Big Data meets Open Political Science:

An Empirical Assessment of Transparency Standards 2008-2019

Introduction

We have witnessed a strong push to make political science research more transparent in recent years. There are several reasons for this. For one thing, increased pressure from public authorities is forcing academics and publishers to consider a variety of open-access publication alternatives (see, e.g., May 2005 and Bull 2016). For another, growing awareness of the potential for fraud, manipulation and/or simple mistakes is encouraging political scientists to call for greater transparency (e.g., Elman and Kapiszewski 2014; Lupia and Elman 2014). As a result, journal editors have begun to publish datasets, along with analyses files, so that the resulting publications can be scrutinized and studied more thoroughly. These developments have spearheaded much critical discussion, and this (in turn) has led to more awareness of the need to protect sensitive sources, and to be aware of concerns about propriety/ownership. Consequently, the discipline is engaged in a healthy dialogue about the need for (and limits to) transparency.

At the same time, some of the most popular fields of political science find themselves under increased critical scrutiny. Once powerful approaches to political behaviour have had difficulty predicting a series of recent and surprising election results—e.g., the UK elections of 2015, the Brexit referendum in the UK, and the 2016 and 2020 US presidential elections. These failures have drawn attention to a number of challenges to traditional polling methods and models (Milliband 2016), sparking concerns about the reliability of election surveys, their representativeness (e.g., Bethlehem 2017), and the possibility of social desirability bias (Coppock 2017; Lax et al. 2016), among other things.[1]

Data scientists have raced to fill the resulting void, employing "Big Data" (BD). The digital exhaust being generated by some of the largest and most influential social media sites and search motors (e.g., Facebook, Twitter, Weibo, Google, and YouTube) can be—and has been—used to provide alternative accounts of voter preferences and attitudes. While we can disagree about whether BD will ever be a reliable alternative to surveys, there can be no doubt that these data are new and different, and these differences matter.

Although we think there is significant potential for using BD in political science (PS) inquiry, we are concerned about how its use may challenge the trajectory of mainstream social science. In particular, we are concerned that much of Big Data Political Science (BDPS) may be occurring beyond the critical gaze of practicing social scientists—i.e., it is not being published in traditional PS venues, and that the nature of BD (and the venues in which BDPS are published) has the potential to threaten the trend of greater openness and transparency. After all, what does replication entail with a dataset that includes "billions of interactions" (Blumenstock et al. 2015, 1073), or which is the legal and secret property of a multinational enterprise?

In this paper, we seek to assess the extent to which BDPS articles provide full replication materials (dataset, code, or any other material necessary to verify the empirical analyses). To do this we develop and analyse an original dataset of 1,555 articles drawn from the Web of Science in the period 2008-2019, which are coded for type (theoretical or empirical), research design (qualitative, quantitative, or mixed methods), and transparency. Using these data,[2] we provide an empirical mapping of the growing field of BDPS; this provides us with a survey of where this new research is being published. We then compare trends in transparency among BDPS articles against a sample of core PS articles that *do not* employ BD-- as these articles provide the baseline or benchmark, against which BDPS articles can be compared. Our analysis confirms that much of the new BDPS work is being published outside traditional PS journals and does not adhere to the transparency standards in the discipline.

The remainder of the paper is structured as follows. We begin with a mapping of recent developments in political science, to demonstrate our increased reliance on both transparency (especially in large-N studies) and BD. This discussion is followed by a description of our research design, including the construction of our dataset. In the subsequent section we outline the main trends in BDPS and assess the level of transparency relative to conventional PS articles. The final section includes a discussion and some concluding remarks.

Transparency and Big Data trends

The past two decades have seen a growing concern about transparency in political science, but the need for, and desirability of greater transparency is not new. Already in 1995, *PS* published a symposium on replication in the social sciences, which addressed issues of transparency (September 1995 issue), and which concluded with a varied and influential group of contributors promoting replication and transparency as essential elements to study science.[3] In 2003, another symposium—this time in *International Studies Perspectives* (Bueno de Mesquita et al. 2003)—took a closer look at the state of replication in the (sub) discipline, and the authors were clearly disappointed by the lack of progress. By 2001, a little less than 1/5 of political science journals had some sort of replications policy.[4]

Concerns for transparency have accelerated in the wake of several high-profile examples of fraud, error, and deceit. Over the past decade, the media have revealed stories about the politically-convenient sloppiness of Reinhart and Rogoff (2010; cf. Herndon et al.

2013) and the astonishing example of deceit and manipulation by LaCour and Green (2014; cf. Broockman et al. 2015 and Singal 2015). As a result, there is growing recognition among political scientists about the need and desire for greater transparency (for recent reviews, see Elman et al. 2018 and Laitin and Reich 2017; see also Miguel et al. 2014). Aware of the costs of being associated with 'fake' science, political scientists doubled down: a new transparency movement took hold, in which mainstream PS journals aimed to increase transparency and facilitate replication (Janz 2018).

In response, a growing number of political science journals have embraced the need for greater transparency, replication (and the data depositories this would require), and pre-registration. This movement was sparked by the 2010 decision of an ad hoc Committee of the American Political Science Association (APSA) to launch the Data Access and Research Transparency (DA-RT) group. This group approached leading PS journal editors in an attempt to have them sign their Journal Editors Transparency Statement (JETS).[5] Consequently, 29 PS journals—including *EPS*—have committed themselves to greater data access and research transparency (see the Online Supplement Material (OSM), D.1) and to implement policies that would make their publications more transparent and accessible. This public push for greater transparency can be seen across the social sciences, and in journals that extend beyond the JETS list.[6]

At the same time, we have seen a phenomenal rise in political scientists' use of "Big Data" (BD) to generate alternative measures of individual preferences, attitudes, and behaviour. We provide evidence of this rise below, in Figure 4. Drawing from a wide variety of technologies—e.g., data-processing hardware and software and a plethora of digitized apparatuses—and a remarkably broad array of sources (e.g., public, commercial, proprietary), political scientists have been able to map multiple aspects of political life by wading through the data produced by these many technologies and sources.

The most famous, or infamous, example comes from the Cambridge Analytica scandal, where millions of Facebook users found their data was being used, without their knowledge, to aid the political campaigns of conservative candidates in the 2016 US presidential campaign, including Donald Trump (see, e.g., Wylie 2019 and Berghel 2018). This example is just one of many political analyses employing data from numerous platforms (e.g., Twitter, Facebook, Weibo, Google, Flickr, YouTube…) and techniques (e.g., automatic content analysis; scraping/web crawling; network analysis; sentiment analysis and topic modelling; machine learning, etc). See OSBM: B.3 for an overview of platforms and approaches uncovered in our research.

While it should not be controversial to note the rise in BDPS research, it is difficult to confirm empirically. While it is common to define BD with reference to 3Vs (velocity, volume, variety),[7] this definition is hard to transform into something measurable. Our solution is to focus on the way that BD employs large, repurposed, datasets.[8] This repurposing provides social scientists with access to (e.g.) sensor data, satellite and measurement data, enormous digital archives, social media exhaust and geotags, that can be "taken" from their original mission and repurposed for subsequent social scientific analysis. While neither large datasets or repurposing data is new to political science, the scope of this BD collection is novel, as is the fact that the datasets produced are often too large for traditional approaches and programs.

Such a definition taps into the new tools that are being used to repurpose data (e.g., scraping and machine-learning that allow us to use "digital footprints" in real time, over a number of platforms), and reflect on the enormous scale (hence "Big") that this repurposing provides (in terms of volume, velocity and variety). But it also introduces at least three new challenges for political scientists that employ Big Data, each of which are relevant to our recent embrace of transparency.[9]

The first challenge concerns *proprietary control*. Many BD companies (e.g. Facebook, Google…) keep their data and algorithms as proprietary trade secrets, resulting in analytical black boxes, over which subsequent users have very little control or insight. In effect, data scientists often work with data controlled by a monopoly holder. Buyers/users of the data do not know the algorithms that go into their parsing and cannot know the details of how that data was created (Dalton and Thatcher 2015, 6).

