000:12, [Tapir].

S&G AND PARTNERS. 2009. *Quantiatative Maintenance Analysis* [Online]. sg-partners.nl: S&G and Partners. Available: http://www.sg-partners.nl/pdf/English%20brochures/Quantitative%20maintenance%20analyses.pdf [Accessed February 15th 2012].

S&G AND PARTNERS. 2012a. *About S&G* [Online]. http://www.sg-partners.nl/: S&G and Partners,. Available: http://www.sg-partners.nl/About_Us.html [Accessed February 1st 2012].

S&G AND PARTNERS. 2012b. *Import export* [Online]. http://www.sg-partners.nl/: S&G and Partners,. Available: http://www.sg-partners.nl/import_export.html [Accessed February 13th 2012].

SCHNEIDER, J., GAUL, A. J., NEUMANN, C., HOGRÄFER, J., WELLßOW, W., SCHWAN, M. & SCHNETTLER, A. 2006. Asset management techniques. *International Journal of Electrical Power &amp; Energy Systems,* 28, 643-654.

SELSETH, I. & RØSTUM, J. 2002. Langsiktig rehabiliteringsbehov for avløpsnettet i Trondheim kommune. *In:* VA, S. K. V. O. (ed.). Trondheim: Sintef Kjemi Vannrensing og VA.

SELSETH, I. & SÆGROV, S. 2001. Fornyelsesbehov for ledningsnettet i Trondheim - prognoser basert på KANEW. Trondheim: SINTEF Vannrensing og VA.

SJØVOLD, F. & SELSETH, I. 2009. Saneringsplan for vannforsyning for Trondheim kommune - forprosjekt fase 2. *In:* BYGGFORSK, S. (ed.). Trondheim: Sintef Byggforsk Vann og miljø.

SUN, S., BERTRAND-KRAJEWSKI, J.-L., LYNGGAARD-JENSEN, A., VAN DEN BROEKE, J., EDTHOFER, F., DO CÉU ALMEDIDA, M., RIBERIO, À. S. & MENAIA, J. 2011. Literature review for data validation methods. *PREPARED.*

SÆGROV, S. 2005. *CARE W - Computer Aided Rehabilitation for Water Networks,* Alliance House, 12 Caxton Street, London UK, International Water Association Publishing.

THAMES WATER UTILITIES LIMITED. 2012. *Thames Water Live* [Online]. www.thameswater.co.uk: www.thameswater.co.uk. Available: http://www.thameswater.co.uk/thameswaterlive/index.htm [Accessed March 8th 2012].

TRONDHEIM BYTEKNIKK 2005. Hovedplan for vannforsyning 2004-2013. *In:* TRONDHEIM KOMMUNE (ed.). Trondheim: Trondheim kommune,.

TRONDHEIM KOMMUNE 2007. Drikkevann - en livsviktig ressurs. Om vannforsyningen i Trondheim kommune. *In:* TRONDHEIM KOMMUNE (ed.). Trondheim: Trondheim kommune,.

TRONDHEIM KOMMUNE. 2011. *Vikelvdalen Vannbehandlingsanlegg* [Online]. trondheim.kommune.no: Trondheim Kommune. Available: http://www.trondheim.kommune.no/content/1117613765/Vikelvdalen-Vannbehandlingsanlegg?language=0 [Accessed March 21st 2012].

UGARELLI, R. 2008. *Asset Management of wastewater networks.* PhD, UNIVERSITA' DI BOLOGNA.

UGARELLI, R. April 20th 2012. *RE: request of meeting.* Type to MUNTHE-KAAS, M. S., SCOLA, S., ROKSTAD, M. M. & SÆGROV, S.

UGARELLI, R., VENKATESH, G., BRATTEBO, H., FEDERICO, V. D. & SAEGROV, S. 2010. Asset Management for Urban Wastewater Pipeline Networks. *Journal of infrastructure systems,* 16, 112-121.

USEPA. 2008. *Asset Management: A Best Practices Guide* [Online]. epa.gov: USEPA. [Accessed November 9th 2011].

USEPA. 2012. *Software That Models the Hydraulic and Water Quality Behavior of Water Distribution Piping Systems* [Online]. epa.gov: epa.gov. Available: http://www.epa.gov/nrmrl/wswrd/dw/epanet.html [Accessed February 16th 2012].

VAN NOORTWIJK, J. M. & FRANGOPOL, D. M. 2004. Two probabilistic life-cycle maintenance models for deteriorating civil infrastructures. *Probabilistic Engineering Mechanics,* 19, 345-359.

VANIER, D., NEWTON, L. & RAHMAN, S. 2006. A Framework for Municipal Infrastructure Management for Canadian Municipalities. National Research Council Canada.

VANIER, D. J. Asset management A to Z. 2001-09-09 2001 Philadelphia, PA. 1-16.

VANIER, D. J. 2006. Chapter 7 - Towards sustainable municipal infrastructure asset management. *Handbook on Urban Sustainability.* Dordrecht, The Netherlands: Springer.

VANIER, D. J., NEWTON, L. A. & RAHMAN, S. 2009. Implementation Details to Support a Generalized Framework for Municipal Infrastructure Management.

VANIER, D. N., L. & RAHMAN, S. 2006. A Framework for Municipal Infrastructure Management for Canadian Municipalities. National Research Council Canada.

VIGON, B. W. & JENSEN, A. A. 1995. Life cycle assessment: data quality and databases practitioner survey. *Journal of Cleaner Production,* 3**,** 135-141.

VITORINO, D. 2012. *RE: Difference between LLIFE and FAIL.* Type to ROKSTAD, M. M.

VLOERBERGH, I. N., SCHIPDAM, R., VAN THIENEN, P. & BEUKEN, R. H. S. 2011. Pipe Fitters assist to predict investment needs for water main rehabilitation. *Leading Edge Asset Management 2011.* Mülheim, Germany: KWR Watercycle Research Institute.

VLOERBERGH, I. N. & VAN THIENEN, P. 2011. Registratie geeft storingen waarde – Implementatie en evaluatie van USTORE. *Bouwstenen Leidingnetbeheer.* KWR.

VMW. 2010. *Feiten en cijfers* [Online]. Available: http://www.vmw.be/nl/collections/18/feiten-en-cijfers.html#Overzichtstabel%202010%20per%20provincie [Accessed September 26 2011].

VMW. 2012a. *Infrastructuur* [Online]. vmw.be: VMW. Available: http://www.vmw.be/nl/content/161/infrastructuur.html [Accessed February 6th 2012].

VMW. 2012b. *Organisatiestructuur* [Online]. vmw.be: vmw.be. Available: http://www.vmw.be/organisation [Accessed March 1st 2012].

VMW OOST-VLAANDEREN 2009. Activiteitenverslag 2009 Provenciale directie Oost-Vlaanderen. *In:* BOUWENS, A. (ed.). VMW Oost-Vlaanderen, Gent: Vlaamse Maatschappij voor Watervoorziening.

VMW OOST-VLAANDEREN 2010. Activiteitenverslag 2010 Provenciale directie Oost-Vlaanderen. *In:* BOUWENS, A. (ed.). VMW Oost-Vlaanderen, Gent: Vlaamse Maatschappij voor Watervoorziening.

VREEBURG, J. H. G., BLOKKER, E. J. M., RIETVELD, J., SLUIJS, A. A. I. & BLOTTERBLOM, S. 2011. Hydrant maintenance as example of targeted Asset Management. *LESAM - Leading Edge Strategic Asset Management 2011.* Mülheim, Germany: International Water Association.

WANG, R. Y., KON, H. B. & MADNICK, S. E. Data quality requirements analysis and modeling. Data Engineering, 1993. Proceedings. Ninth International Conference on, 19-23 Apr 1993 1993. 670-677.

WANG, R. Y., REDDY, M. P. & KON, H. B. 1995a. Toward quality data: An attribute-based approach. *Decision Support Systems,* 13**,** 349-372.

WANG, R. Y., STOREY, V. C. & FIRTH, C. P. 1995b. A framework for analysis of data quality research. *Knowledge and Data Engineering, IEEE Transactions on,* 7**,** 623-640.

WANG, R. Y. & STRONG, D. M. 1996. Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.,* 12**,** 5-33.

WEIDEMA, B. P. & WESNÆS, M. S. 1996. Data quality management for life cycle inventories—an example of using data quality indicators. *Journal of Cleaner Production,* 4**,** 167-174.

WIRAHADIKUSUMAH, R., ABRAHAM, D. & ISELEY, T. 2001. Challenging Issues in Modeling Deterioration of Combined Sewers. *Journal of infrastructure systems,* 7**,** 77-84.

WOOD, A. & LENCE, B. J. 2006. Assessment of water main break data for ASSET MANAGEMENT. *American Water Works Association. Journal,* 98**,** 76-86,10.

WOOD, A., LENCE, B. J. & LIU, W. 2007. Constructing water main break databases for asset management. *American Water Works Association. Journal,* 99**,** 52-65,10.

WWW.IPJ.NO. 2012. *EA Driftskontroll og overvåking* [Online]. www.ipj.no,: Ing.firma Paul Jørgensen as,. Available: www.ipj.no [Accessed April 10th 2012].

# Appendix A Input fields in diary database

**Table 27: Input fields for functional requirement entries**

| Input # | Input name | Input type | Alternatives |
|---|---|---|---|
| 1 | Event ID | Numeric | |
| 2 | Asset ID | Numeric | |
| 3 | Date of entry | Numeric | |
| 4 | Requirement start date | Numeric | |
| 5 | Requirement stop date | Numeric | |
| 6 | Scenario ID | Numeric | |
| 7 | Modelling method | Selection list | InfoWorks, LCC-AM/QM…. |
| 8 | Parameter | Selection list | Value, RSL, expected failure rate, condition, hydraulic criticality, expected time to inspection/maintenance, maintenance/repair effect, performance, cost of reliability increase, cost of condition increase, cost of repair |
| 9 | Parameter value | Numeric | |
| 10 | Attachments | files | |

**Table 28: Input fields for complaint events**

| Input # | Input name | Input type | Alternatives |
|---|---|---|---|
| 1 | Event ID | Numeric | |
| 2 | Date of complaint | Numeric | |
| 3 | Nature of complaint | Selection list | Quality, pressure, leakage, service etc. |
| 4 | Utility reaction | Selection list | Nothing, inspection, inspection and repair |
| 5 | Responsible personnel ID | Numeric | |
| 6 | Resolved date | Numeric | |
| 7 | Resolved through action: Event ID | Numeric | |
| 8 | Attachments | files | |

**Table 29: Input fields for water quality samples**

| Input # | Input name | Input type | Alternatives |
|---|---|---|---|
| 1 | Event ID | Numeric | |
| 2 | Date of sampling | Numeric | |
| 3 | Quality sample status | Selection list | Pending, finished |
| 3 | Address | Alphanumeric | |
| 4 | Customer ID | Numeric | |
| 5 | Quality problem? | Selection list | Yes, no |
| 6 | Array of quality parameters[14] | Numeric | |
| 7 | Recommended action | Selection list | Customer warning, inspection, new sample, nothing |
| 8 | Attachments | files | |

**Table 30: Input fields for inspection events**

| Input # | Input name | Input type | Alternatives |
|---|---|---|---|
| 1 | Event ID | Numeric | |
| 2 | Asset ID | Numeric | |
| 3 | Date of inspection | Numeric | |
| 4 | Inspection status | Selection list | Pending, finished |
| 5 | Inspection priority | Selection list | High, medium, low |
| 6 | Responsible personnel ID | Numeric | |
| 7 | Reason for inspection | Selection list | Random, time-based planning, due to other event |
| 8 | Reason event ID | Numeric | |
| 9 | Inspection method | Selection list | Visual, CCTV, pressure gauging, flow measurement… |
| 10 | Condition measurement | Numeric | |
| 11 | Service disturbance duration | Numeric | |
| 12 | Number of customers affected | Numeric | |
| 13 | Work hours invested | Numeric | |
| 14 | Cost of material | Numeric | |
| 15 | Cost of equipment | Numeric | |
| 16 | Recommended action | Selection list | Nothing, immediate repair, new inspection in x years |
| 17 | Suggested time of recommended action | Numeric | |
| 18 | Attachments | files | |

---

[14] Hygienic, physical, chemical and aesthetical quality

Table 31: Input fields for maintenance events

| Input # | Input name | Input type | Alternatives |
|---|---|---|---|
| 1 | Event ID | Numeric | |
| 2 | Asset ID | Numeric | |
| 3 | Date of maintenance | Numeric | |
| 4 | Maintenance status | Selection list | Pending, finished |
| 5 | Maintenance priority | Selection list | High, medium, low |
| 6 | Responsible personnel ID | Numeric | |
| 7 | Reason for maintenance | Selection | Random, time-based planning, due to other event |
| 8 | Reason event ID | Numeric | |
| 9 | Maintenance type | Selection list | Flushing, swabbing, air scouring, pressure jetting, scraping, relining, instrument calibration, renovation etc. |
| 10 | Number of clients affected | Numeric | |
| 11 | Disturbance duration | Numeric | |
| 12 | Number of customers affected | Numeric | |
| 13 | Work hours invested | Numeric | |
| 14 | Cost of material | Numeric | |
| 15 | Cost of equipment | Numeric | |
| 16 | Recommended action | Selection list | Nothing, immediate repair, new inspection in x years |
| 17 | Time until recommended action | Numeric | |
| 18 | Attachments | Files | |

Table 32: Input fields for failure and repair events

| Input # | Input name | Input type | Alternatives | Comment |
|---|---|---|---|---|
| 1 | Event ID | numeric | | |
| 2 | Asset ID | numeric | | |
| 3 | Date of repair | numeric | | |
| 4 | Repair status | selection list | Pending, finished | |
| 5 | Repair priority | Selection list | High, medium, low | |
| 6 | Responsible personnel ID | Numeric | | |
| 7 | Work hours invested | numeric | | |
| 8 | Cost of material | numeric | | |
| 9 | Cost of equipment | | | |
| 10 | Repair method | selection list | | |
| 11 | Duration of disturbance | numeric | | hours of disturbance |
| 12 | Number of customers affected | numeric | | |
| 13 | Recommended action | selection list | | |
| 14 | Time until recommended action | numeric | | |
| 15 | Attachments | files | | |
| 16 | Company code | selection list | | USTORE |
| 17 | Registration date | numeric | | USTORE |
| 18 | Failure date | numeric | | USTORE |
| 19 | Location: GPS XY-coordinate | numeric | | USTORE |
| 20 | Location: Address | alphanumeric | | USTORE |
| 21 | Failed object type | selection list | | USTORE |
| 22 | Type of joint | selection list | | USTORE |
| 23 | Material | selection list | | USTORE |
| 24 | Material type | selection list | | USTORE |
| 25 | External corrosion protection | selection list | | USTORE |
| 26 | Internal corrosion protection | selection list | | USTORE |
| 27 | Installation year | numeric | | USTORE |
| 28 | Pressure category/wall thickness | numeric | | USTORE |
| 29 | Diameter | alphanumeric | | USTORE |
| 30 | Nature of failure | selection list | | USTORE |
| 31 | Cause 1 | selection list | | USTORE |
| 32 | Cause 2 | selection list | | USTORE |
| 33 | Cause 3 | selection list | | USTORE |
| 34 | Soil type | selection list | | USTORE |
| 35 | Pollution | selection list | | USTORE |
| 36 | Trees | selection list | | USTORE |
| 37 | Groundwater | selection list | | USTORE |
| 38 | Ground coverage | selection list | | USTORE |
| 39 | Traffic | selection list | | USTORE |
| 40 | Need to close valves? | selection list | | USTORE |
| 41 | Did the valves function? | selection list | | USTORE |

