

<p>Not covered as it goes to many new problems</p> <p>I need to include motivation for why I picked real dataset for my experiment</p> <p>Computer forensic, digital forensics (Data corpora, data set, real data corpora, Synthetic data corpora, Forensic corpora, corpora), forensic databases, cybersecurity dataset, information security dataset</p> <p>only I also should document where/how the dataset have been applied.</p>	<p>possible search term: data mining forensics, malware corpora, files corpora, linguistic corpora, password corpora, logs, big data dataset, network traffic, IDS dataset, memory image forensics image, Cyber security datasets, spam corpora, phishing corpora, email dataset, fraud dataset</p>
<p>Similar terminology</p>	
<p>Can divide into categories: real and synthetic and then public/private</p>	<p>Focusing on both big data and non big data datasets?</p>

Public	Digital Corporate Gateway	View this research report
Public	A2Share data	View this research data report data table
Public	Error dataset	View this research data report data table
In-house dataset	Pathways & YourShare reports for AOKS files described in the paper	View this report
In-house dataset	YourShare large-scale collection (working on now) details TBD	View this report

Search ID	Search query (search string within " " and search parameters is comma seperated)	Search resource	Results
1	Digital forensic data set	www.powersploit.com	1 142 000

[See how it works here](#)

(Many of the dataset seems to be under the category of

Public sample (limited access to full)
 By invitation only

[Smartphone dataset \(DataLock vs Morality: A Smartphone Dataset for Cybersecurity Research\)](#)
[VirusShare.com - Because Sharing is Caring \(System currently contains 23,348,476 samples.\)](#)

[http://bigdata.usc.edu/~shankar/](#)
[http://virusshare.com/](#)

[illegible]

compiled list	image and video
G.M.I.E. - Global Medical Emergency Response Initiative http://www.gmie.org/ http://www.gmie.org/en/about/gmie.html http://www.gmie.org/en/about/objectives.html	

[illegible][illegible]

Public	SMS corpus forensic linguistic - NUS corpus	https://oathub.com/3a/1688/nuu-sms-corpus
Public	The Blog Authorship Corpus (Forensic linguistic)	http://csl.mcgill.ca/~loupas/BlogCorpus.htm

[illegible]

Public	Yahoo Password Frequency Corpus	http://pages.southwestern.edu/~password/Frequency_Corpus/vol1.txt
	Search	

Public	Dataset scenarios dataset - Network traffic	https://zenodo.org/record/3890000/files
Public	Khanom Malware Dataset	https://khanom.github.io/2018/05/01/khanom-dataset/

Public	Maffrow - Malware Android application traffic	https://www.cit.uci.edu/infodiv/de/connections/maffrow/
Public	DIT 3MS Spam Dataset	http://www.dit.ie/connections/3ms/3ms-spam-dataset/

Request	current database - http://www.databaseserver.com/secure/secure.htm
Public	password data set http://www.databaseserver.com/secure/secure.htm#password_data_set

Public	Phishing Websites Data Set	https://archive.ics.uci.edu/ml/datasets/Phishing+Websites
By request	Spam Track	https://www.kaggle.com/wordsandnumbers/spam-track

[illegible]

Completed Test

[View Document on the eLibrary website.](#)

[illegible][illegible]