Another important challenge concerns the *selection criteria* or sampling techniques. Traditional social science data collection is often publicly funded, and data is generally collected with the explicit aim of securing a representative sample from a clearly defined population so findings can be generalized. Think of census data, and the effort that lies behind its collection. Unlike traditional social science data, some forms of BD do not need (and make no effort) to be representative, and we know that the collected data are often strongly biased with regard to class, language, and use of technology.

This uniquely commercial nature of the data, in turn, raises several additional issues related to replication and access. The way that BD is often gathered and used makes it nearly impossible for researchers to emulate (forget about replicate) the approach and its results (Longley 2012). Both the structure of the data, and the way they are accessed (through networked streams of data), determine what can be known from that data (Burgess and Bruns 2012). Most firms are leery of sharing their data, whatever the cost—and these costs can be substantial. Independent researchers, outside of key companies that control the data, do not have access to the core proprietary algorithms that process and interpret the data.[10] They are forced to navigate blind.

Our third and final concern lies at the heart of this paper: we are worried that the nature of BD (and the venues in which BDPS are published) *threatens the trend of greater openness and transparency*.

Research design

Our research proceeded in three steps, as outlined in Figure 1. We began by identifying articles as being examples of Political Science (PS), Big Data (BD) or both (BDPS). We then screened the results, to ensure the quality of the resulting samples. Finally, we directed those articles deemed to be BDPS (n=355) into our working database, and those articles that were classified as PS but not BD (n=1,200) into a baseline for comparison. The existence of this benchmark will allow us to compare the relative level of transparency in traditional PS vis-à-vis BDPS articles. The remainder of this section describes how the two search strings were developed and how we proceeded to construct the two datasets from the search results. More details are provided in the OSM. The details of the benchmark are described below, in a subsequent section

Figure 1 about here

Assessing the degree of transparency in the new BDPS literature presents several empirical challenges. The most fundamental challenge is to operationalize both BD (and non-BD) and PS in a consistent manner. To the extent that there is an emerging BDPS literature, it may be conducted by data scientists with no formal background in political science and published outside traditional political science journals.

Figure 2 illustrates the problem. To construct our dataset, we essentially had to identify, from the pool of *all* scientific articles, those articles that employ BD (horizontally-striped oval) to address a political science question (vertically-striped oval). The resulting overlap (the hatched intersection) constitutes the world of BDPS.

Figure 2 about here

For practical purposes, we limit ourselves to journal articles indexed in the Web of Science, published in 2008-2019 and written in English.[11] This database represents the pool of all scientific articles illustrated in Figure 2. Because we are searching for PS articles that are not necessarily published in traditional PS venues, we needed to develop an independent means of categorizing work in the field of political science. Consequently, we developed two different search strings to identify PS and BD articles. The first search is for articles that included *"political science" AND "big data"*, as described below. This search serves to identify the hatched intersection of BDPS articles in Figure 2. The raw search result (n=8,745) was then manually coded, and the majority of articles were excluded as irrelevant (not PS, not BD). We refer to the set of included articles as the BDPS dataset (n=355).

The second search was for articles that included *"political science" NOT "big data"*. For this search, we limited ourselves to articles published in journals categorized as political science journals in the Web of Science. The search results represent the vertically-striped area in Figure 2. From this search we drew a quasi-random sample of articles, stratified by year (n=1,200). This sample serves as a baseline against which the BDPS data is compared. Both searches were conducted as a "topic search" (TS) in Web of Science, which returns results from the article title, abstract, and keyword fields.[12]

The PS search string

Because we are looking for works in political science that *may not* be published in traditional PS journals, we cannot rely on the most common means for defining the discipline. Most of us understand "political science" to be work that is produced and discussed in explicit PS communities, such as PS conferences and/or PS departments, and/or work that is published in journals classified as PS journals. As we sought to identify all BDPS articles, independent of journal outlet, we had to operationalize PS in a Boolean search string, without referencing or selecting from the underling publication venue. To develop the search string, we surveyed the most influential journals in PS, sampled a set of abstracts from each journal, collected the most frequently used words from the selected abstracts, and combined the words in a search string.

To identify key journals in the discipline, our starting point was the journal rankings provided by Giles and Garrand (2007, table 2), but we updated these to include the SCImago Journal Rank (SJR) indicator.[13] This provided us with a total of five rankings for comparison (see OSM: A.1). Twenty-two journals appeared in at least two of the rankings. From these, we selected all journals that were included in at least three of the rankings, i.e., the top 13 journals. In an effort to avoid a bias towards international relations and North American journals, and to ensure sufficient variety in search terms, we added four journals that were included on two rankings,[14] as well as two renowned journals in important sub-fields in the discipline.[15] Table 1 shows the list of journals from which we sampled abstracts. Although readers may differ about the inclusion (or exclusion) of any particular journal on this list, we think it offers a fair summary of mainstream PS.

Table 1 about here

We sampled abstracts from the first ten empirical articles published in 2017. To avoid oversampling of specific issues, we included only the first article in special issues. These 190 abstracts (19 x 10) were extracted and then analysed for salient terms using NVivo. Salient terms that did not describe a PS topic (e.g., 'data', 'survey', 'and', and 'or') were excluded. Figure 3 shows the resulting word cloud of the 50 most common words from our sample of abstracts with terms that cover the discipline of political science.[16]

Figure 3 about here

The corresponding list of most frequent terms was then analyzed for words that could be combined using stemming or wildcard techniques (e.g., 'support* (supporter, supporters, supported')). In the end, we selected 22 most relevant terms/stems: *soci*, politic* OR policy OR policies, state OR states), conflict$, (election$ OR elector*), economic$, media, countr*, war$, govern*, party OR parties, democra*, institution$, civil OR civic, power, citizen$, gender,* and *nation*, interest$, labo$r,* and *opposition$.*

To maximize the possibility that a given article indeed addresses a PS question, we operationalize "PS" as an article in which the "topic", i.e., title, abstract, and keyword field, includes at least *three* of the terms listed above, e.g. *("soci* AND democra* AND institution$) OR ("soci* AND democra* AND power) OR …*, and so forth. Doing so yields 1,771 possible combinations of terms that were used in the search string, as documented in OSM: A.2.[17]

The BD search string

Operationalizing BD in a consistent manner was more difficult than operationalizing PS, as we could not draw on an established literature or foundation to the same extent. The common definition of BD with reference to 3Vs (velocity, volume, variety) is exceedingly hard to transform into something measurable. As a result of this, our "BD" search is much less precise than the "PS" search. Consequently, most of the work to identify "true" BDPS articles took place after the search had been conducted, through a manual coding of all the articles that resulted from the search (see below). Through a series of trial and error, we limited our BD search string to simply include *'social media' OR 'twitter' OR 'facebook' OR 'google' OR 'algorithm' OR 'Web 2.0'*. We recognize this operationalization process produces a rather truncated set of BD articles, in that it selected papers that repurpose data collected from primarily big tech firms. We think this tendency is predominate in the literature but hasten to note that the sample was not limited to papers that draw data from such firms. While the search string could certainly have been more comprehensive, it still yielded a very large result. A broader search would have resulted in an unfeasibly large pool of records that had to be screened manually.

The records resulting from the search *"political science" AND "big data"* (n=8,745)[18] were downloaded and cleaned for duplicates (n=91). The remaining records (n=8,654) were subsequently screened for eligibility: to what extent did a given article address a PS research question using BD? A large share of excluded records – more than 95% – contain both records that clearly do not address a PS research question or records that *do* address a PS research question, but without using BD. In the end, 355 records were deemed valid cases of BDPS. These observations were then coded for transparency (see below) and constitute the BDPS dataset.

Screening the 8,654 retrieved records for inclusion in the BDPS dataset was done in several iterations. As already noted, it was difficult to develop a search string that adequately captured BD, which means that the bulk of the work to classify a BDPS article took place at the screening stage (see Figure 1). As discussed above, we departed from the 3V (velocity, volume, and variety) to focus on *repurposing* and *size*, i.e., the datasets are usually too large for conventional statistical modelling.

To develop the coding practice, we undertook a pilot coding where both authors coded the same random sample of 100 abstracts. Subsequently we compared the coding results and discussed discrepancies until we agreed on all coding decisions. A research assistant (RA) was then employed to code all cases manually. We used a single coder here to ensure consistency. All observations were then coded into one of four mutually exclusive categories: *Include* (the article is both PS and BD), *doubt* (needs further reading), *exclude* (PS, but not BD) and *irrelevant* (completely unrelated). [19] Nearly 60% of our results were found to be PS but not BD, (*exclude*) and another 31.8% were deemed completely irrelevant.[20] We provide more details about this screening in OSM: B.3.