**Table 33: Input fields for decommissioning events**

| Input # | Input name | Input type | Alternatives |
|---|---|---|---|
| 1 | Event ID | Numeric | |
| 2 | Asset ID | Numeric | |
| 3 | Date of decommissioning | Numeric | |
| 4 | Decommissioning status | Selection list | Pending, finished |
| 5 | Reason 1 for decommissioning | Selection list | Third party reason, reliability problems, technical service life ended,<br>economic service life ended, change in performance requirements (capacity increase), asset redundant |
| 6 | Reason 2 for decommissioning | Selection list | Third party reason, reliability problems, technical service life ended,<br>economic service life ended, change in performance requirements (capacity increase), asset redundant |
| 7 | Replaced by asset ID | Numeric | |
| 8 | Remaining asset value | Numeric | |
| 9 | Asset value evaluation method | Selection list | Estimate, model, etc. |
| 10 | Related inspection IDs | Numeric | |
| 11 | Related model results IDs | Numeric | |
| 12 | Attachments | files | |

**Table 34: Input fields for model results**

| Input # | Input name | Input type | Alternatives |
|---|---|---|---|
| 1 | Model ID | Numeric | |
| 2 | Asset ID | Numeric | |
| 3 | Date of modelling | Numeric | |
| 4 | Date of result | Numeric | |
| 5 | Model scenario ID | Numeric | |
| 6 | Modelling method | Selection list | InfoWorks, LCC-AM/QM…. |
| 7 | Modelled parameter | Selection list | Value, RSL, expected failure rate, condition, hydraulic criticality, expected time to inspection/maintenance, maintenance/repair effect, performance, cost of reliability increase, cost of condition increase, cost of repair |
| 8 | Modelled value | Numeric | |
| 9 | Modelled uncertainty | Numeric | |
| 10 | Attachments | files | |

# Appendix B Data documentation for cost-benefit model

**Table 35: Data classes, items, dependencies and unit costs**

| Class | Items | Dependent on | Unit cost [min] |
|---|---|---|---|
| Asset Basic | Asset ID, coordinates of asset, date of installation, height (s), length, material, nominal diameter, orientation, type of asset, friction factor, roughness | None | 20.0 |
| Asset Cost | Acquiring cost | Asset Basic | 4.0 |
| Asset Extended | Connection type, corrosion protection, network type, street name | Asset Basic | 2.5 |
| Asset Status | Abandoned/decommissioned/open/closed | Asset Basic | 1.0 |
| Asset external Basic | Bedding material, consumptions, depth of cover, groundwater level, number of connections, soil type, surface material, traffic | Asset Basic | 15.0 |
| Complaint Basic | Customer ID, customer address, date of complaint, nature of complaint, utility reaction | None | 8.0 |
| Customer Basic | Customer ID, consumption, customer connection point coordinates | None | 10.0 |
| Decommissioning Basic | Asset ID, date of decommissioning | Asset Basic | 5.0 |
| Decommissioning Extended | Asset value evaluation method, decommissioning status, reason 1 for decommissioning, Reason 2 for decommissioning, related inspection ID, remaining asset value, replaced by asset ID | Decommissioning Basic | 20.0 |
| Failure Basic | Asset ID, company code, diameter, failed object type, failure date, installation year, Location: Address, material, nature of failure registration date | Asset Basic | 5.0 |
| Failure Extended | Did valves function?, external/internal corrosion protection, location: GPS coordinates, material type, need to close valves?, pressure category/wall thickness, type of joint | Failure Basic | 10.0 |
| Failure Mode | Cause 1, cause 2, cause 3, ground coverage, groundwater, pollution, soil type, traffic, trees | Failure Basic | 6.5 |
| Inspection Basic | Date of inspection, inspection method, inspection result: condition/performance | Asset Basic | 4.5 |
| Inspection Cost | Cost of equipment, cost of material, work hours invested | Inspection Basic | 7.0 |
| Inspection Extended | Reason for inspection, recommended (time of) action | Inspection Basic | 2.0 |
| Inspection Disturbance | disturbance duration, number of customers affected | Inspection Basic | 4.0 |
| Maintenance Basic | Asset ID, date of maintenance | None | 2.0 |
| Maintenance Cost | Cost of equipment, cost of material, work hours invested | Maintenance Basic | 7.0 |
| Maintenance Extended | Reason for maintenance, maintenance type, recommended (time of) action | Maintenance Basic | 1.0 |
| Maintenance Disturbance | disturbance duration, number of customers affected | Maintenance Basic | 5.0 |
| Operational Basic | Gauging, open/closed, active/not active etc. | Asset Basic | 60.0 |
| Rehabilitation Cost | Estimated unit cost for rehabilitation techniques | Asset Basic | 30.0 |
| Renewal Cost | Estimated unit cost for renewal techniques | Asset Basic | 30.0 |
| Repair Basic | Asset ID, date of repair | Failure Basic | 3.0 |
| Repair Cost | Cost of equipment, cost of material, work hours invested | Repair Basic | 7.0 |
| Repair Extended | Repair method, repair priority, repair status, responsible personnel ID, recommended (time of) action | Repair Basic | 3.5 |

| | | | |
|---|---|---|---|
| Repair Disturbance | disturbance duration, number of customers affected | Repair Basic | 2.0 |
| Replacement Cost | Estimated unit cost for replacement of component cohort | Asset Basic | 30.0 |
| Request Basic | Customer ID, date of request, nature of request | None | 10.0 |
| Water quality Basic | Tapping connection point, date of sampling, recommended action, water quality array | Asset Basic | 12.5 |

**Table 36: Information items in the cost-benefit model**

| Information | Aggregation level | Strategic | Tactical | Operational | What do you own? | What is it worth? | What is the deferred maintenance? | What is the condition? | What is the remaining service life? | What do you fix first? | Goal/regulatory compliance? | Resource planning? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inventory | Asset | 🟩 | 🟩 | 🟩 | 🟩 | | | | | | | |
| Acquiring cost/depreciation | Cohort | | 🟩 | | | 🟩 | | | | | | |
| Cost of replacement | Asset | | 🟩 | | | 🟩 | | | | | | |
| Value of performance (deprival cost) | Asset | | 🟩 | | | 🟩 | | | | | | |
| LCC | Asset | | 🟩 | 🟩 | | 🟩 | | | 🟩 | 🟩 | | |
| LCC | Cohort | | 🟩 | | | 🟩 | | | 🟩 | 🟩 | | |
| Reliability-cost relationship | Cohort | 🟩 | 🟩 | | | | 🟩 | | | | | |
| Condition-cost relationship | Cohort | | 🟩 | | | | 🟩 | | | | | |
| Cost of maintenance | Asset | | 🟩 | 🟩 | | | 🟩 | | 🟩 | 🟩 | | |
| Cost of maintenance | Cohort | | 🟩 | | | | 🟩 | | 🟩 | 🟩 | | |
| Structural condition | Asset | | 🟩 | 🟩 | | | | 🟩 | | | | |
| Structural condition | Cohort | | 🟩 | | | | | 🟩 | | | | |
| Performance | Asset | | 🟩 | 🟩 | | | | 🟩 | | | | |
| Performance | Cohort | 🟩 | 🟩 | | | | | 🟩 | | | | |
| Reliability | Cohort | 🟩 | 🟩 | | | | | 🟩 | 🟩 | 🟩 | 🟩 | |
| Cost of failure | Asset | | 🟩 | 🟩 | | | | | 🟩 | 🟩 | | |
| Cost of failure | Cohort | | 🟩 | | | | | | 🟩 | 🟩 | | |
| Hydraulic criticality | Asset | 🟩 | 🟩 | 🟩 | | | | | 🟩 | 🟩 | 🟩 | |
| Hydraulic risk | Asset | 🟩 | 🟩 | 🟩 | | | | | 🟩 | 🟩 | 🟩 | |
| Economical risk | Asset | 🟩 | 🟩 | 🟩 | | | | | 🟩 | 🟩 | 🟩 | |
| Level of service | DMA | 🟩 | 🟩 | | | | | 🟩 | 🟩 | 🟩 | | |
| Customer-minutes lost | DMA | 🟩 | 🟩 | | | | | | | | | |
| Cost of inspection | Asset | | 🟩 | 🟩 | | | | | 🟩 | 🟩 | | |
| Cost of inspection | Cohort | | 🟩 | | | | | | 🟩 | 🟩 | | |
| Work order | Asset | | | 🟩 | | | | | | | | 🟩 |
| Work orders | Cohort | | 🟩 | 🟩 | | | | | | | | 🟩 |
| Capacity | Asset | | 🟩 | 🟩 | | | | | | | 🟩 | |
| Pressure | Asset | | 🟩 | | | | | | | | 🟩 | |
| Fire water availability | Asset | | 🟩 | 🟩 | | | | | | | 🟩 | |
| Customer satisfaction | DMA | 🟩 | | | | | | | | | 🟩 | |
| Economical condition | DMA | 🟩 | | | | | | | | | 🟩 | |
| Water quality | DMA | 🟩 | 🟩 | | | | | | | | 🟩 | |
| Investment need | CL | | 🟩 | | | | | | | | | 🟩 |
| Strategic investment need | DMA | 🟩 | | | | | | | | | | 🟩 |
| Prioritised investments | Cohort | | 🟩 | | | | | | | | | 🟩 |

# Appendix C User instruction for cost-benefit model

## Appendix C1 Introduction

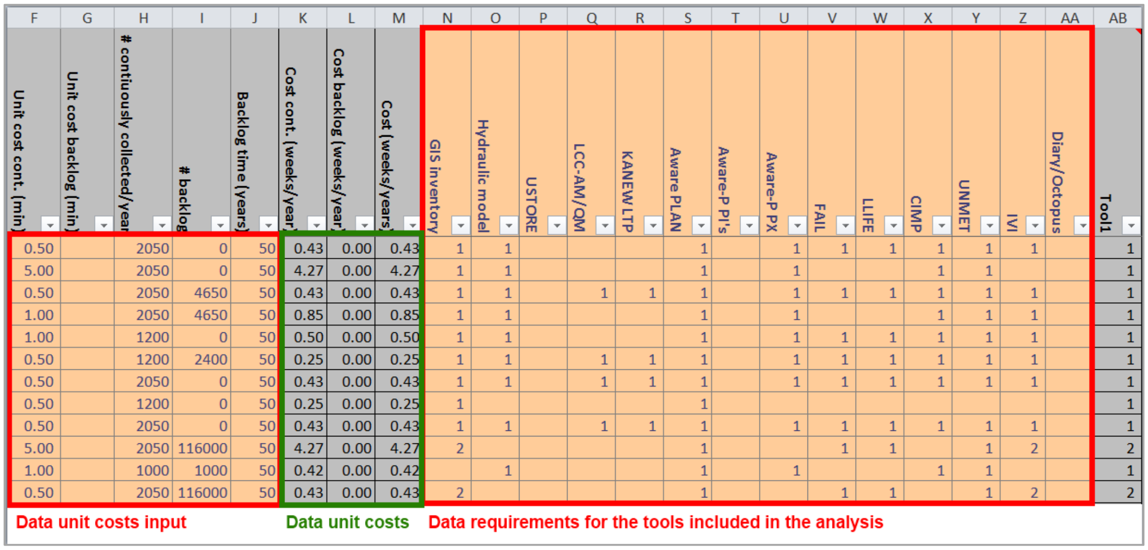
The purpose of this instruction is to show how the cost-benefit spread sheet model is configured, and how to it is used. The model is made to help utilities evaluate the relationship between cost of data collection, and the informational benefits it gives through the use of asset management (AM) tools.

All fields with changeable input are marked with a light colour in the spread sheet. Inputs and outputs are framed with a strong colour in this instruction. All input fields in the model sheet are coloured pinkish.

The model has five sheets, each of which will be explained separately in the following sections. In sheet 4 the objective is to find the most suitable configuration of tools to utilise in the utility, and estimate the absolute and relative costs of using these tools. In sheet 5, the direct cost of collecting different data classes is related to the benefits, and thus shows which data are most valuable. The approach of evaluating tool combinations is a practical one, while the approach of evaluating direct cost-benefit relationships is a more generic and theoretical one.

## Appendix C2 Explanation of sheets

### Appendix C2-1 DataCost sheet



| F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unit cost cont. (min) | Unit cost backlog (min) | # contiuously collected/year | # backlog | Backlog time (years) | Cost cont. (weeks/year) | Cost backlog (weeks/year) | Cost (weeks/years) | GIS inventory | Hydraulic model | USTORE | LCC-AM/QM | KANEW LTP | Aware PLAN | Aware-P PI's | Aware-P PX | FAIL | LLIFE | CIMP | UNMET | IVI | Diary/Octopus | Tool1 |
| 0.50 | | 2050 | 0 | 50 | 0.43 | 0.00 | 0.43 | 1 | 1 | | | | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 |
| 5.00 | | 2050 | 0 | 50 | 4.27 | 0.00 | 4.27 | 1 | 1 | | | | 1 | | 1 | | | 1 | 1 | | | 1 |
| 0.50 | | 2050 | 4650 | 50 | 0.43 | 0.00 | 0.43 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 |
| 1.00 | | 2050 | 4650 | 50 | 0.85 | 0.00 | 0.85 | 1 | 1 | | | | 1 | | 1 | | | 1 | 1 | 1 | | 1 |
| 1.00 | | 1200 | 0 | 50 | 0.50 | 0.00 | 0.50 | 1 | 1 | | | | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 |
| 0.50 | | 1200 | 2400 | 50 | 0.25 | 0.00 | 0.25 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 |
| 0.50 | | 2050 | 0 | 50 | 0.43 | 0.00 | 0.43 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 |
| 0.50 | | 1200 | 0 | 50 | 0.25 | 0.00 | 0.25 | 1 | | | | | 1 | | | | | | | | | 1 |
| 0.50 | | 2050 | 0 | 50 | 0.43 | 0.00 | 0.43 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 |
| 5.00 | | 2050 | 116000 | 50 | 4.27 | 0.00 | 4.27 | 2 | | | | | 1 | | | 1 | 1 | | 1 | 2 | | 2 |
| 1.00 | | 1000 | 1000 | 50 | 0.42 | 0.00 | 0.42 | | 1 | | | | 1 | | 1 | | | 1 | 1 | | | 1 |
| 0.50 | | 2050 | 116000 | 50 | 0.43 | 0.00 | 0.43 | 2 | | | | | 1 | | | 1 | 1 | | 1 | 2 | | 2 |

Figure 25: Input and output of the DataCost sheet

The purpose of the DataCost sheet is to set up the cost of data for each combination of tools that is included in the analysis.

There are two groups of input in the DataCost sheet.