For the included observations, the RA coded additional information about the type of data (source and size) employed in the analysis. When the RA had finished, the authors went through the doubts and recoded them into *include* (BDPS) or *exclude* (not BD and PS). The *include* observations constitute what we refer to as the BDPS dataset.

Constructing the benchmark group

In order to gauge the relative transparency of this new BDPS work, we need to establish the level of transparency enjoyed by other types of published work in political science. To do this we created a benchmark or baseline for comparison: we selected a random sample of non-BD political science articles (using the same PS selection criteria noted above). We then had two new coders trace their degree of transparency, following the procedure described below. The results provide us with a baseline transparency level for all political science articles published, against which we can compare the BDPS transparency levels.

As noted above, our second Boolean search was for articles where the topic included *"political science" NOT "big data"*, published in journals classified as PS in Web of Science. This search yielded 44,014 observations from 171 different journals (see OSM: C.1.). The large number of records made it necessary to draw a sample to be able to code the observations for transparency. In the next step, we drew a representative sample of articles, stratified by year (n=1,200). For each publication year, we sorted the observations by relevance[21] and selected the 100 first articles that were published each year (100 x 12 years).[22] These records constitute the observations in the benchmark or baseline against which the BDPS data is compared, and were subsequently coded for transparency and merged with the BDPS dataset.

Coding transparency

Having assembled the two datasets – the BDPS dataset and the benchmark – two RAs coded all observations in both datasets for transparency. Both datasets were randomly split in two and assigned to an RA, so that each RA coded half of the observations in each dataset. These coders were trained to read and code the materials in a way that makes subsequent replication possible (see OSM: D.2). In particular, each RA followed a four-step process to code each article (see OSM: D.3 for details):

1. The article was traced back to its publication site to search for supporting replication files;
2. The article was skimmed to determine whether it contained empirical analyses (or if it was a theoretical contribution, a literature review, or similar);
3. Empirical articles were further coded as quantitative, qualitative, or mixed methods depending on the research design employed; and
4. Empirical articles were subsequently coded across five categories (yes or no):
    a) Dataset is available;
    b) Code or script to reproduce analyses is available;
    c) Article states that the replication material is available;
    d) Replication material is available on request; and
    e) Author(s) state that there are restrictions on the shared material (e.g., ethical concerns).

In short, coders conducted a thorough search of each article publication site, as well as the homepage of the corresponding author, to search for replication materials. Coders also combed through the article drafts to search for explicit references to how (or if) replication materials were available on request, or at an explicit third-party site.

Comments and doubts were then checked and revised by the authors. The two complete datasets were subsequently combined, and a dummy variable indicates whether an observation is *BDPS* or belongs to the benchmark. Finally, we added two dummy variables at the journal level: *PS journal* (yes or no)[23] and *JETS signatory* (yes or no),[24] and checked for reliability (see OSM: D.4). To measure *transparency*, we created a dummy variable that takes the value of 1 if all necessary replication material is available (both dataset and code/script).

In the analysis, we focus exclusively on empirical (i.e., quantitative, qualitative, and mixed research designs) political science. This excludes 181 observations that do not contain any form of empirical analysis (e.g., theoretical work). We believe this constitutes the most correct benchmark for comparison, as the BDPS articles are empirical (and quantitative) by definition. In some of the analyses, we also limit ourselves to articles with a quantitative research design only, as this group most closely resembles the BDPS articles.

Results

Our analysis is mainly visual, and we rely largely on predicted transparency scores with corresponding confidence intervals.[25] To corroborate the visual analysis, we rely on a series of logistic regression models that allow us to test our propositions more formally. The tests also include a model using coarsened exact matching (Iacus et al. 2012), where 'treated' (i.e., BDPS) observations and the benchmark are matched on publication year, journal, and coder id (i.e., which of the two RAs coded the article for transparency).

The section presents our results in two parts. First, we describe trends in BDPS publishing during the last decade. Then we describe transparency trends in the benchmark, i.e., conventional political science.

Trends in BDPS publishing 2008-2019

Figure 4 displays when and where the identified BDPS articles were published. Here we are struck by a sharp increase in the total number of BDPS articles from 2014 and onwards. In the first part of the period, we find hardly any BDPS articles at all. Notably, the first BDPS article in our material appears in 2009. Before 2014, there are only 4 cases of BDPS in our dataset.

Figure 4 about here

Secondly, we note the relative irrelevance of traditional PS journals. In particular, it would seem that the rapid growth in BDPS publications is essentially taking place outside of the discipline's core journals. Of the 355 observations, only 54 records, or about 15 percent, were published in journals classified as PS; the remaining 301 records were published in journals not categorized as PS. In total, only 19 articles were published in a JETS journal.

A closer look at the non-PS journals reveals significant heterogeneity. The 301 articles are published in 152 different journals. In our material, the six most common publication outlets published about 21% of all BDPS articles, and about 25% of BDPS articles in non-PS journals. Table 2 lists the six most frequent PS and non-PS journals.[26]

Table 2 about here

It is worth noting, however, that even if most BDPS articles are published outside of mainstream PS outlets, many are published in interdisciplinary journals or in social science journals, broadly defined.[27] We doubt that many political scientists are regular readers of *Social Science Computer Review* and *Plos One*—but these are clearly the most fruitful outlets for this type of work, with a focus that is obviously tangential to the discipline. Producers and consumers of BDPS would be well advised to follow these outlets closely. On the other hand, a number of BDPS articles appear in journals far from the typical political scientist's selection of staple journals (e.g., *EPJ Data Science, IEEE Access*, and *Nuclear Engineering and Technology*, to mention but a few).

Next, we turn to transparency among BDPS and non-BDPS articles.

Transparency among BDPS and non-BDPS articles

Starting with the benchmark group, Figure 5 shows transparency trends among empirical articles published in PS journals. We break down the results to focus on articles with a quantitative (including mixed methods) research design and articles published in so-called JETS journals.

The trends in Figure 5 are clear: While there seems to be an overall increase in transparency—defined by the availability of full replication materials—in empirical PS, the increase is mainly driven by quantitative work. This is as expected, given that it is easier (practically as well as ethically) to share quantitative replication materials. The finding also echoes the concerns voiced about non-quantitative data and confidentiality that we described above. More interesting, however, is the gap between journals that have signed the JETS statement and those that have not. Among the non-BD articles with a purely quantitative research design, complete replication materials were provided in 35% of the cases. If we distinguish between JETS and non-JETS journals, the share is about 66% and 16%, respectively. In other words, compared to articles in non-JETS journals, JETS articles are about four times more likely to provide full replication materials.[28]

Figure 5 about here

As argued above, the best baseline for evaluating the transparency of BDPS articles is a subsample of observations that employs purely quantitative research designs. In the remainder of the analysis, we therefore limit ourselves to this group, while also keeping in mind the difference between JETS signatories and other PS journals.

Figure 6 compares transparency trends among BDPS articles and non-BDPS articles, as well as for the subset of non-BDPS articles that are published in JETS journals.[29] The figure clearly demonstrates that BDPS articles are less likely to provide full replication materials. If anything, this discrepancy is growing over time as PS journals increasingly expect replication materials to accompany the published work.

Figure 6 about here

As Figure 5 showed, there is a notable difference between PS journals that are part of the JETS initiative and those that are not. In the lower panel of Figure 6, we split the benchmark group in two, depending on whether the article is published in a JETS journal (or not). The results indicate that while articles published in non-JETS PS journals do not follow the same strict transparency standards (as articles published in JETS journals), they are more likely to provide replication material than the average BDPS article.

To corroborate the visual analyses, we conducted a series of more formal tests using logistic regression. The results are displayed in Table 3. Model 1 is the simplest model which estimates the difference between BDPS and non-BDPS articles controlling for publication year only. Model 2 repeats Model 1, but with standard errors clustered by journal. In Model 3 we also control for whether the journal is a JETS signatory. Model 4 is identical to Model 2 except it excludes non-empirical work, while Models 5 and 6 are limited to articles with a purely quantitative research design. In Model 6, we employ coarsened exact matching, matching observations by publication year, journal, and coder identity. All models confirm the results already illustrated above. No matter how the model is specified, BDPS articles are significantly less likely to provide full replication material. This finding holds also when controlling for JETS journals (Model 3). As expected, the difference between BDPS and mainstream PS articles increases if we focus on empirical work, and particularly purely quantitative work.

Table 3 about here

Discussion

Our study illustrates some of the opportunities and challenges associated with extracting data from the Web of Science and similar publication databases, and our results are both surprising and worrisome for the discipline. To begin with, we demonstrate a significant rise in BDPS after 2014. Much of this new research is *not* being published in traditional PS journals. This trend may offer political scientists greater opportunities to collaborate across disciplines, as long as we are aware of where this work is being published. There is clearly much potential to leverage technical and methodological expertise in ways that provide new approaches to solving old problems. The opportunity to introduce a broader diversity of perspectives to both BD and PS is clearly exciting (see e.g., D'Ignazio and Klein, 2020). But it is also clear that much of this work is being published beyond the gaze of mainstream political science, and we need to ensure that this work receives the critical attention it deserves.

At the same time, we demonstrate that much of the BDPS work does not abide to the high transparency standards that our discipline has tried to encourage over recent decades. Indeed, our results indicate there is substantial variation in the employment of the discipline's transparency standards—both within the mainstream discipline journals (generally); and relative to BDPS articles (in particular). The first variation, within mainstream PS journals, is not particularly surprising and can even be expected: there is significant variation in commitment to transparency and replication across articles that are empirical and quantitative, as opposed to theoretical and/or qualitative. Recognizing this, we have been very careful in how we have defined our control cases (i.e., with an eye on focusing mostly on similarly empirical and quantitative approaches). In the doing, we uncovered substantial variation in commitment to the needs of transparency and replication across articles published in JETS (as opposed to non-JETS) journals. We are both surprised and bothered to find that BDPS articles are significantly less likely to be accompanied by full replication materials.

We conclude by hoping that our findings can spark a discussion within the discipline about how to deal with the transparency challenge in the face of a growing wave of BDPS research.

Notes

---

[1] In the wake of the 2020 US Presidential elections, there was already much media speculation about what might have gone wrong. See, e.g., Bump (2020), Cohen (2020) and Tufekci (2020).

[2] Replication material for the paper can be found at [redacted]. Scholars who disagree with our coding choices are invited to dialogue, and we will update the data accordingly.

[3] Gary King's (1995) contribution is probably the best known of these.

[4] Gleditsch and Metelits (2003) list 27 journals in political science. See https://academic.oup.com/isp/article/4/1/72/1930641#29683436 for a list and brief discussion.

[5] See the Journal Editors Transparency Statement, https://www.dartstatement.org/2014-journal-editors-statement-jets.

[6] It should be noted that the move toward greater transparency is not embraced by everyone, and a long string of prominent political scientists, and earlier presidents of the APSA, wrote a letter to journal editors to ask them to think carefully about signing the JETS. See Powell et al. (2016).

[7] See Laney (2001).

[8] More detailed coding rules and procedures will follow and can be found in the OSM.

[9] We have a fourth concern, but it is not directly linked to the discussion of transparency. This is a concern about the role of *theory* in a context of apparently boundless reams of data. Some BD advocates suggest that the sheer amount of data can mean an "end of theory" (e.g., Anderson 2008). In other words, this quantitative bonanza can be used as an excuse to avoid theory, and the theoretical awareness that usually accompanies a trained social scientist. To the extent that BD lends itself to naïve induction, it represents yet another challenge to contemporary social science.

[10] We would be negligent if we didn't mention Social Science One in this context: Harvard's attempt to access and share Facebook data. See https://socialscience.one/blog/social-science-one-announces-access-facebook-dataset-publicly-shared-urls. For a more critical view, see Tromble (2021) and Dommett and Tromble (2022).

[11] Our first search covered 2008-2018 but was later updated to also include 2019.

[12] See https://clarivate.libguides.com/woscc/searchtips.

[13] This platform takes its name from the SCImago Journal Rank (SJR) indicator, developed by SCImago from the widely known algorithm Google PageRank™. This indicator shows the visibility of the journals contained in the Scopus® database from 1996. See Jensen and Moses (2020) for a discussion of challenges associated with using SCImago to rank PS journals.

[14] *Journal of Political Economy*; *Political Geography*; *British Journal of Political Scien*ce and *Comparative Politics*.

[15] *Political Psychology* and *Political Communication*. *Political Psychology* appears in one of the rankings.

[16] For a similar approach, see Cooper et al. (2009).

[17] Ideally, we would have combined more than three words and included more terms. However, through trial and error we found that this search string was the most complicated that the Web of Science search engine could handle. Even so, we had to conduct the search in several iterations.

[18] For a distribution of the results over time, see OSM: B.1.

[19] For examples of records coded as irrelevant and excluded, see OSM: B.2.

[20] Given the tedious nature of this work, we experimented with machine learning to develop an algorithm that could take our manually-coded results and use them to learn how to read an abstract and decide for itself when it

employed "PS" and "BD". In the end, the experiment failed, as the training dataset included a share of BDPS observations that was too small for the machine's needs. For more information, see OSM: B.4.

[21] Recall that our search string was generated by choosing the most common words used in the abstracts from a list of prominent journals in the field. By giving preference to "relevant" records, we choose those articles that best meet our definition/operationalization of PS. We think this is an appropriate approach, as it allows us to focus on what we might think of as the gold standard in the discipline, from the most cited journals. To the extent that our definition (and, quite possibly, the discipline) prioritizes quantitative approaches, then the resulting benchmark group will tend to be biased in the direction of including a higher share of transparency articles than the profession at large. In other words, the benchmark group is likely to include a higher share of transparencies—providing a higher benchmark for comparisons with the BDPS data. As noted by one of the anonymous reviewers, we recognize that a stronger conceptual framework—one that does a better job of linking the intersection of actual data qualities and our hypotheses—could have made our search and our failed attempt at machine-reading more viable.

[22] Given the limitations in the Web of Science search platform, a completely random sample would have been very cumbersome. Web of Science only allowed us to download 50 records at a time, so downloading 44,014 observations would require about 880 separate downloads.

[23] Extracted from Web of Science.

[24] Source for the JETS coding was the DA-RT statement, available at https://www.dartstatement.org/2014-journal-editors-statement-jets; see also OSM: D.1.

[25] We use Stata's *twoway qfitci* function, which plots predicted values of the dependent variable (in this case, transparency) as a linear function of X and X^2 (here: year and year^2) with a corresponding confidence interval.

[26] An extended table with all journals in our material is available in OSM: E.1.

[27] We recognize that the Web of Science categorization of 'political science journals' can appear somewhat arbitrary. For reasons of consistency, we choose to rely on Web of Science's classification throughout the data collection and analysis.

[28] For detailed descriptive statistics, see OSM: E.2.

[29] Note that the first BDPS article in the dataset appears in 2009. The wide confidence interval for the BDPS articles during the first part of the period reflects the fact that there are hardly any observations in the dataset prior to 2012.

References

Anderson, C. 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 23 June.

Berghel, H. 2018. Malice domestic: The Cambridge analytica dystopia. *Computer* 51(5): 84-89.

Bethlehem, J. 2017. The Representativity of Election Polls. *Statistics, Politics and Policy* 8: 1-12.

Blumenstock, J., G. Cadamuro and R. On. 2015. Predicting poverty and wealth from mobile phone metadata. *Science* 350(6264): 1073-6.

Broockman, D., J. Kalla and P. Aronow. 2015. Irregularities in LaCour (2014). 19 May. Available from: http://stanford.edu/~dbroock/broockman_kalla_aronow_lg_irregularities.pdf. (accessed 14 March 2018).

Bueno de Mesquita, B., N.P. Gleditsch, P. James, G. King, C. Metelits, J.L. Ray, B. Russett, H. Strand, and B. Valeriano. 2003 Symposium on Replication in International Studies Research. *International Studies Perspectives* 4: 72–107.