1. The input for cost calculation of each data item:

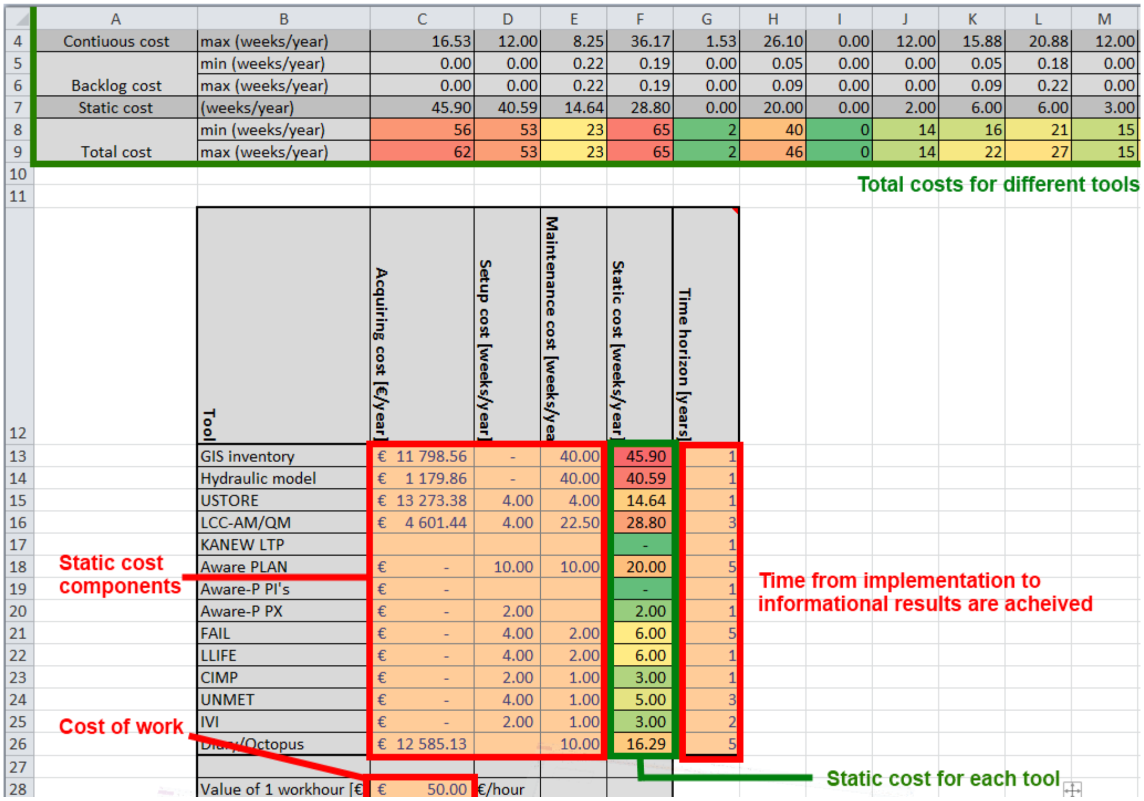
- The unit costs of each data item [column F]
- The unit costs of backlogged data items [column G]
- The number of entries of each data item every year [column H]
- The number of backlogged items [column I]
- The time horizon in which backlogged items are expected to be itemized [column J]

2. The input to determine the data requirement of each tool. Each tool has one column, where required data are marked as *1*, optional data as *2,* and non-relevant data as *"empty"*.

See Figure 25.

The only output from the data-cost sheet is the unit cost for each data item (column [K] to [M]).

## Appendix C2-2 ToolCost sheet



| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | Contiuous cost | max (weeks/year) | 16.53 | 12.00 | 8.25 | 36.17 | 1.53 | 26.10 | 0.00 | 12.00 | 15.88 | 20.88 | 12.00 |
| 5 | | min (weeks/year) | 0.00 | 0.00 | 0.22 | 0.19 | 0.00 | 0.05 | 0.00 | 0.00 | 0.05 | 0.18 | 0.00 |
| 6 | Backlog cost | max (weeks/year) | 0.00 | 0.00 | 0.22 | 0.19 | 0.00 | 0.09 | 0.00 | 0.00 | 0.09 | 0.22 | 0.00 |
| 7 | Static cost | (weeks/year) | 45.90 | 40.59 | 14.64 | 28.80 | 0.00 | 20.00 | 0.00 | 2.00 | 6.00 | 6.00 | 3.00 |
| 8 | | min (weeks/year) | 56 | 53 | 23 | 65 | 2 | 40 | 0 | 14 | 16 | 21 | 15 |
| 9 | Total cost | max (weeks/year) | 62 | 53 | 23 | 65 | 2 | 46 | 0 | 14 | 22 | 27 | 15 |

Total costs for different tools

| | Tool | Acquiring cost [€/year] | Setup cost [weeks/year] | Maintenance cost [weeks/yea | Static cost [weeks/year] | Time horizon [years] |
|---|---|---|---|---|---|---|
| 13 | GIS inventory | € 11 798.56 | - | 40.00 | 45.90 | 1 |
| 14 | Hydraulic model | € 1 179.86 | - | 40.00 | 40.59 | 1 |
| 15 | USTORE | € 13 273.38 | 4.00 | 4.00 | 14.64 | 1 |
| 16 | LCC-AM/QM | € 4 601.44 | 4.00 | 22.50 | 28.80 | 3 |
| 17 | KANEW LTP | | | | - | 1 |
| 18 | Aware PLAN | € - | 10.00 | 10.00 | 20.00 | 5 |
| 19 | Aware-P PI's | € - | | | - | 1 |
| 20 | Aware-P PX | € - | 2.00 | | 2.00 | 1 |
| 21 | FAIL | € - | 4.00 | 2.00 | 6.00 | 5 |
| 22 | LLIFE | € - | 4.00 | 2.00 | 6.00 | 1 |
| 23 | CIMP | € - | 2.00 | 1.00 | 3.00 | 1 |
| 24 | UNMET | € - | 4.00 | 1.00 | 5.00 | 3 |
| 25 | IVI | € - | 2.00 | 1.00 | 3.00 | 2 |
| 26 | Di[illegible]/Octopus | € 12 585.13 | | 10.00 | 16.29 | 5 |
| 27 | | | | | | |
| 28 | Value of 1 workhour [€ | € 50.00 | €/hour | | | |

Static cost components

Cost of work

Time from implementation to informational results are acheived

Static cost for each tool

**Figure 26: Input and output for the ToolCost sheet**

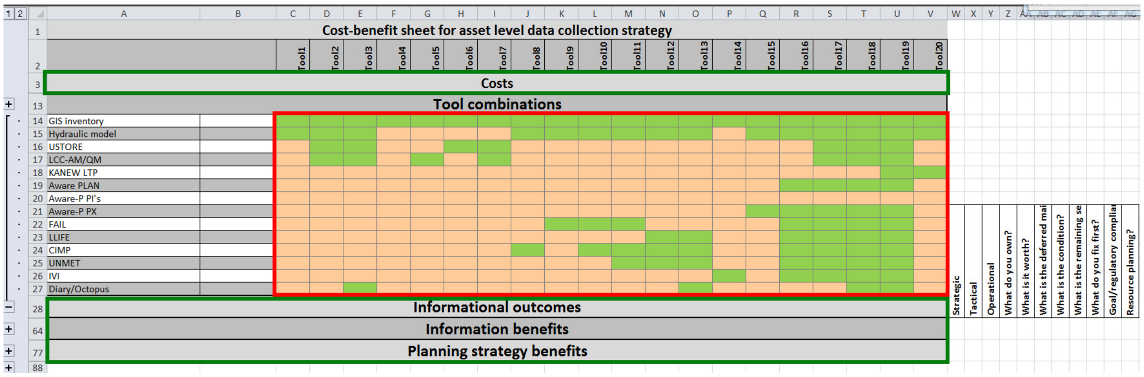
In ToolCost, the base costs of using the different tools included in the analysis are to be calculated. Also, the cost of data for each individual tool is displayed as output in the sheet.

There are three input areas in the ToolCost sheet:

1. The base cost components, consisting of the acquiring cost of the tool (distributed over the period the tool is used), the work hours needed to set the tool up, and the work hours needed to maintain the tool
2. The time from implementation to informational outcomes are achieved, which is used as input to calculate the total cost (not per year) until the informational outcomes are achieved
3. The cost of 1 work hour

The output from the DataCost sheet is the tool-specific costs (both data costs and base costs).

## Appendix C2-3 CostBenefit sheet



Cost-benefit sheet for asset level data collection strategy

Costs

Tool combinations

| | Tool1 | Tool2 | Tool3 | Tool4 | Tool5 | Tool6 | Tool7 | Tool8 | Tool9 | Tool10 | Tool11 | Tool12 | Tool13 | Tool14 | Tool15 | Tool16 | Tool17 | Tool18 | Tool19 | Tool20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GIS inventory | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| Hydraulic model | 🟩 | 🟩 | 🟩 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| USTORE | 🟧 | 🟩 | 🟩 | 🟧 | 🟧 | 🟩 | 🟩 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟧 |
| LCC-AM/QM | 🟧 | 🟩 | 🟩 | 🟧 | 🟩 | 🟧 | 🟩 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟧 |
| KANEW LTP | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 |
| Aware PLAN | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 |
| Aware-P PI's | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 |
| Aware-P PX | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 |
| FAIL | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 |
| LLIFE | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 |
| CIMP | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 |
| UNMET | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟧 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 |
| IVI | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟧 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 |
| Diary/Octopus | 🟧 | 🟧 | 🟩 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟧 | 🟧 | 🟧 | 🟧 | 🟩 | 🟩 | 🟧 |

Informational outcomes

Information benefits

Planning strategy benefits

| Strategic | Tactical | Operational | What do you own? | What is it worth? | What is the deferred mai | What is the condition? | What is the remaining se | What do you fix first? | Goal/regulatory complia | Resource planning? |
|---|---|---|---|---|---|---|---|---|---|---|

Figure 27: Input field in CostBenefit sheet

The purpose of the CostBenefit sheet is to allow the user to select the tool combinations that are to be incorporated in the analysis, and to show the raw results of costs and benefits aligned with the different tool combinations.

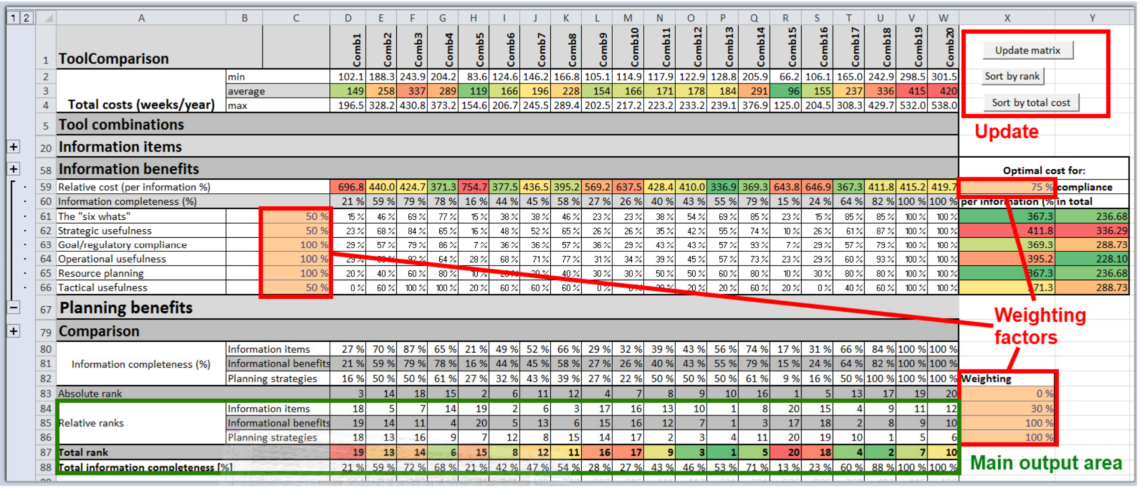
The CostBenefit sheet has one main input area: The area where the tools that are included in each tool combination are set; see the red frame in Figure 27. The user can include a tool into any of the tool combinations (Tool1-Tool20) by indicating the cell with *1*. All tools indicated with *1* in a tool combination column will be included in the analysis.

(The tools that are included in the analysis can be changed; then the user have to change the data requirements of that tool (DataCost), the base costs of the tool (ToolCost) and the rules for the benefits the tool produces (CostBenefit). This has to be done manually.)

The CostBenefit sheet has four main output areas, all organised in drop-down menus (can be opened by clicking on the "+"-buttons on the far left):

1. The **costs** for each tool combination
2. The achieved **informational outcomes** for each tool combination
3. The **information benefits** emerging from the informational outcomes in each tool combination
4. The **planning strategies** that are possible to achieve with the informational outcomes and benefits for each tool combination