Bull, M. J. 2016. Introduction; open access in the social and political sciences: threat or opportunity? *EPS* 15: 151-7.

Bump, P. 2020. It’s important to ask whey 2020 polls were off. It’s more important to ask what will happen next. *The Washington Post*, 16 November.

Burgess, J. and A. Bruns. 2012. Twitter archives and the challenges of ‘Big Social Data’ for media and communication research. *M/C Journal* 15(5).

Cohen, N. 2020. What went wrong with Polling? Some early theories. *The New York Times*, 10 November.

Cooper, C., T.A. Collins, and H.G. Knotts. 2009. Picturing political science. *PS* 42: 365-365.

Coppock, A. 2017. Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment. *Statistics, Politics and Policy* 8: 29-40.

Dalton, C. and J. Thatcher. 2015. Inflated granularity: Spatial ‘Big Data’ and geodemographics. *Big Data & Society* (July-December): 1-15.

D'Ignazio, C. and L.F. Klein. 2020. *Data Feminism*. Cambridge: MIT Press.

Doctorow, C. 2008. Big data: Welcome to the petacenter. *Nature* 455: 16–21.

Dommett, K. and R. Tromble. 2022. Advocating for Platform Data Access: Challenges and Opportunities for Academics Seeking Policy Change. *Politics and Governance* 10(1): 220-229.

Elman, C. and D. Kapiszewski. 2014. Data access and research transparency in the qualitative tradition. *PS* 47: 43-47.

Elman, C., D. Kapiszewski and A. Lupia. 2018. Transparent Social Inquiry: Implications for Political Science. *Annual Review of Political Science* 21: 29-47.

Giles, M.W. and J.C. Garrand. 2007. Ranking Political Science Journals: Reputational and Citational Approaches. *PS* 40: 741-51.

Gleditsch, N.P. and C. Metelits. 2003. Symposium on Replication in International Studies Research. *International Studies Perspectives* 4: 72-107.

Herndon, T., M. Ash and R. Pollin. 2013. Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff, April 15. PERI Working Paper No. 322.

Iacus, S.M., G. King and G. Porro. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20: 1–24.

Janz, N. 2018. Replication and transparency in political science—did we make any progress? Political Science Replication blog. 14 July. Available from: https://politicalsciencereplication.wordpress.com/2018/07/14/replication-and-transparency-in-political-science-did-we-make-any-progress/ (accessed 15 June 2022).

Jensen, M.R. and J.W. Moses. 2021. The state of political science, 2020. *EPS* 20: 14-33.

King, G. 1995. Replication, Replication. *PS* 28: 444-452.

LaCour, M.J. and D.P. Green. 2014. When contact changes minds: An experiment on transmission of support for gay equality. *Science* 346: 1366-1369.

Laney, D. 2001. 3D data management: Controlling data volume, velocity, variety. Application Delivery Strategies Meta Group File 949. Available from: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed 14 May 2015).

Laitin, D.D. and R. Reich. 2017. Trust, transparency, and replication in political science. *PS* 50: 172-175.

Lax, J.R., J.H. Phillips and A.F. Stollwerk. 2016. Are survey respondents lying about their support for same-sex marriage? Lessons from a list experiment. *Public opinion quarterly* 80: 510-533.

Longley, P. 2012. Geodemographics and the practices of geographic information science. *International Journal of Geographical Information Science* 26: 2227–2237.

Lupia, A. and C. Elman. 2014. Openness in political science: data access and research transparency—introduction. *PS* 47: 19-42.

May, C. 2005. The academy's new electronic order? Open source journals and publishing political science. *EPS* 4: 14-24.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K.M. Esterling, A. Gerber and D. Laitin. 2014. Promoting transparency in social science research. *Science* 343: 30-31.

Milliband, D. 2016. How the pollsters got the US election wrong—just like Brexit. *The Telegraph*, 9 November.

Moher, D., A. Liberati, J. Tetzlaff, & D.G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine* 6: e1000097.

Powell, B. et al. 2016. Letter from distinguished political scientists urging nuanced journal interpretation of JETS policy guidelines. 13 January. Available from: https://politicalsciencenow.com/letter-from-distinguished-political-scientists-urging-nuanced-journal-interpretation-of-jets-policy-guidelines/ (accessed 9 December 2020).

Reinhart, C. and K.K. Rogoff. 2010. Growth in a Time of Debt. *American Economic Review: Papers & Proceedings* 100: 573–578.

Singal, J. 2015. The Case of the Amazing Gay-Marriage Data: How a Graduate Student Reluctantly Uncovered a Huge Scientific Fraud. Science of Us blog. 29 May. Available from: http://nymag.com/scienceofus/2015/05/how-a-grad-student-uncovered-a-huge-fraud.html (accessed on 14 March 2018).

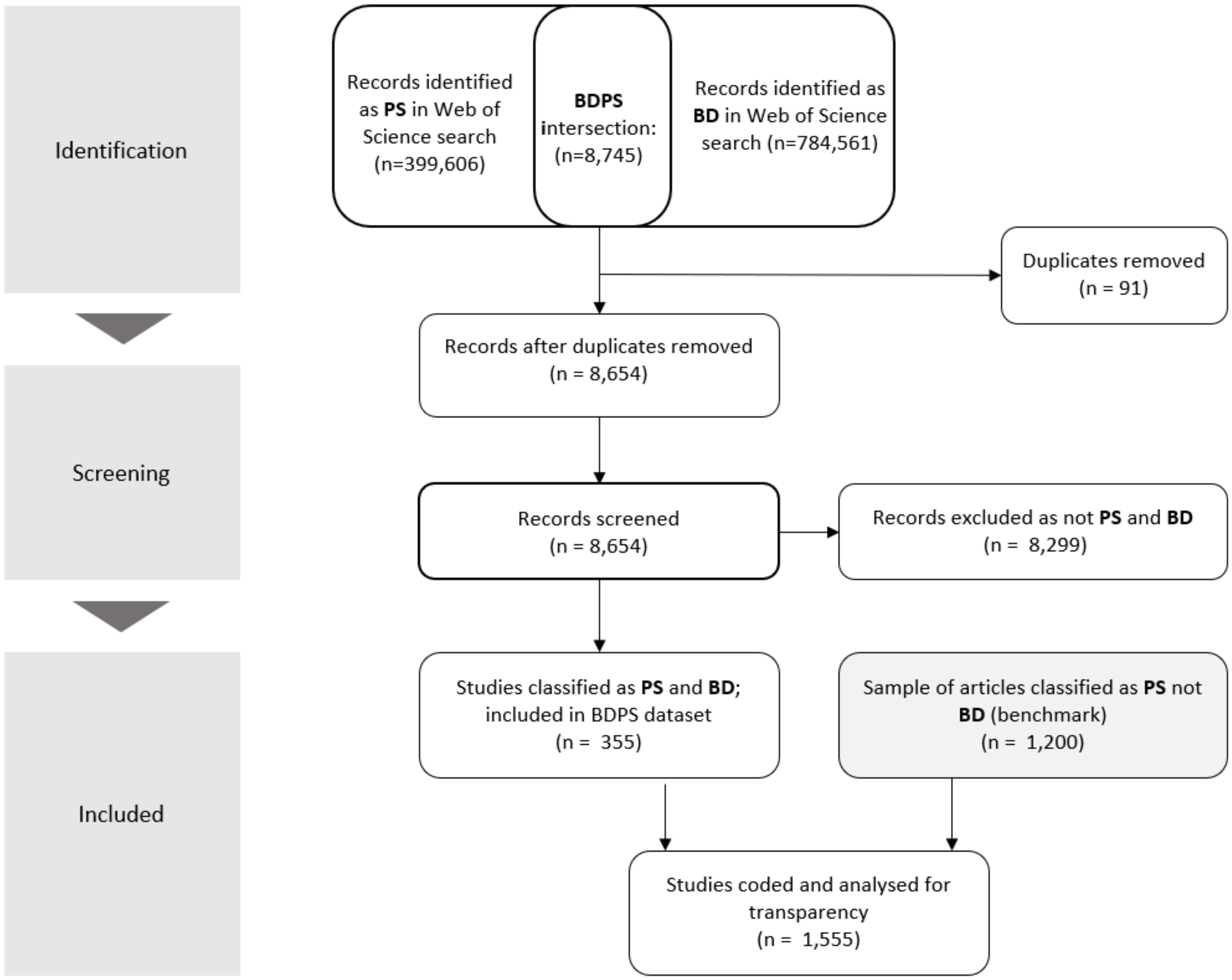
Tromble, R. 2021. Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age. *Social Media and Society* 1-8.

Tufekci, Z. 2020. Can we finally agree to ignore election forecasts? *The New York Times*, 1 November.

Wylie, C. 2019. *Mindf*ck: Cambridge Analytica and the plot to break America*. New York: Random House.

We would like to thank Erlend Gylver, Christoffer Kongsness, Eirik Lidsheim, Glenn Adrian Nordhus, and Alessia Reina for excellent research assistance, and Lisa Reutter, Melanie Magin, and Amélie Godefroidt for helpful comments on earlier versions of the manuscript.

Figure 1: Simplified workflow to construct the dataset



Note: Adapted from the PRISMA standard (Moher et al. 2009).

Figure 2: The terrain to be mapped

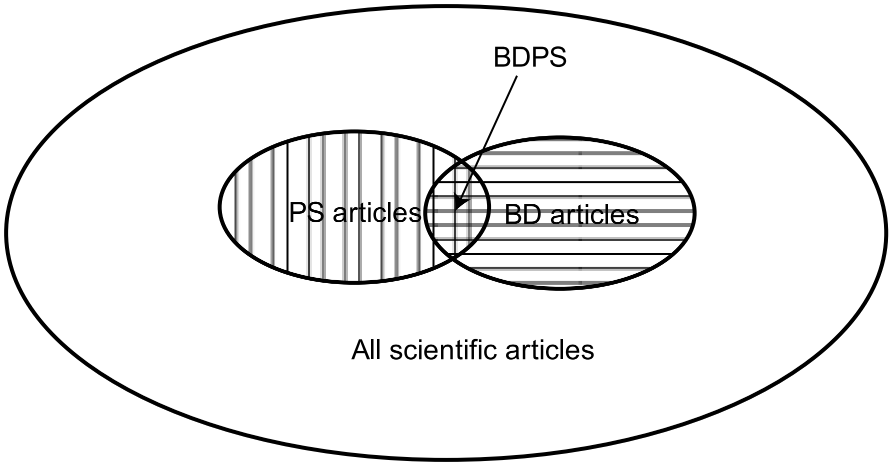


Figure 3: PS Word Cloud (Top 50)

Figure 4: BDPS publication trends over time, total, in PS journals, and in JETS journals, 2008-2019

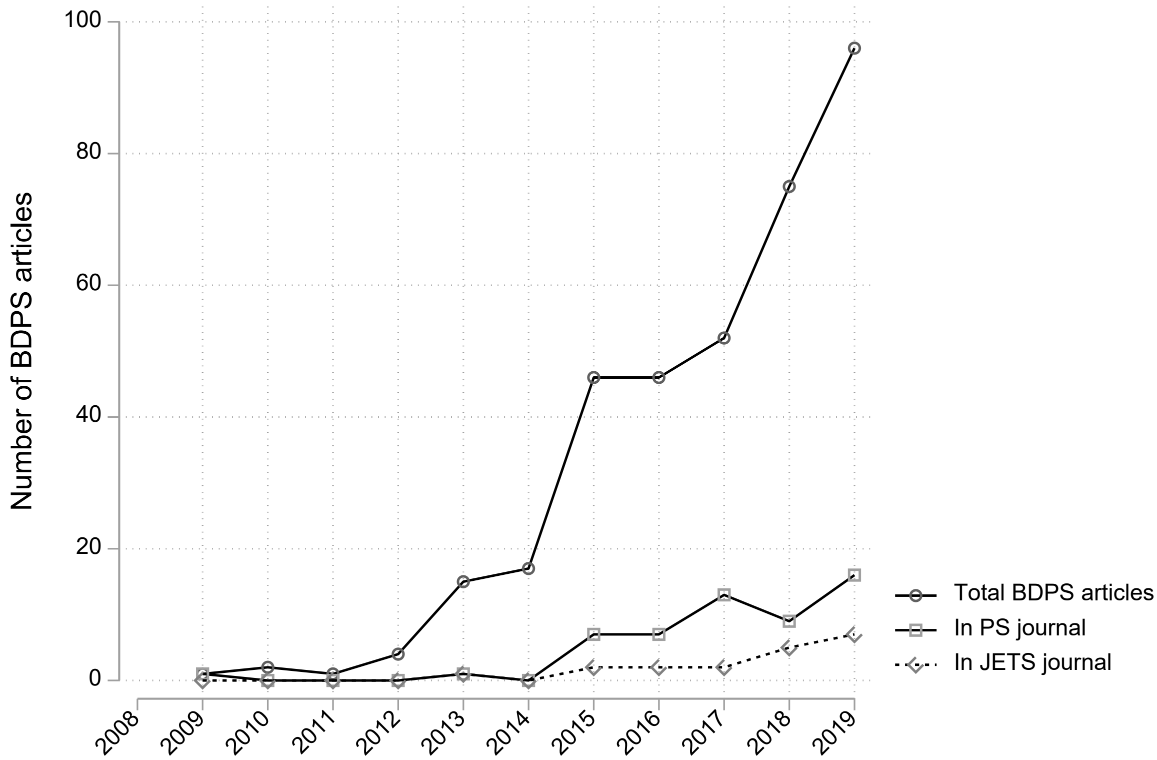
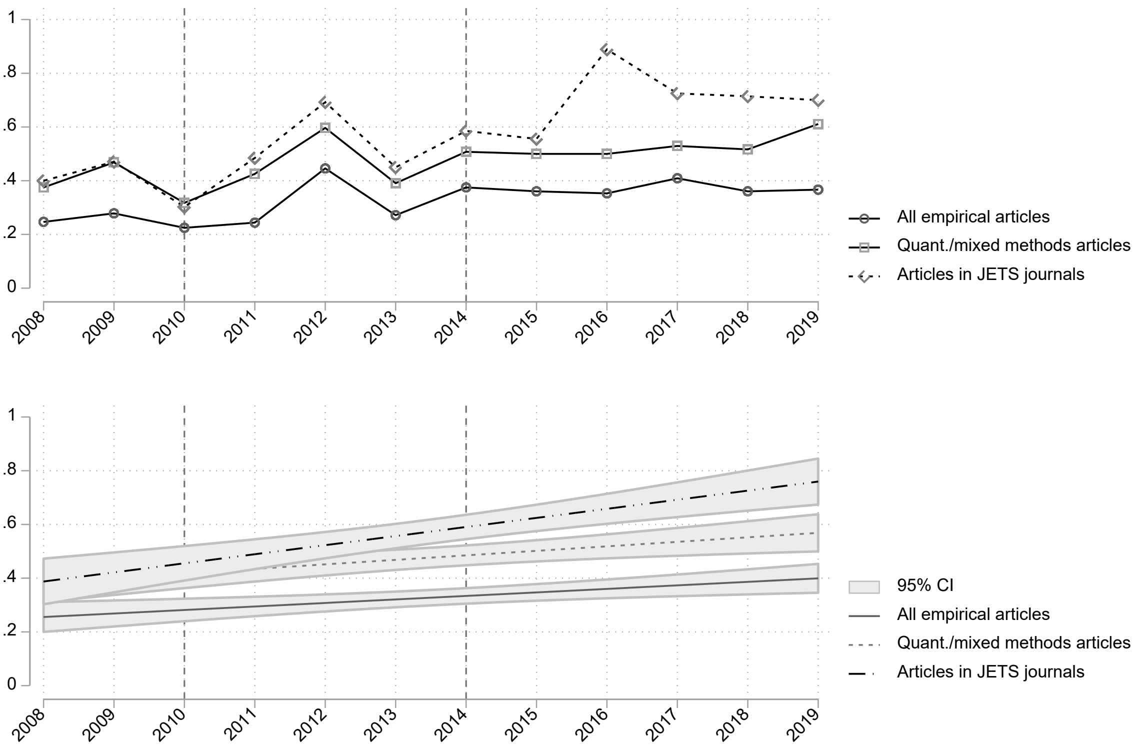
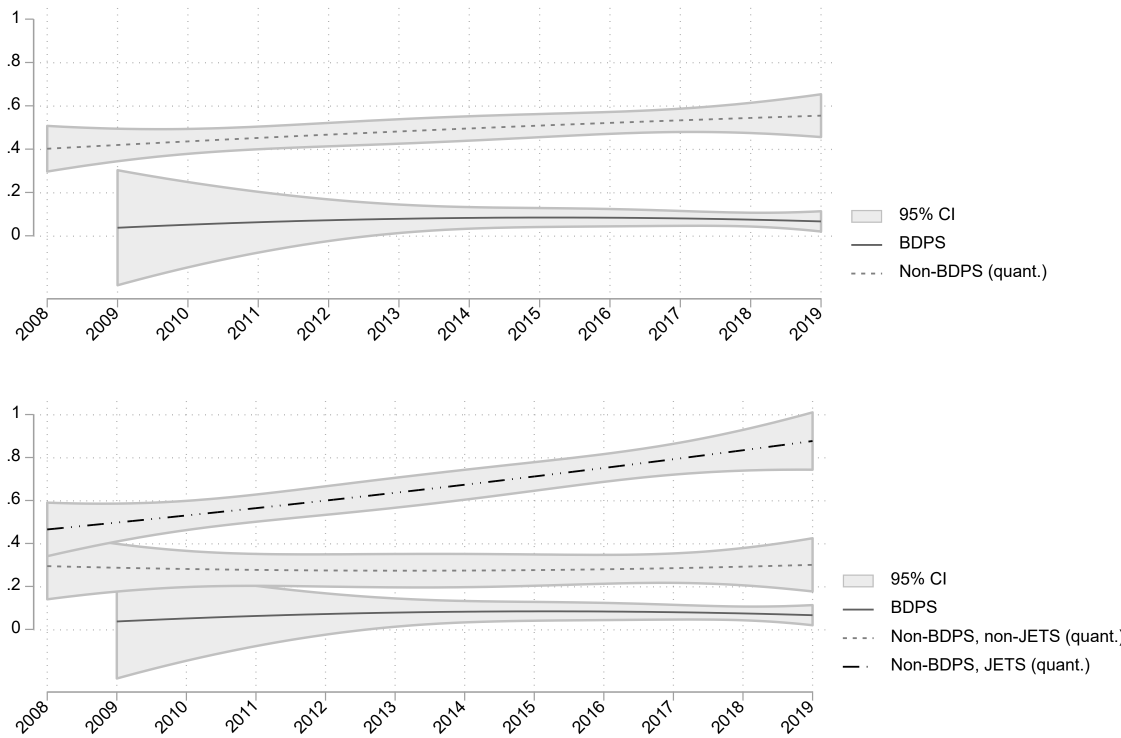