## Appendix C2-4 ToolComparison



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ToolComparison | | | Comb1 | Comb2 | Comb3 | Comb4 | Comb5 | Comb6 | Comb7 | Comb8 | Comb9 | Comb10 | Comb11 | Comb12 | Comb13 | Comb14 | Comb15 | Comb16 | Comb17 | Comb18 | Comb19 | Comb20 | | |
| 2 | | min | | 102.1 | 188.3 | 243.9 | 204.2 | 83.6 | 124.6 | 146.2 | 166.8 | 105.1 | 114.9 | 117.9 | 122.9 | 128.8 | 205.9 | 66.2 | 106.1 | 165.0 | 242.9 | 298.5 | 301.5 | | |
| 3 | | average | | 149 | 258 | 337 | 289 | 119 | 166 | 196 | 228 | 154 | 166 | 171 | 178 | 184 | 291 | 96 | 155 | 237 | 336 | 415 | 420 | | |
| 4 | Total costs (weeks/year) | max | | 196.5 | 328.2 | 430.8 | 373.2 | 154.6 | 206.7 | 245.5 | 289.4 | 202.5 | 217.2 | 223.2 | 233.2 | 239.1 | 376.9 | 125.0 | 204.5 | 308.3 | 429.7 | 532.0 | 538.0 | | |
| 5 | Tool combinations | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 | Information items | | | | | | | | | | | | | | | | | | | | | | | | |
| 58 | Information benefits | | | | | | | | | | | | | | | | | | | | | | | Optimal cost for: | |
| 59 | Relative cost (per information %) | | | 696.8 | 440.0 | 424.7 | 371.3 | 754.7 | 377.5 | 436.5 | 395.2 | 569.2 | 637.5 | 428.4 | 410.0 | 336.9 | 369.3 | 643.8 | 646.9 | 367.3 | 411.8 | 415.2 | 419.7 | 75 % | compliance |
| 60 | Information completeness (%) | | | 21 % | 59 % | 79 % | 78 % | 16 % | 44 % | 45 % | 58 % | 27 % | 26 % | 40 % | 43 % | 55 % | 79 % | 15 % | 24 % | 64 % | 82 % | 100 % | 100 % | per information (% | in total |
| 61 | The "six whats" | | 50 % | 15 % | 46 % | 69 % | 77 % | 15 % | 38 % | 38 % | 46 % | 23 % | 23 % | 38 % | 54 % | 69 % | 85 % | 23 % | 15 % | 85 % | 85 % | 100 % | 100 % | 367.3 | 236.68 |
| 62 | Strategic usefulness | | 50 % | 23 % | 68 % | 84 % | 65 % | 16 % | 48 % | 52 % | 65 % | 26 % | 26 % | 35 % | 42 % | 55 % | 74 % | 10 % | 26 % | 61 % | 87 % | 100 % | 100 % | 411.8 | 336.29 |
| 63 | Goal/regulatory compliance | | 100 % | 29 % | 57 % | 79 % | 86 % | 7 % | 36 % | 36 % | 57 % | 36 % | 29 % | 43 % | 43 % | 57 % | 93 % | 7 % | 29 % | 57 % | 79 % | 100 % | 100 % | 369.3 | 288.73 |
| 64 | Operational usefulness | | 100 % | 29 % | [illegible] | 92 % | 64 % | 28 % | 68 % | 71 % | 77 % | 31 % | 34 % | 39 % | 45 % | 57 % | 73 % | 23 % | 29 % | 60 % | 93 % | 100 % | 100 % | 395.2 | 228.10 |
| 65 | Resource planning | | 100 % | 20 % | 40 % | 60 % | 80 % | 10 % | 20 % | 20 % | 40 % | 30 % | 30 % | 50 % | 50 % | 60 % | 80 % | 10 % | 30 % | 80 % | 80 % | 100 % | 100 % | 367.3 | 236.68 |
| 66 | Tactical usefulness | | 50 % | 0 % | 60 % | 100 % | 100 % | 20 % | 60 % | 60 % | 60 % | 0 % | 0 % | 20 % | 20 % | 20 % | 60 % | 20 % | 0 % | 40 % | 60 % | 100 % | 100 % | 371.3 | 288.73 |
| 67 | Planning benefits | | | | | | | | | | | | | | | | | | | | | | | | |
| 79 | Comparison | | | | | | | | | | | | | | | | | | | | | | | | |
| 80 | Information completeness (%) | Information items | | 27 % | 70 % | 87 % | 65 % | 21 % | 49 % | 52 % | 66 % | 29 % | 32 % | 39 % | 43 % | 56 % | 74 % | 17 % | 31 % | 66 % | 84 % | 100 % | 100 % | | |
| 81 | | Informational benefits | | 21 % | 59 % | 79 % | 78 % | 16 % | 44 % | 45 % | 58 % | 27 % | 26 % | 40 % | 43 % | 55 % | 79 % | 15 % | 24 % | 64 % | 82 % | 100 % | 100 % | | |
| 82 | | Planning strategies | | 16 % | 50 % | 50 % | 61 % | 27 % | 32 % | 43 % | 39 % | 27 % | 22 % | 50 % | 50 % | 50 % | 61 % | 9 % | 16 % | 50 % | 100 % | 100 % | 100 % | Weighting | |
| 83 | Absolute rank | | | 3 | 14 | 18 | 15 | 2 | 6 | 11 | 12 | 4 | 7 | 8 | 9 | 10 | 16 | 1 | 5 | 13 | 17 | 19 | 20 | 0 % | |
| 84 | Relative ranks | Information items | | 18 | 5 | 7 | 14 | 19 | 2 | 6 | 3 | 17 | 16 | 13 | 10 | 1 | 8 | 20 | 15 | 4 | 9 | 11 | 12 | 30 % | |
| 85 | | Informational benefits | | 19 | 14 | 11 | 4 | 20 | 5 | 13 | 6 | 15 | 16 | 12 | 7 | 1 | 3 | 17 | 18 | 2 | 8 | 9 | 10 | 100 % | |
| 86 | | Planning strategies | | 18 | 13 | 16 | 9 | 7 | 12 | 8 | 15 | 14 | 17 | 2 | 3 | 4 | 11 | 20 | 19 | 10 | 1 | 5 | 6 | 100 % | |
| 87 | **Total rank** | | | 19 | 13 | 14 | 6 | 15 | 8 | 12 | 11 | 16 | 17 | 9 | 3 | 1 | 5 | 20 | 18 | 4 | 2 | 7 | 10 | | |
| 88 | **Total information completeness [%]** | | | 21 % | 59 % | 72 % | 68 % | 21 % | 42 % | 47 % | 54 % | 28 % | 27 % | 43 % | 46 % | 53 % | 71 % | 13 % | 23 % | 60 % | 88 % | 100 % | 100 % | | |

**Figure 28: Main input and output of the ToolComparisonS sheet**

The purpose of the ToolComparison sheet is to be able to evaluate the cost-benefit relationship, based on the three different ways of expressing the benefits: (1) the achieved informational items, (2) the achieved informational benefits, and (3) the possible planning strategies.

There are four drop-down lists in the sheet, which all require input:

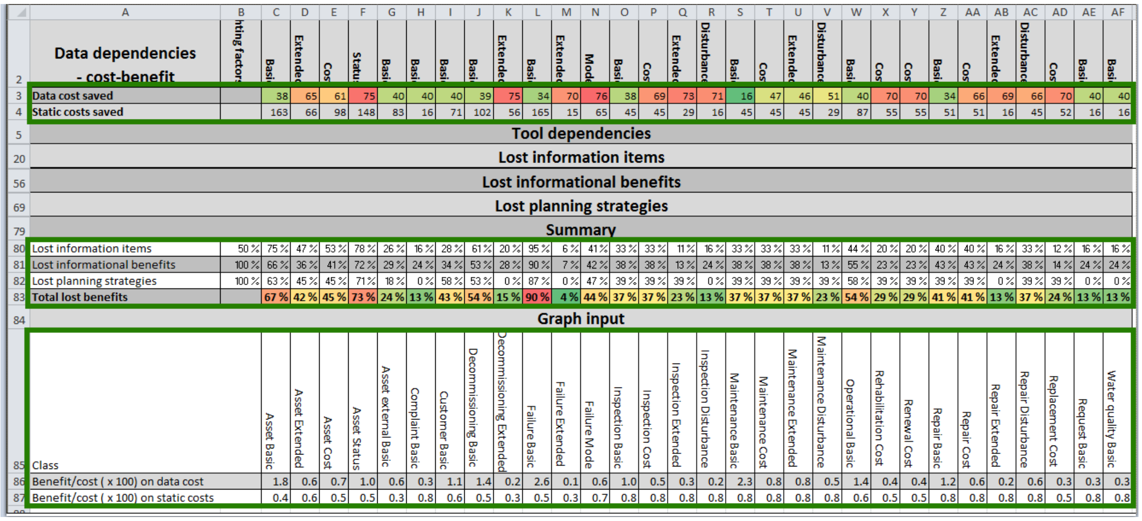
1. The Tool combinations, which show the different tools which are included in each combination. This list requires no input.
2. The information items list, where the different achieved informational items are listed. This list requires the user to put in weighting factors [column C]
3. The information benefits list, where information benefits are shown (strategic planning, tactical planning, achievement of "the six whats"). This list also requires weighting factors [column C]. In addition, the user needs to set in a desired benefit compliance level [cell X59], i.e. the percentage over which optimal costs will be calculated.
4. The achieved planning levels list, where the planning levels are listed. Also here weighting factors are applied.

The output of the sheet is:

1. Total and outcome-relative costs for the different tool combinations [row 21, 59 and 68]
2. The optimal (minimal average cost) for all the different outcomes [column X and Y]
3. Rankings for the different tool combinations [cells A83:W88]

**Before evaluating the content of the ToolComparison sheet, it is necessary to update the sheet by pressing "Update matrix".**

## Appendix C2-5 DataComparison



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Data dependencies - cost-benefit | nting factor | Basi | Extende | Cos | Statu | Basi | Basi | Basi | Basi | Extende | Basi | Extende | Mode | Basi | Cos | Extende | Disturbanc | Basi | Cos | Extende | Disturbanc | Basi | Cos | Cos | Basi | Cos | Extende | Disturbanc | Cos | Basi | Basi |
| 3 | Data cost saved | | 38 | 65 | 61 | 75 | 40 | 40 | 40 | 39 | 75 | 34 | 70 | 76 | 38 | 69 | 73 | 71 | 16 | 47 | 46 | 51 | 40 | 70 | 70 | 34 | 66 | 69 | 66 | 70 | 40 | 40 |
| 4 | Static costs saved | | 163 | 66 | 98 | 148 | 83 | 16 | 71 | 102 | 56 | 165 | 15 | 65 | 45 | 45 | 29 | 16 | 45 | 45 | 45 | 29 | 87 | 55 | 55 | 51 | 51 | 16 | 45 | 52 | 16 | 16 |
| 5 | Tool dependencies | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 | Lost information items | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 56 | Lost informational benefits | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 69 | Lost planning strategies | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 79 | Summary | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 80 | Lost information items | 50 % | 75 % | 47 % | 53 % | 78 % | 26 % | 16 % | 28 % | 61 % | 20 % | 95 % | 6 % | 41 % | 33 % | 33 % | 11 % | 16 % | 33 % | 33 % | 33 % | 11 % | 44 % | 20 % | 20 % | 40 % | 40 % | 16 % | 33 % | 12 % | 16 % | 16 % |
| 81 | Lost informational benefits | 100 % | 66 % | 36 % | 41 % | 72 % | 29 % | 24 % | 34 % | 53 % | 28 % | 90 % | 7 % | 42 % | 38 % | 38 % | 13 % | 24 % | 38 % | 38 % | 38 % | 13 % | 55 % | 23 % | 23 % | 43 % | 43 % | 24 % | 38 % | 14 % | 24 % | 24 % |
| 82 | Lost planning strategies | 100 % | 63 % | 45 % | 45 % | 71 % | 18 % | 0 % | 58 % | 53 % | 0 % | 87 % | 0 % | 47 % | 39 % | 39 % | 39 % | 0 % | 39 % | 39 % | 39 % | 39 % | 58 % | 39 % | 39 % | 39 % | 39 % | 0 % | 39 % | 39 % | 0 % | 0 % |
| 83 | **Total lost benefits** | | **67 %** | **42 %** | **45 %** | **73 %** | **24 %** | **13 %** | **43 %** | **54 %** | **15 %** | **90 %** | **4 %** | **44 %** | **37 %** | **37 %** | **23 %** | **13 %** | **37 %** | **37 %** | **37 %** | **23 %** | **54 %** | **29 %** | **29 %** | **41 %** | **41 %** | **13 %** | **37 %** | **24 %** | **13 %** | **13 %** |
| 84 | Graph input | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 85 | Class | | Asset Basic | Asset Extended | Asset Cost | Asset Status | Asset external Basic | Complaint Basic | Customer Basic | Decommissioning Basic | Decommissioning Extended | Failure Basic | Failure Extended | Failure Mode | Inspection Basic | Inspection Cost | Inspection Extended | Inspection Disturbance | Maintenance Basic | Maintenance Cost | Maintenance Extended | Maintenance Disturbance | Operational Basic | Rehabilitation Cost | Renewal Cost | Repair Basic | Repair Cost | Repair Extended | Repair Disturbance | Replacement Cost | Request Basic | Water quality Basic |
| 86 | Benefit/cost ( x 100) on data cost | | 1.8 | 0.6 | 0.7 | 1.0 | 0.6 | 0.3 | 1.1 | 1.4 | 0.2 | 2.6 | 0.1 | 0.6 | 1.0 | 0.5 | 0.3 | 0.2 | 2.3 | 0.8 | 0.8 | 0.5 | 1.4 | 0.4 | 0.4 | 1.2 | 0.6 | 0.2 | 0.6 | 0.3 | 0.3 | 0.3 |
| 87 | Benefit/cost ( x 100) on static costs | | 0.4 | 0.6 | 0.5 | 0.5 | 0.3 | 0.8 | 0.6 | 0.5 | 0.3 | 0.5 | 0.3 | 0.7 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.6 | 0.5 | 0.5 | 0.8 | 0.8 | 0.8 | 0.8 | 0.5 | 0.8 | 0.8 |

**Figure 29: Three main output areas of the DataComparison sheet**

In the previous sheets, the goal has been to find the most suitable combination of tools for the utility. However, the question about what data are actually most valuable is still unanswered. The DataComparison sheet answers this. Here, the cost of data collection for each data class that has been defined is directly linked to the same informational outcomes as in the previous sheets, and ultimately used to calculate the cost-benefit ratio for each data class.

The DataComparison sheet has no input. The weighting factors are extracted from the previous sheet, and are hence the same.

The output of the DataComparison sheet is made in a similar fashion as the CostBenefit sheet. However, in the DataComparison sheet the benefits are expressed as "lost benefits", i.e. a certain benefit is considered as "lost" if a necessary data class is taken away. The two main outputs are:

1. The saved data cost [row 3-4], which expresses the work hours saved by not collecting the data in the data class.
2. The lost benefits [row 80-84], which expresses the benefits that are lost if the data in the data class is not collected.

In addition, the cost-benefit ratio (lost benefits divided by saved data costs) is displayed graphically.

**NB! The graphs have to be updated by pressing the "Update graphs" button. The graph contents can also be sorted.**

# Appendix D Sensitivity and uncertainty analysis results from cost-benefit model

## Appendix D1 Weighting factors

The situation with differentiated weighting factors was tested against the situation of uniform weighting factors (all outputs weighted equally). The weighting factors are enclosed in Table 37.

**Table 37: Weighting factors for model outcomes**

| Item | Level | Weight | Item | Level | Weight |
|---|---|---|---|---|---|
| Inventory | Asset | 100 % | Fire water availability | Asset | 30 % |
| Acquiring cost/depreciation | Cohort | 50 % | Customer satisfaction | DMA | 100 % |
| Cost of replacement | Asset | 30 % | Economic condition | DMA | 100 % |
| Value of performance (deprival cost) | Asset | 30 % | Water quality | DMA | 100 % |
| LCC | Asset | 30 % | Investment need | CL | 50 % |
| LCC | Cohort | 50 % | Strategic investment need | DMA | 100 % |
| Reliability-cost relationship | Cohort | 50 % | Prioritised investments | Cohort | 50 % |
| Condition-cost relationship | Cohort | 30 % | Strategic usefulness |  | 100 % |
| Cost of maintenance | Asset | 30 % | Tactical usefulness |  | 50 % |
| Cost of maintenance | Cohort | 50 % | Operational usefulness |  | 30 % |
| Structural condition | Asset | 30 % | The "six whats" |  | 50 % |
| Structural condition | Cohort | 50 % | Goal/regulatory compliance |  | 100 % |
| Performance | Asset | 30 % | Resource planning |  | 100 % |
| Performance | Cohort | 50 % | Reactive |  | 30 % |
| Reliability | Cohort | 50 % | Time-based |  | 30 % |
| Cost of failure | Asset | 30 % | Performance-based |  | 50 % |
| Cost of failure | Cohort | 50 % | Condition-based |  | 50 % |
| Hydraulic criticality | Asset | 30 % | Criticality-based |  | 50 % |
| Hydraulic risk | Asset | 50 % | Proactive reliability-based |  | 100 % |
| Economical risk | Asset | 50 % | Proactive risk-based |  | 100 % |
| Level of service | DMA | 100 % | Predictive LCC-based |  | 100 % |
| Customer-minutes lost | DMA | 100 % | LCC-risk-based |  | 100 % |
| Cost of inspection | Asset | 30 % | Rank Information items |  | 50 % |
| Cost of inspection | Cohort | 50 % | Rank Informational benefits |  | 100 % |
| Work order | Asset | 30 % | Rank Planning strategies |  | 50 % |
| Work orders | Cohort | 50 % |  |  |  |
| Capacity | Asset | 30 % |  |  |  |
| Pressure | Asset | 30 % |  |  |  |

# Appendix D2 Sensitivity of data unit costs

The sensitivity of data unit costs were assessed by increasing the unit costs of each data class individually, by 25 %, and calculating the deviations created by the change in each data class.