Figure 5: Share of PS articles that provide complete replication material, 2008-2019



Note: The upper panel shows the share of fully transparent articles per year in each category, while the lower panel shows fitted time trends with corresponding confidence intervals. Vertical, dashed lines indicate APSA's launch of the Data Access and Research Transparency DA-RT group (2010) and the JETS (2014).

Figure 6: Transparency trends among BDPS and quantitative non-BDPS articles



Note: The upper panel compares fitted time trends with corresponding confidence intervals BDPS with quantitative BDPS articles. In the lower panel, the group of quantitative non-BDPS articles is split into JETS and non-JETS journals.

Table 1: PS journals employed to sample abstracts

# of listings	Journal
5	*International Organization*
5	*World Politics*
5	*American Political Science Review*
5	*Journal of Conflict Resolution*
5	*International Security*
5	*International Studies Quarterly*
4	*American Journal of Political Science*
4	*European Journal of International Relations*
4	*Comparative Political Studies*
4	*Journal of Politics*
3	*Journal of Peace Research*
3	*Public Opinion Quarterly*
3	*Legislative Studies Quarterly*
2	*Journal of Political Economy*
2	*Political Geography*
2	*British Journal of Political Science*
2	*Comparative Politics*
1[1]	*Political Psychology*
0	*Political Communication*

[1] Included in SCImago, but not in Giles and Garand (2007, Table 2).

Table 2: Journals that published most BDPS articles, by Web of Science categorization of political science – non-political science

Non-PS journals	N	PS journals	N
Social Science Computer Review	21	*Journal of Information Technology & Politics*	12
Plos One	20	*American Political Science Review*	3
Information Communication & Society	14	*International Journal of Press-Politics*	3
New Media & Society	13	*Policy and Internet*	3
Social Media + Society	8	*Political Analysis*	3
*American Behavioral Scientist**	6	*Political Communication*	3

Note: * *EPJ Data Science, Government Information Quarterly,* and *Social Network Analysis and Mining* are also listed with 6 entries each.

Table 3: Logistic regression model of transparency

	(1)	(2)	(3)	(4)	(5)	(6)
BDPS	-1.775	-1.775	-1.094	-1.964	-2.557	-2.520
	(8.10)**	(4.37)**	(2.21)*	(4.85)**	(6.23)**	(11.68)**
Year	0.067	0.067	0.107	0.057	0.052	
	(3.58)**	(2.38)*	(3.03)**	(1.99)*	(1.59)	
JETS			2.459			
			(3.62)**			
Constant	-134.894	-134.894	-218.461	-114.919	-104.537	0.116
	(3.61)**	(2.39)*	(3.04)**	(2.00)*	(1.58)	(1.47)
Standard errors clustered by journal	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*No*
Non-empirical work excluded	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Pure quantitative work only	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Coarsened exact matching	*No*	*No*	*No*	*No*	*No*	*Yes*
N	1,555	1,555	1,555	1,374	988	973

Note: * p<0.05; ** p<0.01; logit coefficients.

Online Supplemental Material:

Big Data meets Open Political Science: An Empirical Assessment of Transparency Standards 2008-2019

A. Developing the political science search string

A.1. Comparison of journal rankings

SCImago	ISI (2007)	PS impact (2007)	WPS Impact	GG-Impact
AJPS	*International Organization*	*International Organization*	*International Organization*	*APSR*
International Organization	*Journal of Political Economy*	*APSR*	*APSR*	*AJPS*
World Politics	*World Politics*	*World Politics*	*World Politics*	*Journal of Politics*
QJPS	*APSR*	*AJPS*	*AJPS*	*World Politics*
APSR	*International Security*	*International Security*	*Journal of Conflict Resolution*	*International Organization*
Political Analysis	*EJIR*	*Journal of Conflict Resolution*	*International Studies Quarterly*	*BJPS*
Journal of Conflict Resolution	*Journal of Law and Economics*	*International Studies Quarterly*	*International Interactions*	*Comparative Politics*
International Security	*Public Opinion Quarterly*	*Journal of Politics*	*International Security*	*Comparative Political Studies*
Journal of Peace Research	*JLEO*	*EJIR*	*Journal of Peace Research*	*PS: Political Science and Politics*
REO	*AJIL*	*Journal of Peace Research*	*CMPS*	*Political Research Quarterly*
Western European Politics	*Political Geography*	*International Interactions*	*Journal of Politics*	*International Studies Quarterly*
EJIR	*Law and Society Review*	*Comparative Political Studies*	*Comparative Political Studies*	*Political Science Quarterly*
International Studies Quarterly	*Politics and Society*	*Post Soviet Affairs*	*EJIR*	*Public Opinion Quarterly*
CMPS	*Journal of Conflict Resolution*	*Legislative Studies Quarterly*	*Legislative Studies Quarterly*	*Journal of Conflict Resolution*
Post-Soviet Affairs	*International Studies Quarterly*	*Public Opinion Quarterly*	*BJPS*	*International Security*
Security Dialogue	*Urban Studies*	*Public Administration Review*	*Journal of Theoretical Politics*	*Legislative Studies Quarterly*
Philosophy and Public Affairs	*World Development*	*CMPS*	*Comparative Politics*	*Political Theory*
ITJIPLP	*Comparative Political Studies*	*Journal of Democracy*	*Political Research Quarterly*	*Public Administration Reviw*
China Quarterly	*Theory and Society*	*Political Geography*	*Security Studies*	*Journal of Political Economy*
Political Psychology	*Journal of Politics*	*Electoral Studies*	*SCPID*	*Polity*