**Table 38: Data class sensitivity results**

| Changed class | | Change in cost | | Change in rank | | Change in cost-benefit-ratio | | Total cost |
|---|---|---|---|---|---|---|---|---|
| | | Maximal | Average | Maximal | Average | Maximal | Average | [weeks/year] |
| Asset | Basic | 3.2 % | 1.6 % | 0 | 0.00 | 20.0 % | 13.0 % | 12.4 |
| Asset | Cost | 0.9 % | 0.2 % | 1 | 0.10 | 5.1 % | 0.2 % | 3.4 |
| Asset | Extended | 0.2 % | 0.1 % | 0 | 0.00 | 2.6 % | 0.1 % | 1.5 |
| Asset | Status | 0.2 % | 0.1 % | 0 | 0.00 | 1.6 % | 0.1 % | 0.9 |
| Asset external | Basic | 1.3 % | 0.7 % | 0 | 0.00 | 10.7 % | 0.4 % | 11.3 |
| Complaint | Basic | 0.3 % | 0.1 % | 1 | 0.15 | 20.0 % | 0.7 % | 3.3 |
| Customer | Basic | 0.0 % | 0.0 % | 0 | 0.00 | 20.0 % | 0.7 % | 9.6 |
| Decommissioning | Basic | 0.1 % | 0.0 % | 0 | 0.00 | 0.6 % | 0.0 % | 0.3 |
| Decommissioning | Extended | 0.1 % | 0.1 % | 1 | 0.10 | 2.8 % | 0.1 % | 1.7 |
| Failure | Basic | 0.6 % | 0.2 % | 0 | 0.00 | 4.3 % | 0.8 % | 2.7 |
| Failure | Extended | 1.0 % | 0.2 % | 1 | 0.15 | 7.0 % | 0.2 % | 6.0 |
| Failure | Mode | 0.8 % | 0.2 % | 0 | 0.00 | 4.5 % | 0.2 % | 3.5 |
| Inspection | Basic | 0 % | 0 % | 1 | 0.1 | 1.2 % | 0.2 % | 0.7 |
| Inspection | Cost | 0 % | 0 % | 1 | 0.15 | 2.0 % | 0.1 % | 1.5 |
| Inspection | Extended | 0 % | 0 % | 1 | 0.15 | 2.5 % | 0.2 % | 0.3 |
| Inspection | Disturbance | 0 % | 0 % | 1 | 0.15 | 2.0 % | 0.2 % | 0.3 |
| Maintenance | Basic | 2 % | 0 % | 1 | 0.15 | 20.0 % | 2.0 % | 9.2 |
| Maintenance | Cost | 1 % | 0 % | 1 | 0.15 | 10.9 % | 0.5 % | 8.8 |
| Maintenance | Extended | 0.7 % | 0.2 % | 1 | 0.15 | 7.9 % | 0.4 % | 3.8 |
| Maintenance | Disturbance | 0.9 % | 0.3 % | 1 | 0.15 | 11.9 % | 0.5 % | 10.8 |
| Operational | Basic | 0.2 % | 0.1 % | 1 | 0.15 | 2.2 % | 0.2 % | 1.3 |
| Rehabilitation | Cost | 0.2 % | 0.1 % | 1 | 0.1 | 2.0 % | 0.2 % | 0.3 |
| Renewal | Cost | 0.2 % | 0.1 % | 1 | 0.15 | 2.0 % | 0.2 % | 0.3 |
| Repair | Basic | 0.4 % | 0.2 % | 1 | 0.1 | 2.4 % | 0.4 % | 1.6 |
| Repair | Cost | 0.7 % | 0.3 % | 1 | 0.15 | 4.4 % | 0.3 % | 3.8 |
| Repair | Extended | 0.2 % | 0.1 % | 1 | 0.15 | 2.4 % | 0.2 % | 2.3 |
| Repair | Disturbance | 0.3 % | 0.1 % | 1 | 0.15 | 2.0 % | 0.2 % | 1.1 |
| Replacement | Cost | 0.2 % | 0.1 % | 1 | 0.1 | 2.0 % | 0.2 % | 0.3 |
| Request | Basic | 1.0 % | 0.2 % | 1 | 0.15 | 20.0 % | 0.9 % | 8.3 |
| Water quality | Basic | 0.8 % | 0.2 % | 1 | 0.15 | 7.9 % | 0.5 % | 6.5 |

## Appendix D3 Tool cost sensitivity

The sensitivity of the base tool costs was assessed by changing the base cost for each tool individually, by 25 %, and then calculating the deviations it produced.

**Table 39: Tool cost sensitivities**

| Changed tool | Change in cost | | | | Change in rank | | | | Total base costs |
|---|---|---|---|---|---|---|---|---|---|
| | Maximum | | Average | | Maximum | | Average | | |
| | Sim. #1 | Sim. #2 | Sim. #1 | Sim. #2 | Sim. #1 | Sim. #2 | Sim. #1 | Sim. #2 | |
| GIS inventory | 18.0 % | 5.1 % | 8.8 % | 4.7 % | 6 | 1 | 1.00 | 0.10 | 45.9 |
| Hydraulic model | 10.2 % | 4.5 % | 5.5 % | 4.1 % | 6 | 1 | 0.85 | 0.10 | 40.6 |
| USTORE | 4.6 % | 1.6 % | 0.8 % | 1.5 % | 1 | 1 | 0.05 | 0.10 | 14.6 |
| LCC-AM/QM | 6.4 % | 3.1 % | 1.6 % | 2.9 % | 2 | 1 | 0.55 | 0.10 | 28.3 |
| KANEW LTP | 0.4 % | 0.3 % | 0.0 % | 0.1 % | 0 | 2 | 0.00 | 0.35 | 3.0 |
| Aware PLAN | 3.2 % | 2.0 % | 0.4 % | 0.9 % | 2 | 3 | 0.50 | 0.70 | 20.0 |
| Aware-P PX | 1.0 % | 0.4 % | 0.1 % | 0.1 % | 1 | 0 | 0.05 | 0.00 | 4.0 |
| FAIL | 1.4 % | 0.6 % | 0.4 % | 0.1 % | 1 | 1 | 0.15 | 0.10 | 6.0 |
| LLIFE | 1.2 % | 0.6 % | 0.2 % | 0.2 % | 1 | 2 | 0.15 | 0.45 | 6.0 |
| CIMP | 0.7 % | 0.3 % | 0.2 % | 0.1 % | 1 | 2 | 0.10 | 0.35 | 3.0 |
| UNMET | 1.1 % | 0.5 % | 0.2 % | 0.1 % | 1 | 2 | 0.25 | 0.40 | 5.0 |
| IVI | 1.2 % | 0.3 % | 0.1 % | 0.1 % | 1 | 2 | 0.05 | 0.20 | 3.0 |
| Diary/Octopus | 4.7 % | 4.0 % | 1.0 % | 3.7 % | 1 | 1 | 0.20 | 0.10 | 36.3 |

## Appendix D4 Sensitivity results for simulation #2

**Table 40: Sensitivity results for simulation #2**

| Input parameter | Method | Output parameter | Maximum change | Average change | Most sensitive input |
|---|---|---|---|---|---|
| **Weighting factors** | Weighting factors versus no weighting factors | Rank | 3 | 0.55 | Relative rank weights |
| | | Total information completeness | 7.0 % | 3.4 % | |
| | | Benefit/cost-ratio | 26.8 % | 12.2 % | |
| **Tool costs** | 25 % increase in base cost for each tool | Total cost | 5.1 % | 1.4 % | High acquiring cost tools |
| | | Rank | 3 | 0.2 | |

## Appendix D5 Uncertainty

**Table 41: Uncertainty limits (absolute change in *work weeks/year*)**

| Comb | Uncertainty | | Comb | Uncertainty | | Tool | Base cost uncertainty | |
|---|---|---|---|---|---|---|---|---|
| | % | Absolute | | % | Absolute | | % | Absolute |
| 1 | 3.7 % | 3.9 | 21 | 4.2 % | 10.7 | GIS inventory | 6.2 % | 7.9 |
| 2 | 2.7 % | 5.3 | 22 | 3.9 % | 10.5 | Hydraulic model | 0.0 % | 0.0 |
| 3 | 4.2 % | 10.7 | 23 | 3.9 % | 10.5 | USTORE | 5.4 % | 2.7 |
| 4 | 7.8 % | 17.4 | 24 | 4.1 % | 10.7 | LCC-AM/QM | 0.0 % | 0.0 |
| 5 | 5.9 % | 5.3 | 25 | 3.9 % | 10.5 | KANEW LTP | 0.0 % | 0.0 |
| 6 | 3.1 % | 3.9 | 26 | 3.8 % | 10.5 | Aware PLAN | 9.8 % | 9.9 |
| 7 | 3.5 % | 5.3 | 27 | 3.8 % | 10.5 | Aware-P PX | 0.0 % | 0.0 |
| 8 | 2.3 % | 3.9 | 28 | 4.1 % | 10.7 | FAIL | 22.3 % | 9.5 |
| 9 | 3.6 % | 3.9 | 29 | 3.8 % | 10.5 | LLIFE | 17.3 % | 9.5 |
| 10 | 4.1 % | 5.0 | 30 | 4.1 % | 10.7 | CIMP | 0.0 % | 0.0 |
| 11 | 4.0 % | 5.0 | 31 | 3.7 % | 10.5 | UNMET | 15.7 % | 9.5 |
| 12 | 3.9 % | 5.0 | 32 | 4.4 % | 11.6 | IVI | 28.2 % | 8.4 |
| 13 | 3.7 % | 5.0 | 33 | 3.6 % | 10.5 | Diary/Octopus | 13.6 % | 27.3 |
| 14 | 7.8 % | 17.4 | 34 | 3.6 % | 10.5 | | | |
| 15 | 7.0 % | 5.0 | 35 | 3.6 % | 10.5 | | | |
| 16 | 3.6 % | 3.9 | 36 | 3.6 % | 10.5 | | | |
| 17 | 2.9 % | 5.0 | 37 | 3.5 % | 10.5 | | | |
| 18 | 2.1 % | 5.2 | 38 | 3.6 % | 10.5 | | | |
| 19 | 3.4 % | 10.5 | 39 | 3.5 % | 10.5 | | | |
| 20 | 3.4 % | 10.5 | 40 | 3.4 % | 10.5 | | | |

# Appendix E Paper on cost-benefit analysis

NB! The appendices to this paper have been omitted since they all exist in similar form within the thesis. The references for the paper coincide with the references in the thesis' main text (section 8). The elements in this paper are all extracted from the thesis main text.

## *Data collection for water infrastructure asset management: A cost-benefit analysis approach*

Marius Møller Rokstad (1), Rita Ugarelli (2), Sveinung Sægrov (3)
*Dep.of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology, SP. Andersens veg 5, 7491 Trondheim, Norway*

*(1) + 47 980 36 599, marius.rokstad@gmail.com; (2) + 47 22 96 55 51, rita.ugarelli@sintef.no; (3) +47 73 59 47 65, sveinung.sagrov@ntnu.no*

**Abstract:**

This paper contains a proposed methodology for assessing the cost-benefit relationship between data collection and data utilisation for infrastructure asset management tools. In this methodology, the costs are expressed as the work hours invested in collecting the data, while the benefits are expressed as informational outcomes. The model has been developed as a spread sheet model, and is demonstrated with a fictive water utility. The paper shows how such a model can be used by a utility in order to assess (1) current cost-benefit situation for their asset management data, and (2) how the benefits can be improved. The methodology seems to be a promising concept, but the model still needs to be improved by including measured values of the unit costs for infrastructure data collection.

**Keywords:** Infrastructure asset management, cost-benefit, data collection

## 1 Introduction

Traditionally, it has been economically feasible for water utilities to maintain a desired level of service by repairing or replacing assets whenever they fail (D'Água et al., 2007). However, the peak of capital investment intensity has passed for the water utilities, which means that the assets in the urban water infrastructure are ageing. At the same time, the constraints from level of service, economic, population growth, and urbanisation are getting increasingly stringent (Alegre and Matos, 2009). The modern society is systemically dependent on water infrastructure; therefore the need for utility owners to consider and minimise risk of service failure is growing (Aven, 2010). For these reasons the traditional reactive approach will not be feasible for water utilities of the future – better management of the water utility assets is needed.

There exists a wide array of tools aiding more foresighted asset management strategies. Software and analytical procedures can, through knowledge about economy, engineering and science, assist more rational decision-making processes within the water utilities. However, they all have one thing in common: they rely on data about the assets. A lot of the data that is required by the decision-support tools have traditionally not been collected by the water utilities, because they have not been needed, and data availability often stands as an impeding factor for these tools (Halfawy and Figueroa, 2006).

When implementing new tools for asset management in a water utility, one is often faced with the fact that data collection is a resource-consuming process; a balance between collection of data and consumption of data (to produce useful informational outcomes) needs to be established. In this balancing process, data providers are often tempted to want to simplify models and tools so that the data requirements will be reduced, while analysts are often tempted to want to incorporate an excess of factors, bearing no regard of the high data requirements it imposes (Alegre et al., 2006a). The challenge is to find a suitable level of data collection, where the data collection efforts are in proportion with the value of the informational outcomes they can provide – a cost-benefit analysis of data collection.

This paper describes a methodology for assessing the relationship between the cost of collecting data (expressed as work hours invested), and the benefits the data yield (expressed as informational outcomes).

Some of the questions one would hope to answer with such an analysis are: What will be the cost of data for a given combination of tools? What combination of asset management tools is optimal for the utility, with respect to data cost and information output? What data are most valuable? What data should one prioritise to collect when a new collection system is introduced?

## 2 Theory review

### 2.1 What is IAM?

A definition of infrastructure asset management (AM) that is applicable for water utilities is:

*"…asset management can be recognised as a set of management, financial, economic, engineering activities, systematic and coordinated, to optimally manage the physical assets and their associated performance, risks and expenditures over their life cycle with the objective of ensuring level of service in the most cost-effective manner." (Ugarelli et al., 2010)*

One may thus understand AM as a set of optimisation activities, in which performance and level of service is maximised, whilst risk and life cycle cost is minimised. *An asset* is understood as any property of a person or organisation that has a value and/or is revenue-generating (Amadi-Echendu et al., 2010).

It is often referred to three levels of planning in AM of water utilities (Alegre and Covas, 2010):

4. At the **strategic level** the general direction of the planning is set, through defining the goals and objectives of the utility. The desired level of service, overall condition and performance, and general financial needs are determined at the strategic level; how the objectives should be "benchmarked" should also be decided on the strategic level. Strategic planning occurs at a network level, and typically has a long time horizon (10-20 years).
5. At the **tactical level** different alternatives to achieve the strategic goals are evaluated and compared. In a tactical planning situation are individual assets evaluated with respect to the objectives defined at the strategic level. A tactical plan yields a set of prioritised assets that need to be renewed or rehabilitated (projects). The time horizon for tactical planning is typically 3-5 years.
6. At the **operational level**, different technologies to realise the projects, which have been selected on the tactical level, are evaluated and selected. The operational plan is a short-term plan (1-2 years) and contains the description of how the projects should be implemented

Further, it is stated that AM is a multidisciplinary pursuit, requiring competences such as *engineering, management* and *information* (Alegre et al., 2006a).

### 2.2 What is the role of data in AM?

The level of AM implementation one is able to achieve is dependent on what data that is available. A low level of implementation only requires an asset registry, accompanied by some method of asset valuation that does not necessarily account for the asset condition (Lemer, 1998, Vanier, 2001, Vanier and Rahman, 2006). A high level of asset management, on the other hand, requires an array of different data, keeping track of the physical condition, value and cost associated with the asset. The models that exist for assisting the assessment of the AM decision criteria are also very much dependent on data over a certain observation period, and very often are the quality of the results impeded by the quality and amount of the input data. Vanier (2001) argues that "efficient information management is the key to better decision-making for municipal infrastructure".

Several other authors identify lack of data (or quality data) as one of the main challenges or impediments for implementation of high-level level asset management (Halfawy and Figueroa, 2006, Halfawy et al., 2006, Lemer, 1998, Lin et al., 2006, Lin et al., 2007) and identifies data as the foundation for operational, tactical and strategic planning and resource management – complex models for (life cycle cost, condition, reliability, remaining service life) cannot be successfully implemented without adequate background data (Wood and Lence, 2006). This statement from Halfawy and Figueroa (2006) summarises asset management's dependence on life cycle data management:

*"Successful implementation of asset management strategies largely depends on: (1) the efficiency to share, access, and manage and manage the asset life-cycle data; and (2) the ability to efficiently support and coordinate the multi-disciplinary work processes at the operational and strategic levels." (Halfawy and Figueroa, 2006)*

## 2.3 AM tools and models

The use of models to aid the decision support processes inherent in asset management has become increasingly more important the past decades. Models for assessing and predicting reliability of buried assets, based on statistical and/or physical principles, have been utilised as decision support (Rajani and Kleiner, 2001, Kleiner and Rajani, 2001). Models can be used to assess performance, condition, criticality, failure rates, risk, life cycle costs and other decision-influencing factors.

Models for asset management can be understood as tools that rationalise and transform data into useful information. The models vary greatly in what underlying principles they utilise, what data they require, and in which form the results emerge:

- On one hand, a utility will usually rely on a hydraulic model portfolio, which is useful for assessing the importance of components, hydraulic performance, the effect of changes in the hydraulic conditions etc. A hydraulic model is based on simplified physical laws and basic characteristics of the components in the water distribution network (Rossman, 2000).
- On the other hand the utility may utilise models that assess the deterioration of the state of the network, in the form of condition, reliability, etc. – such models utilise asset characteristics data and historical data (repairs, maintenance, inspections, decommissioning) to forecast the "health" of the components, either by using mechanical or statistical principles. Deterioration models usually yield results on a cohort or asset aggregation level. (Kleiner and Rajani, 2001, Rajani and Kleiner, 2001, Røstum, 2000)
- Both models for assessing importance or performance, and deterioration models may be the basis for decision-making, which can be aided by decision-support systems. Decision-support models may be based on criticality, reliability, hydraulic or economical risk, life cycle costing etc. Decision-support is based on both the aforementioned model classes, as well as data about the different alternatives that are relevant to choose between. (Hadzilacos et al., 2000)

## 2.4 Tool examples

The different tools described here will be included in the cost-benefit model. In addition an inventory software will also be present in the analysis.

### 2.4.1 The Aware-P toolbox

As an example of the range of different modelling tools that can be used for asset management, the *Aware-P infrastructure asset management software* will be used (from here on denoted *Aware-P toolbox*) (baseform.org, 2012a). This toolbox contains both a hydraulic component, for assessing component importance and performance, a deterioration component for assessing reliability and service life, and a decision-support system for assessing different investment scenarios. The Aware-p toolbox is thus a broad example of how models can help to provide information necessary for rational planning.

The different modules of the Aware-P toolbox are reviewed in section 2.3.2.

### 2.4.2 KANEW LTP

KANEW is a methodology that was developed in the CARE-W project, and is an example of a long-term planning (LTP) tool. KANEW is a cohort survival model (Liu et al., 2012), and makes an analysis based on cohorts of pipes with similar characteristics and expected service lives, where survival functions are assigned to these cohorts. The tool is thus able to produce prognoses for the expected amount of pipe length needing to be replaced in a given year and cohort. Further, if assumptions about different rehabilitation and replacement alternatives are entered into KANEW, the software will be able to produce the expected costs associated with the rehabilitation scenario. (Baur and Herz, 1999, Herz, 1994, Herz, 1996)

### 2.4.3 USTORE

USTORE (*uniform storingsregistratiesysteem [Dutch]*) is a failure registry database developed at KWR Water Research Institute and consists of a shared database of failure incidents gathered from several participating water utilities. KWR has developed a uniform manner of registering failure data, either by standard paper forms or by web applications, which are used for registering failure events in the water utility database; further the data is exchanged (or copied) to KWR's central database, via a web application (USTOREweb), and can be utilised for statistical analysis (both by KWR and the utilities participating), rendering greater possibilities than the data of each utility alone. In USTORE, four levels of data are collected from a failure incident: (1) details about the failure, (2) asset characteristics, (3) surroundings, and (4) situational factors. (Vloerbergh et al., 2011)

### 2.4.4 LCC software

LCC software is an AM tool that aids the asset owners to make decisions based on the life cycle cost of the assets. In order to make LCC- based considerations, acquiring, maintenance, inspection, failure costs etc. all have to be taken into account. An example of such a software may be the *LCC-AM/QM* from S&G and Partners (S&G and Partners, 2009); the input and output of LCC-AM/QM has been the basis for the LCC tool described in the cost-benefit model in this paper. In this paper it is assumed that the LCC software is reliability-based, and that it does not consider component criticality.

### 2.4.5 Diary functions

A utility can keep track of the assets' life cycles through keeping a diary of all the different events that occur with the assets throughout their service lives; all events and interventions (repairs, inspections, maintenance, decommissioning etc.) will then be obtainable for each asset. In the cost-benefit analysis such functionalities will be exemplified by two different tools:

- Asset intervention diary, including water quality and (leakage) inspection results registration
- Complaint diary, where complaints are registered and spatio-temporally fixed

An example of a water utility diary function can be found in the Norwegian software Gemini VA (Powel, 2011)

## 3 Model specifications

The objective of the cost-benefit model described here is to establish a relationship between:

3. The cost associated with collection of different classes of data, which are related to the asset management of buried infrastructure assets
4. … And the benefit these data yield, measured as achieved informational outcomes.

An example of a data cost may be the work hours required to fill in and process data after a repair has been completed. An informational outcome may for instance be the reliability of a specific type of asset.

### 3.1 Model structure

The analysis will be carried out with *asset management tools* as the common denominator. In this context, an asset management tool is an instrument that transforms raw data into useful informational outcomes. The assumption is that a certain set of input data classes, treated through a certain set of tools will produce a certain set of informational outcomes. The combination of data classes (or sources) and tools may be synergetic, i.e. the production of informational outcomes may be greater when tools and data sources are combined, than when separated.

The cost-benefit model is built as a spread sheet model. The actual model has three sheets:

6. **The data collection cost sheet** (DataCost). Here, all the different data items are described and assigned costs. In this sheet the data requirement for different tools are also identified
7. **The base cost sheet** (ToolCost). In this sheet the base costs of the different tools are calculated.
8. **Results sheet** (Results). In this sheet costs and benefits of the different tools are summarised, based on the two cost sheets. For each informational outcome, certain rules and criteria have been defined to determine whether or not they emerge from the tool combination in question. The sheet is made to

display the preliminary results, which are to be used in the evaluation.

In addition to the three model sheets, there are two sheets

9. **The tool comparison sheet** (ToolComparison), where the different tool combinations can be evaluated with respect to the benefits as information items, informational benefits or achieved planning strategies. Within each group, weighting factors can be assigned in order to differentiate the importance of the different outcomes. The different tool combinations are ranked based on the outcomes and the weighting factors.
10. **The data comparison sheet** (DataComparison), where the cost-benefit ratio of different classes of data can be evaluated graphically, based on the weighting factors from the previous sheets. In the DataComparison sheet a number of data dependencies have also been defined; if the utility collect one class of data, it must also collect another – for instance, it makes no sense for a utility to collect specific data about the cost of an inspection, if the basic data about the inspection is not accounted for.

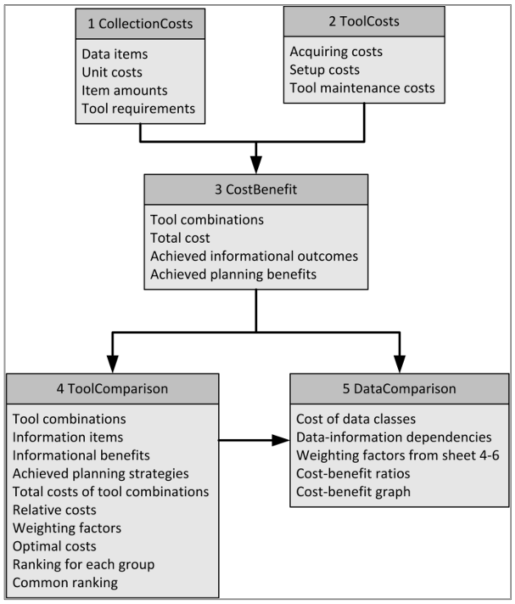


**Figure 1: Structure of spread sheet model**

In sheets 4 the objective is to evaluate and select a suitable combination of tools for the utility. In sheet 5 the objective is to evaluate the cost-benefit ratio for different data classes directly, in order to make out what data are most valuable, regardless of tool combinations. The structure of the spread sheet model is shown in Figure 1.

## 3.2 Costs

There are two types of costs associated with the production of informational benefits: (1) base costs and (2) data collection costs.

**The base costs** are costs associated with purchasing tools, treating data for analysis, and maintaining the tools, and are calculated according to Eq. 1.

For the acquiring cost, both one-time purchase and running license costs for software is included, and distributed over the expected lifetime of the software (usually 10 years). Setup costs are also distributed over the lifetime of the tool.

$$C_{base} = \frac{C_{tool}}{C_{workhour}} + \frac{WH_{setup}}{UL_{tool}} + WH_{main}$$

where:

$C_{base}$ = the base cost $\left[{}^{hours}\!/_{year}\right]$

$C_{tool}$ = the acquiring cost of the tool $\left[{}^{€}\!/_{year}\right]$

$C_{workhour}$ = the average cost of a work hour $\left[{}^{€}\!/_{hours}\right]$

$WH_{setup}$ = the time needed to set up the tool $[hours]$

$UL_{tool}$ = the usable lifespan of the tool $[years]$

$WH_{main}$ = the time needed to maintain the tool $\left[{}^{hours}\!/_{year}\right]$

Eq. 1

**The data collection costs** are costs associated with the collection of data. In order to calculate the costs associated with data collection for different scenarios, different types of data have been organised according to Table 1. Data are organised into classes and sub-classes - A data class is defined as *a set of data items that are internally similar with respect to what kind of information they are describing, their aggregation level, and how they are collected.*

The data classes and items have been defined in Appendix B. The cost to collect one data item per year can then be calculated according to Eq. 2.

**Table 1: Data class definitions**

| Attribute | Explanation | Example |
|---|---|---|
| Item # | Identification number of the item | 3 |
| Class name | Name of the data class the entry belongs to | Failure |
| Sub-class name | Identifies a sub-branch of the data class | Basic |
| Data aggregation level | The level at which the entry is linked. District meter area, cohort, address or asset levels are possible. | Asset |
| Item | The name of the data item | Date of failure |
| Number of entries per year | The number of events occurring each year, that need to be continuously recorded (new assets, repairs, inspections, complaints, decommissioned assets etc.) | 200 |
| Data backlog | Backlog of missing or unrecorded data (missing diameters, materials etc.), and the assumed time horizon for collecting these | 1600 |
| Backlog horizon | The time horizon under which the utility will spend resources to fill in data backlog in order to make records complete. If the utility does not want to invest resources in data backlogs, this field can be excluded from the model. | 15 years |
| Unit work hour cost | The estimated number of work hours associated with collecting one item of the data class. The unit cost is different (higher) for collection missing data, than for collection of current/new data. | 1 min |
| Tools using the data item | Each data point has a list of tools that use the data field. Data can be indicated as absolutely required (1) or optional (2). | LLIFE, FAIL, USTORE… |

$$c_{\text{data item } i} = n_{\text{cont.},i} \cdot t_{\text{cont.},i} + \frac{n_{\text{backlog},i}}{BH_i} \cdot t_{\text{backlog},i}$$

where:

$c_{\text{data item } i}$ = the cost of data item $i$ $\left[{}^{\text{min}}/_{\text{year}}\right]$

$n_{\text{cont.},i}$ = the number of item $i$ registered each year Eq. 2

$t_{\text{cont.},i}$ = the unit cost to collect item $i$ $[\text{min}]$

$n_{\text{backlog},i}$ = the number of backlogged item $i$'s

$t_{\text{backlog},i}$ = the unit cost for a backlogged item $i$ $[\text{min}]$

$BH_i$ = the backlog correction horizon $[\text{years}]$

Interest rates and inflation are not included in the calculation in Eq. 2. The cost of collecting data required by a certain combination of tools can be calculated according to Eq. 3.

For each tool the data requirements have been defined. Each data item is indicated as *1* if they are required by the tool, *2* if they are optional, and *3* if they are not relevant for the tool. When there is more than one tool in a tool combination, the minimum of the indicators are selected for each data item, when the costs are calculated. Data items indicated as *1* are included in the calculation of $C_{min}$, whereas data items indicated as *1* or *2* are included in the calculation of $C_{max}$.

$$C_{min} = \sum_{i}^{\text{all data required}} \left(c_{\text{data item } i}\right)$$

$$C_{max} = \sum_{i}^{\text{all data required}} \left(c_{\text{data item } i}\right) + \sum_{i}^{\text{all optional data}} \left(c_{\text{data item } i}\right)$$

where:

$C_{min}$ = the cost of collecting all data items that are required by the tool combination $\left[{}^{\text{min}}/_{\text{year}}\right]$ Eq. 3

$C_{max}$ = the cost of collecting data that are both required or optional in the tool combination $\left[{}^{\text{min}}/_{\text{year}}\right]$

**The cost of data** is hence dependent on the cost of setting up and maintaining the tools that treat the data, the number of items occurring each year, the backlog of unregistered items and the unit cost to record these items.

## 3.3 Benefits

The benefits in this model are expressed as informational outcomes. In this context, an informational outcome is defined as an *independent set of information at a certain aggregation level, which can be used to assess a certain aspect of the assets' characteristics or performance*. The informational outcomes are listed in Appendix B. The outcomes will also emerge in the results from the analysis.

The outcomes will be expressed in three different groups:

4. As information items (such as *inventory, hydraulic performance* or *reliability*)
5. As informational benefits of the information items (which of the planning levels (strategic, tactical and/or operational), whether or not it helps to control goal or regulatory compliance, and whether or not the information is instrumental for resource planning).
6. As the planning strategies the information items allow the utility to plan by.

It is when information is used as rationales for planning, that the information receives a true value. Therefore, realised planning strategies (point 3) have been included as benefits. The strategies that have been identified are: (1) reactive, (2) time-based, (3) performance-based, (4) condition-based, (5) criticality-based, (6) proactive reliability-based, (7) proactive risk-based, (8) predictive or LCC-based.

As previously stated, three different groups of expressing the benefit of information has been selected – as information items, as the informational benefits these items give, and the planning strategies the information items enable. Within each of these groups, the benefit will be evaluated as the completeness of all the possible items, as expressed in Eq. 4:

$$\text{Completeness}[\%] = \frac{\sum_{i=1}^{n}(w_i \cdot x_i)}{\sum_{i=1}^{n} w_i}$$

where: Eq. 4

$w_i =$ the weighting factor for benefit $i$ $[0,100\%]$

$$x_i = \begin{cases} 1 & \text{if benefit } i \text{ is present} \\ 0 & \text{if benefit } i \text{ is not present} \end{cases}$$

$n =$ the number of possible benefits in the group

In the end, the three benefit types will be calculated as an aggregated measure, the **Total informational completeness**, which is calculated in the same way as in Eq. 4.

## 4 Model calibration

The parameters that need to be set in the model are:

- The number of data-generating events
- The unit costs for the data classes
- The static tool costs
- Weighting factors

### 4.1 Number of data-generating events

Imagine "Utility X", an ordinary European drinking water utility, providing water to a population of 500 000 people, through a 5 000 km network of water distribution pipes. Table 2 shows some characteristic data for the distribution assets owned by *Utility X*. The data about number of assets and the number of different events every year were all used as input to calculate the data costs of the different data classes in the model.

**Table 2: Characteristic data for *Utility X***

| | |
|---|---|
| Number of clients [#] | 500000 |
| Number of connections [#] | 215000 |
| Network length [km] | 5000 |
| Number of pipes [#] | 65000 |
| Number of valves [#] | 25000 |
| Number of hydrants [#] | 30000 |
| Failure rate [#/km/year] | 0.26 |
| New assets in inventory [#/year] | 2100 |
| Decommissioned asset [#/year] | 210 |
| Repairs / unplanned interventions [#/year] | 1300 |
| Complaints [#/year] | 1000 |
| New connections [#/year] | 2100 |
| Maintenance operations [#/year] | 3500 |
| Flushing [#/year] | 7000 |
| Leak detections [#/year] | 400 |
| Repair history backlog [#] | 1600 |
| Water quality samples [#/year] | 1200 |

### 4.2 Data unit costs

No measured records of the average time spent to collect different data items in water utilities have been found. Therefore, qualified estimations on the unit costs have been made.

For this analysis, it has been assumed that *Utility X* has no backlogged data. The unit costs for each data class are shown in Appendix D2.

### 4.3 Base tool costs

**Table 3: Base costs**

| Tool | Setup cost | Maintenance cost | Base cost [weeks/ year] |
|---|---|---|---|
| Inventory | - | 50 | 50 |
| EPANET | - | 40 | 40 |
| USTORE | 4 | 4 | 8 |
| LCC-software | 4 | 20 | 24 |
| KANEW LTP | 3 | | 3 |
| Aware PLAN | 5 | 10 | 15 |
| Aware-P PX | 2 | 2 | 4 |
| FAIL | 4 | 2 | 6 |
| LLIFE | 4 | 2 | 6 |
| CIMP | 2 | 1 | 3 |
| UNMET | 4 | 1 | 5 |
| IVI | 2 | 1 | 3 |
| Complaint register | 5 | 10 | 15 |
| Diary (w/ water quality) | 5 | 10 | 15 |

As with the data unit costs, base tool costs have also been assigned according to qualified judgements. For *Utility X*, it is assumed that the purchase price for all the different tool modules is zero. The base costs are given in Table 3. One may note that it is assumed that the inventory and the EPANET setup costs have been set to zero, assuming that some sort of inventory already exists.

### 4.4 Weighting factors

Weighting factors have been selected according to subjective judgements. The following rationales have been used to weight the information items:

- All "strategic" information items (the inventory and all district metering area level information) are given a weight of 100 %
- Information items on a cohort level, i.e. information that is obtained through modelling, are given a weight of 50 %
- Information items on an asset level are given a weight of 30 %. An exception is the two risk dimensions (hydraulic and economic), which are actually a composite of cohort and asset level information – these items are given a weight of 50%.

The same rules apply for informational benefits. By applying these rationales one ensures that information that applies at a higher, more general level, are given more emphasis than information that applies at a low level, thus reflecting that information that apply for the whole portfolio of assets is more useful (in a decision-making process) than information that only apply to individual assets.

For the calculation of the **Total information completeness**, the information items are weighted 100 %, and the informational benefits and the planning strategies are weighted 50 %.

The weighting factors used are shown in Appendix D1.

## 5 Results and results discussion

### 5.1 Tool combinations

The spread sheet model is capable of comparing 20 different tool combinations in simultaneously. Two simulations were run to demonstrate the results of the model. The first simulation (*Simulation 1*) is set up to show the wide variety in cost-benefit situations a utility can have. In the second simulation (*Simulation 2*) it is assumed that *Utility X* already has a tool combination in *Simulation 1* (Comb5), and that *Utility X* wants to use the cost-benefit model to assess which tools they should add in order to maximise the benefit for minimal cost. The simulation scheme is illustrated in Table 4.

**Table 4: Simulation scheme**

| **Simulation 1** | Comb1 | Comb2 | Comb3 | Comb4 | Comb5 | Comb6 | Comb7 | Comb8 | Comb9 | Comb10 | Comb11 | Comb12 | Comb13 | Comb14 | Comb15 | Comb16 | Comb17 | Comb18 | Comb19 | Comb20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inventory | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| EPANET |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| USTORE |  |  |  |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |  | 🟩 |  |  | 🟩 | 🟩 | 🟩 | 🟩 |  | 🟩 | 🟩 |
| LCC-software |  |  |  |  |  |  |  |  |  | 🟩 | 🟩 |  |  | 🟩 | 🟩 | 🟩 | 🟩 |  | 🟩 | 🟩 |
| KANEW LTP |  |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |  | 🟩 |  |  | 🟩 | 🟩 | 🟩 | 🟩 |  | 🟩 | 🟩 |
| Aware PLAN |  |  |  |  |  |  |  |  |  |  |  | 🟩 | 🟩 | 🟩 |  |  |  | 🟩 |  |  |
| Aware-P PX |  |  |  |  |  |  |  |  | 🟩 |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |  |  |
| FAIL |  |  |  | 🟩 |  | 🟩 | 🟩 |  |  |  |  |  |  |  |  |  |  | 🟩 |  |  |
| LLIFE |  |  |  |  |  |  |  | 🟩 | 🟩 |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| CIMP |  |  |  |  |  |  | 🟩 | 🟩 | 🟩 |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |  |  |
| UNMET |  |  |  |  |  |  |  | 🟩 | 🟩 |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |  |  |
| IVI |  |  |  |  |  |  |  |  |  |  |  |  | 🟩 | 🟩 |  |  |  | 🟩 |  |  |
| Complaint register |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 🟩 |  | 🟩 | 🟩 | 🟩 |  |
| Diary (w/ water quality) |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 🟩 | 🟩 |  | 🟩 |  | 🟩 |
| **Simulation 2** | Comb21 | Comb22 | Comb23 | Comb24 | Comb25 | Comb26 | Comb27 | Comb28 | Comb29 | Comb30 | Comb31 | Comb32 | Comb33 | Comb34 | Comb35 | Comb36 | Comb37 | Comb38 | Comb39 | Comb40 |
| Inventory | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| EPANET | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| USTORE | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| LCC-software |  |  |  |  |  |  |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |  |  |  | 🟩 | 🟩 | 🟩 |
| KANEW LTP | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| Aware PLAN |  |  |  |  |  |  | 🟩 |  |  |  |  |  |  |  |  |  |  |  |  | 🟩 |
| Aware-P PX |  |  |  |  |  | 🟩 | 🟩 | 🟩 |  |  | 🟩 |  |  |  |  |  |  |  | 🟩 | 🟩 |
| FAIL |  | 🟩 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| LLIFE |  |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |  | 🟩 | 🟩 | 🟩 |  |  | 🟩 | 🟩 |  | 🟩 | 🟩 | 🟩 | 🟩 |
| CIMP |  |  |  | 🟩 | 🟩 | 🟩 | 🟩 |  | 🟩 | 🟩 | 🟩 |  |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| UNMET |  |  |  |  | 🟩 | 🟩 | 🟩 |  |  | 🟩 | 🟩 |  |  | 🟩 | 🟩 |  | 🟩 | 🟩 | 🟩 | 🟩 |
| IVI |  |  |  |  |  |  | 🟩 |  |  |  |  |  |  |  |  |  |  |  |  | 🟩 |
| Complaint register |  |  |  |  |  |  | 🟩 |  |  |  |  |  | 🟩 |  | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| Diary (w/ water quality) | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |

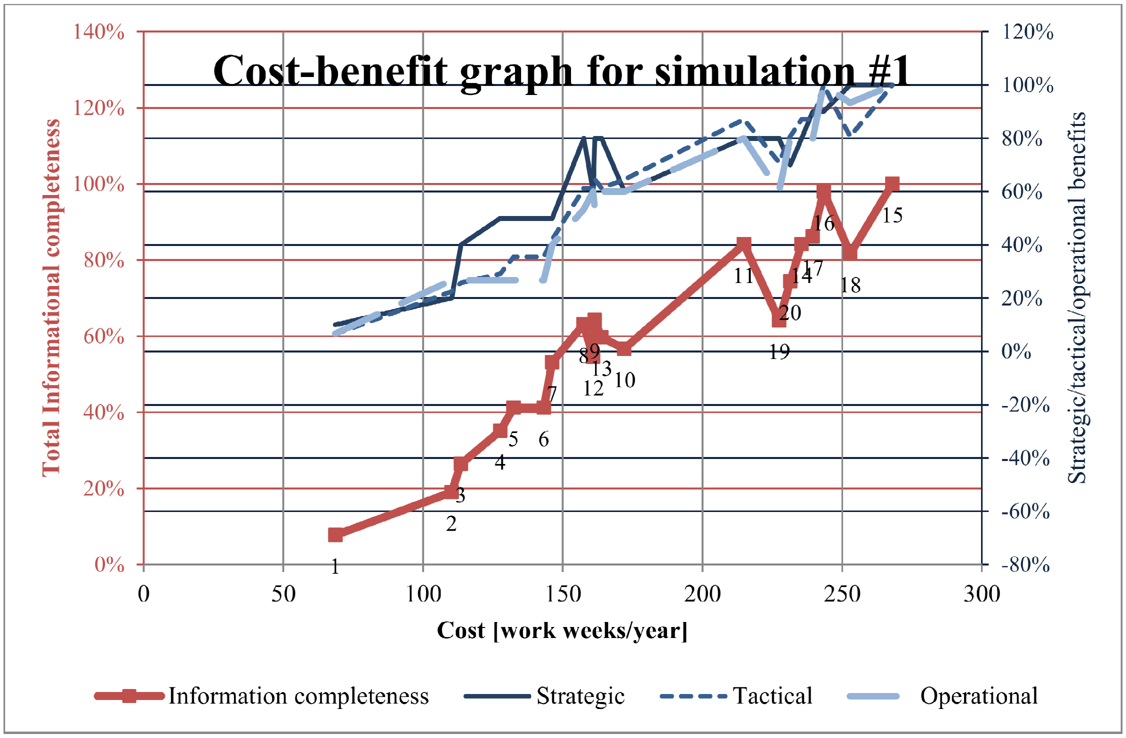


**Figure 2: Results from cost-benefit simulation #1 (the labels refer to the tool combinations in Table 4)**

## 5.2 Tool combination results

**Comments on simulation #1:** The results from the first cost-benefit consideration are shown in Figure 2. Firstly, one may notice that a utility that only utilises an inventory and (alternatively) a hydraulic model (Comb1 and 2) is in a low-benefit situation. The base cost of an inventory and EPANET is quite high, but gives virtually no information about the "health" of the assets. A utility with only an inventory and EPANET is "standing at the foot of the mountain" both with respect to the costs and the benefits.

The combinations from 3 to 6 represent the introduction of KANEW LTP, USTORE and FAIL in different combinations. Not surprisingly, the introduction of KANEW LTP yields an increase in the strategic benefit, without a high cost increase. USTORE gives a greater benefit than FAIL, but USTORE also comes at a slightly higher cost. Combining USTORE and FAIL gives no synergetic effects. Introducing the concept of hydraulic importance (CIMP) to the tool portfolio gives a high increase in the benefits at a low cost (Comb 7). Replacing FAIL with LLIFE whilst also including UNMET, will increase the benefits, mostly at a strategic level because the UNMET allows the utility to assess the level of service (continuity). The Aware-P PX tool represents a low-cost investment (Comb9) that increases the tactical level benefits (PX allows hydraulic performance-based decisions). Adding LLIFE, UNMET, CIMP, PX and KANEW LTP to a tool portfolio with only EPANET and an inventory implies that the data cost will increase by 47 %, but that the benefits will increase from 18 % to 67 %.

Comb10 represents a situation where the utility has invested in an inventory, EPANET and LCC software. Compared to the situation with only an inventory and EPANET, the LCC software represents a major data collection investment (56 %), but also a high benefit increase from 18 % to 57 %. A LCC software is data-demanding, in the sense that failure, repair, inspection, maintenance and decommissioning data are required, both basic data and cost data, but the LCC software does not use this data optimally (for instance by linking reliability to component importance, and thus forming a hydraulic risk measure). Combining the LCC software with the different hydraulic tools in Aware-P and USTORE (Comb11) seems like a better solution. Further, one may see that adding PLAN and IVI to Comb11 (Comb14) does not give any additional benefit. This is due to the fact that many of the outcomes in PLAN and the LCC software are overlapping.

The introduction of diary functions and complaint management are shown in Comb15-20. The complaint register has a lower data cost than the diary, and from the Comb15-20 one may see that the benefits may vary depending on the combinations. Adding only the diary function (Comb16) gives a greater benefit (on tactical and operational level), than only adding the complaint register (Comb17). However, if one wants to keep the high level of informational outcomes at tactical and operational level, and increase the benefits on a strategic level, both the diary and the complaint register must be implemented, even though this implies a higher data cost.

**Simulation #1** shows the wide range in cost and benefits that the different data collection and utilisation. Now, in **Simulation #2** one imagines that *Utility X* in fact has a current tool combination (inventory, EPANET, USTORE, KANEW LTP and the diary), and that *Utility X* is attempting to establish which tools they should add in order to achieve a good cost-benefit yield.

**Comments on simulation #2:** Figure 3 shows the cost-benefit relationship for different additions to *Utility X's* current tool portfolio. From this graph *Utility X* will be able to evaluate the costs and benefits of different data collection and utilisation scenarios. One may see that *Utility* X is in a position of already spending much resources on data collection (193 work weeks/year), whilst only yielding moderate results (58 % total informational completeness). Some good alternatives for *Utility X* are:

- Comb24-26: Increasing data costs by 7.2 %, by adding LLIFE and CIMP to the tool portfolio, will increase the total informational completeness 17 % points (from 60 % to 77 %) for *Utility X*. Alternatively can UNMET (Comb25) and PX (Comb26) be added, for a slightly lower benefit-cost ratio. Including UNMET and PX will give a high yield on the strategic and tactical level.
- Comb30-32: Increasing the costs by 21 %, by adding LLIFE, CIMP and the LCC software, will increase the total informational

completeness to 97 %. Further, by adding UNMET (Comb31) and PX (Comb32) the benefits can be increased to 98 %, for a moderate cost increase (26 % compared to status quo). In these situations there is a balance between the strategic benefit is generally lower than the tactical and operational.

- Comb38-39: Comb38 represent the situation where *Utility X* adds a LCC software, LLIFE, CIMP, UNMET and a complaint register; in Comb39 PX is also added. This tool combination requires a cost increase of 37-39 %, but will render *Utility X* with a total informational completeness of 99-100 %, depending on whether Comb38 or Comb39 is selected.

The model also shows that the most expensive information is information on customer satisfaction, level of service, LCC, and structural condition.

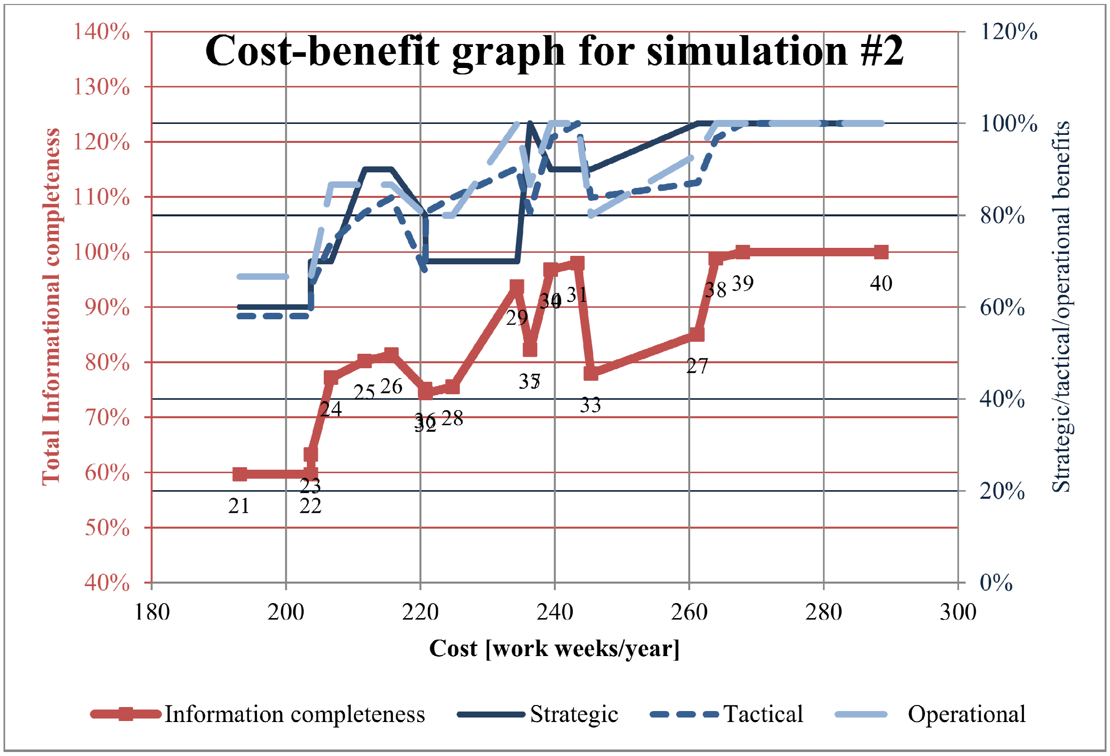


**Figure 3: Results from cost-benefit simulation #2 (the labels refer to the tool combinations in Table 4).**

## 5.3 Data class results

In the previous section, the question of which tool combination is most beneficial was debated. Now the question remains: what data classes are the most useful (has the highest cost-benefit ratio)? The results from the DataComparison sheet are illustrated in Figure 4, where the benefit-cost ratio for each data class is graphed for each planning level, and sorted descending. This is based on data costs only, and does hence not include costs of tools – the idea here is to make considerations that are more independent of the tool combination costs.

As one may see, it is in general the *basic* data that yields the highest benefit-cost ratio. This is not surprising: even though the basic data classes have quite high data costs, they are still the data classes that are most widely used. The *complaint basic* class ranges highest because it is useful strategic information that, if organised right, will have a low cost.

It is interesting to see that the *inspection basic* class has a quite low benefit-cost ratio. This is due to the fact that very few tools actually make use of performance and condition evaluations – it is only LCC-AM/QM that takes inspection data into account. It seems that there is a mismatch between high resources invested in inspection data collection, and low benefits in terms of decision incentives from the inspection data (Murphy et al., 2008). Condition-related information is the “most expensive information”, according to the model.

For most drinking water utilities, most inspection data are leakage detection results – leakage data is usually not utilised to extrapolate information from an asset level to a cohort level. The development of deterioration models for pipes that take inspection data into account could increase the benefits from such data.

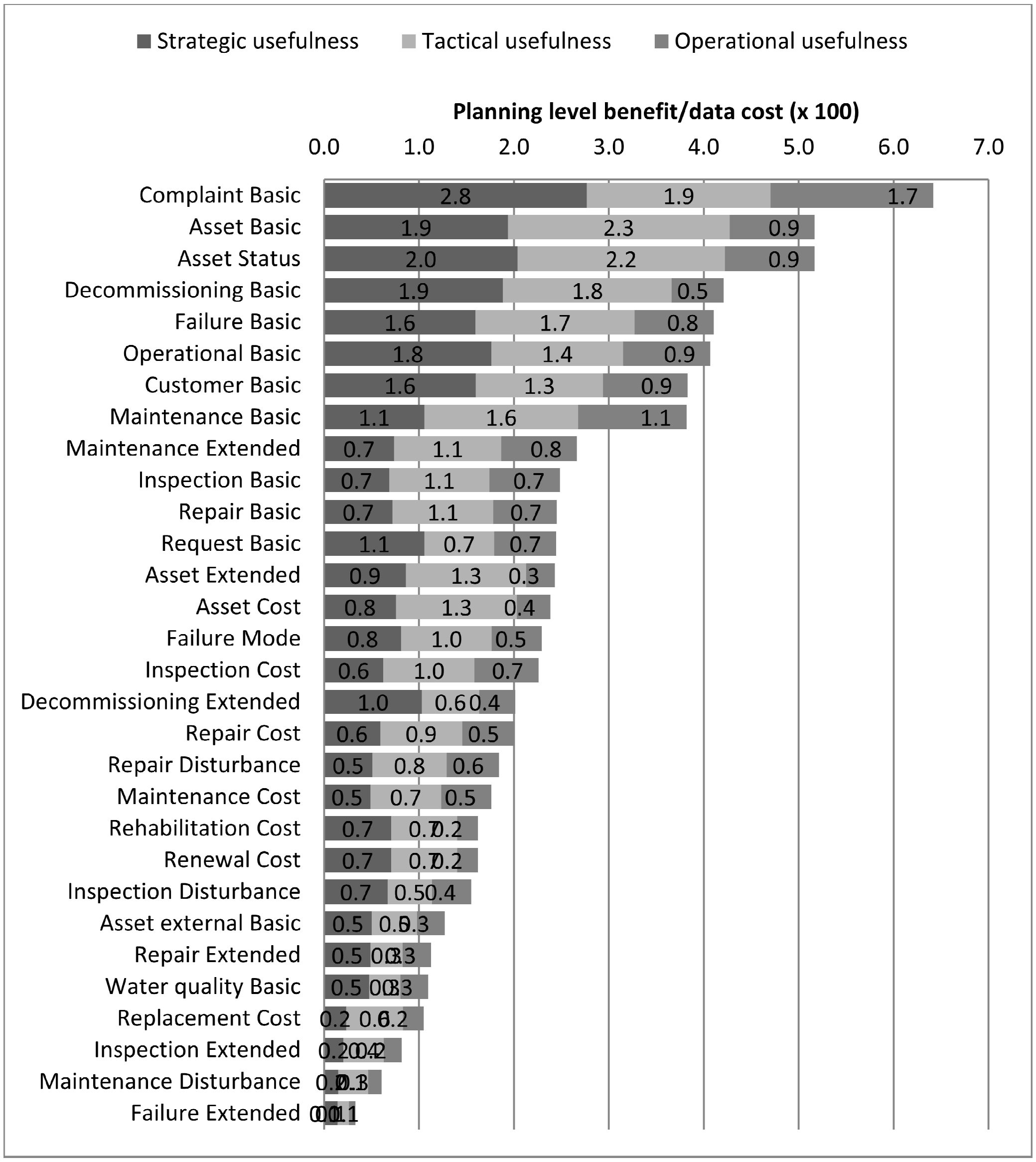


**Figure 4: Data class cost-benefit ratios**

Data about disturbances (duration and number of clients affected) generally score low. However, repair disturbance data score higher, because this data can be used in a wider context.

Further, it is noted that extended data classes in general score low. Although extended data classes can be used, they are seldom a requirement for modelling tools. For instance can extended external factors about a failure (e.g. traffic density) incident *can* prove to be very useful in a failure model, yet it not said that it *is* useful. The usefulness can only be proved by testing.

There seems to be a balance between the strategic, tactical and operational usefulness of the different data classes. The exceptions are the *complaint basic* and the *decommissioning extended* classes. The former has a high strategic impact because it reflects customer satisfaction; the latter has a high strategic impact because the extended decommissioning class contains the reasons for decommissioning, which reflects the strategic issue of *why* assets are replaced.

## 5.4 Sensitivity and uncertainty

### 5.4.1 Sensitivity analysis

The sensitivity of the model was assessed with respect to the following input parameters:

- Weighting factors
- Data unit costs
- Tool costs

The results from the sensitivity analysis for simulation #1 are summarized in Table 5. As one may see the weighting factors (luckily) have a high impact on the different output parameters. This is a positive finding, because it shows that the model is able to convey a cost-benefit consideration based on an array of different priorities that the utility may have, and the user can express these priorities through the weighting factors.

For a 25 % increase for each data class (or each tool class), the model is not alarmingly sensitive; 25 % change in any one of the different data classes, will maximally change the ranking of a tool combination with 2 places. However, for the static tool costs, the rank of a tool combination can change as much as *4* places, which is a large difference when considering only 20 combinations. This is mostly due to the fact that the base cost of tools are quite high, compared to the data cost of the individual data classes. It is therefore considered as more important to achieve an accurate picture of what are the actual base costs (acquiring, set-up and maintenance) of each tool, than to assess unit costs for the data classes accurately, at least for the tools that have a high base cost.

### 5.4.2 Uncertainty analysis

Since the model is not yet based on observed data unit cost values, the uncertainty of the model results is difficult to assess. However, if observed unit costs are included in the model at a later stage, with uncertainties, the uncertainty of the output can most certainly be assessed. For now, uncertainty measures can be based on the minimum and maximum data costs (see section 3.2): the minimum and maximum data costs make out the uncertainty limits. The relative uncertainty measures are calculated according to Eq. 5.

$$\text{Relative uncertainty } [\%] = \frac{C_{max} - C_{min}}{C_{average}} \qquad \text{Eq. 5}$$

On average the uncertainty is 4.6 %. The largest relative uncertainty in tool combination costs is for Comb18 with 7.6 % uncertainty. The tools that contribute most to both the relative and the absolute uncertainty are the diary and the failure models, due to the fact that these tools have many optional data that are quite costly.

The uncertainty results are enclosed in Appendix D5, also for simulation #2; simulation #2 has somewhat smaller sensitivity than #1 due to a more uniform tool selection

# 6 Conclusions

The methodology shown for the cost-benefit assessment suggested in this article represents a step in the direction of formalisation and rationalisation of an evaluation process that should always be present before selecting asset management tools in a utility. The spread sheet model that has been developed is a suitable tool for

**Table 5: Sensitivity analysis results**

| Input parameter | Method | Output parameter | Maximum change | Average change | Most sensitive input |
|---|---|---|---|---|---|
| **Weighting factors** | Weighting factors versus no weighting factors | Rank | 7 | 1.7 | Relative rank weights |
| | | Total information completeness | 5.3 % | 3.2 % | |
| | | Benefit/cost-ratio | 27.8 % | 13.0 % | |
| **Data unit costs** | 25 % increase on each data class | Rank | 2 | 0.1 | Data classes with many entries per year |
| | | Total cost | 4.5 % | 0.3 % | |
| | | Benefit/cost-ratio | 20 % | 0.8 % | |
| **Tool costs** | 25 % increase in base cost for each tool | Total cost | 4 | 0.28 | High acquiring cost tools (inventory, EPANET, LCC) |
| | | Rank | 18.2 % | 1.3 % | |

evaluating and comparing a limited set of tools in a limited set of combinations.

Through the example study of *Utility X* it is shown how such a model can be used to interpret the costs and benefits associated with different data collection and utilisation scenarios:

- Firstly, a utility can use the model to assess its current position with respect to costs and benefits.
- Secondly, a utility can use the model to evaluate which tool combinations that will be most suitable for the utility to invest in (optimal cost-benefit points).
- Thirdly, the model can be used to identify data classes that have yield low benefits compared with the investments associated with them; these results can again be used to assess where efforts should be made to utilise data better.

The results show that it is indeed possible to assess costs and benefits of data collection within a systematic framework, and further that such assessments may be useful for the utilities.

It was found that inspection data yield a low benefit-cost-ratio. Nevertheless is there no doubt that inspection data is a necessary source for assessing actual conditions in water distribution networks. The challenge is to utilise inspection data better. Failure data that is utilised in models generate added value, because it allows the utility to extrapolate and forecast the reliability estimates. If models for extrapolating and forecasting inspection results are developed, the value of inspection data will be much greater. The development of new inspection techniques may also change the benefit-cost relationship for inspection data.

It must still be stressed that data collection is a long-term investment and that, even though certain data classes have a low cost-benefit ratio with the tools that are available today, the development of new tools may invoke greater needs for data sources that have not yielded high benefits in the past.

# 7 Recommendations for further work

The model already has basic capabilities, but the unit data costs still need to be verified through measurements. However, the sensitivity analysis shows that the most prominent factor for costs are the base cost of the tools, thus is it more critical to determine the base costs then the data unit costs. If the utility has good estimates for the base costs and fairly good estimates for the unit costs, the cost-benefit considerations will still be useful.

Another issue that has not been considered in the model is temporal changes. The model works well to predict data costs for a situation similar to status quo, but fails to account for changing conditions in the utility – i.e. changes in failure rates, maintenance intensity etc. If the way of managing the utility changes the costs and benefits of data will also change. Oddly enough, one needs data in order to be able to forecast such changes.

Some suggestions for further development of the model are:

- Better assessment (measurements) of the unit costs for the different data classes
- Including a wider array of different AM tools
- Including temporal change considerations
- Including an automatic sensitivity testing of model results
- Expressing benefits as monetarily

Expressing benefits in monetary units could be achieved in a case study where the reduced economic risk of achieved renewal projects were to be assessed under different decision rationales (which require different data), and calculate the value of data, based on the variation in reduced economic risk between the decision rationales.

The two most important suggestions for further work towards data utilisation in general are: (1) to develop models that can increase the value of inspection data by extrapolation from asset to cohort level and forecasting, and (2) to achieve a better understanding of what extended data classes from asset interventions that are most valuable.