Note: All but the SCImago index are from Giles and Garrand (2007, p. 744). The Scimago ranking is available at https://www.scimagojr.com/journalrank.php?category=3320&area=3300&type=j&min=0&min_type=cd, accessed 22 Oct 2018. *AJIL: American Journal of International Law; AJPS: American Journal of Political Science; APSR: American Political Science Review; BJPS: British Journal of Political Science; CMPS: Conflict Management and Peace Science; EJIR: European Journal of International Relations; ITJIPLP: International Theory: A Journal of International Politics, Law and Philosophy; JLEO: Journal of Law, Economics and Organization; QJPS: Quarterly Journal of Political Science; REO: Review of International Organizations, SCPID: Studies in Comparative and International Development.*

A.2. The political science search string

(TS =((election$ OR elector*) AND democra* AND institution$) OR TS =((election$ OR elector*) AND democra* AND (civil OR civic)) OR TS =((election$ OR elector*) AND democra* AND power) OR TS =((election$ OR elector*) AND democra* AND citizen$) OR TS =((election$ OR elector*) AND democra* AND gender) OR TS =((election$ OR elector*) AND democra* AND nation*) OR TS =((election$ OR elector*) AND democra* AND interest$) OR TS =((election$ OR elector*) AND democra* AND labo$r) OR TS =((election$ OR elector*) AND democra* AND opposition$) OR TS =((election$ OR elector*) AND institution$ AND (civil OR civic)) OR TS =((election$ OR elector*) AND institution$ AND power) OR TS =((election$ OR elector*) AND institution$ AND citizen$) OR TS =((election$ OR elector*) AND institution$ AND gender) OR TS =((election$ OR elector*) AND institution$ AND nation*) OR TS =((election$ OR elector*) AND institution$ AND interest$) OR TS =((election$ OR elector*) AND institution$ AND labo$r) OR TS =((election$ OR elector*) AND institution$ AND opposition$) OR TS =((election$ OR elector*) AND (civil OR civic) AND power) OR TS =((election$ OR elector*) AND (civil OR civic) AND citizen$) OR TS =((election$ OR elector*) AND (civil OR civic) AND gender) OR TS =((election$ OR elector*) AND (civil OR civic) AND nation*) OR TS =((election$ OR elector*) AND (civil OR civic) AND interest$) OR TS =((election$ OR elector*) AND (civil OR civic) AND labo$r) OR TS =((election$ OR elector*) AND (civil OR civic) AND opposition$) OR TS =((election$ OR elector*) AND power AND citizen$) OR TS =((election$ OR elector*) AND power AND gender) OR TS =((election$ OR elector*) AND power AND nation*) OR TS =((election$ OR elector*) AND power AND interest$) OR TS =((election$ OR elector*) AND power AND labo$r) OR TS =((election$ OR elector*) AND power AND opposition$) OR TS =((election$ OR elector*) AND citizen$ AND gender) OR TS =((election$ OR elector*) AND citizen$ AND nation*) OR TS =((election$ OR elector*) AND citizen$ AND interest$) OR TS =((election$ OR elector*) AND citizen$ AND labo$r) OR TS =((election$ OR elector*) AND citizen$ AND opposition$) OR TS =((election$ OR elector*) AND gender AND nation*) OR TS =((election$ OR elector*) AND gender AND interest$) OR TS =((election$ OR elector*) AND gender AND labo$r) OR TS =((election$ OR elector*) AND gender AND opposition$) OR TS =((election$ OR elector*) AND nation* AND interest$) OR TS =((election$ OR elector*) AND nation* AND labo$r) OR TS =((election$ OR elector*) AND nation* AND opposition$) OR TS =((election$ OR elector*) AND interest$ AND labo$r) OR TS =((election$ OR elector*) AND interest$ AND opposition$) OR TS =((election$ OR elector*) AND labo$r AND opposition$) OR TS =(countr* AND war OR wars AND govern*) OR TS =(countr* AND war OR wars AND (party OR parties)) OR TS =(countr* AND war OR wars AND democra*) OR TS =(countr* AND war OR wars AND institution$) OR TS =(countr* AND war OR wars AND (civil OR civic)) OR TS =(countr* AND war OR wars AND power) OR TS =(countr* AND war OR wars AND citizen$) OR TS =(countr* AND war OR wars AND gender) OR TS =(countr* AND war OR wars AND nation*) OR TS =(countr* AND war OR wars AND interest$) OR TS =(countr* AND war OR wars AND labo$r) OR TS =(countr* AND war OR wars AND opposition$) OR TS =(countr* AND govern* AND (party OR parties)) OR TS =(countr* AND govern* AND democra*) OR TS =(countr* AND govern* AND institution$) OR TS =(countr* AND govern* AND (civil OR civic)) OR TS =(countr* AND govern* AND power) OR TS =(countr* AND govern* AND citizen$) OR TS =(countr* AND govern* AND gender) OR TS =(countr* AND govern* AND nation*) OR TS =(countr* AND govern* AND interest$) OR TS =(countr* AND govern* AND labo$r) OR TS =(countr* AND govern* AND opposition$) OR TS =(countr* AND (party OR parties) AND democra*) OR TS =(countr* AND (party OR parties) AND institution$) OR TS =(countr* AND (party OR parties) AND (civil OR civic)) OR TS =(countr* AND (party OR parties) AND power) OR TS =(countr* AND (party OR parties) AND citizen$) OR TS =(countr* AND (party OR parties) AND gender) OR TS =(countr* AND (party OR parties) AND nation*) OR TS =(countr* AND (party OR parties) AND interest$) OR TS =(countr* AND (party OR parties) AND labo$r) OR TS =(countr* AND (party OR parties) AND opposition$) OR TS =(countr* AND democra* AND institution$) OR TS =(countr* AND democra* AND (civil OR civic)) OR TS =(countr* AND democra* AND power) OR TS =(countr* AND democra* AND citizen$) OR TS =(countr* AND democra* AND gender) OR TS =(countr* AND democra* AND nation*) OR TS =(countr* AND democra* AND interest$) OR TS =(countr* AND democra* AND labo$r) OR TS =(countr* AND democra* AND opposition$) OR TS =(countr* AND institution$ AND (civil OR civic)) OR TS =(countr* AND institution$ AND power) OR TS =(countr* AND institution$ AND citizen$) OR TS =(countr* AND institution$ AND gender) OR TS =(countr* AND institution$ AND nation*) OR TS =(countr* AND institution$ AND interest$) OR TS =(countr* AND institution$ AND labo$r) OR TS =(countr* AND institution$ AND opposition$) OR TS =(countr* AND (civil OR civic) AND power) OR TS =(countr* AND (civil OR civic) AND citizen$) OR TS =(countr* AND (civil OR civic) AND gender) OR TS =(countr* AND (civil OR civic) AND nation*) OR TS =(countr* AND (civil OR civic) AND interest$) OR TS =(countr* AND (civil OR civic) AND labo$r) OR TS =(countr* AND (civil OR civic) AND opposition$) OR TS =(countr* AND power AND citizen$) OR TS =(countr* AND power AND gender) OR TS =(countr* AND power AND nation*) OR TS =(countr* AND power AND interest$) OR TS =(countr* AND power AND labo$r) OR TS =(countr* AND power AND opposition$) OR TS =(countr* AND citizen$ AND gender) OR TS =(countr* AND citizen$ AND nation*) OR TS =(countr* AND citizen$ AND interest$) OR TS =(countr* AND citizen$ AND labo$r) OR TS =(countr* AND citizen$ AND opposition$) OR TS =(countr* AND gender AND nation*) OR TS =(countr* AND gender AND interest$) OR TS =(countr* AND gender AND labo$r) OR TS =(countr* AND gender AND opposition$) OR TS =(countr* AND nation* AND interest$) OR TS =(countr* AND nation* AND labo$r) OR TS =(countr* AND nation* AND opposition$) OR TS =(countr* AND interest$ AND labo$r) OR TS =(countr* AND interest$ AND opposition$) OR TS =(countr* AND labo$r AND opposition$) OR TS =(war OR wars AND govern* AND (party OR parties)) OR TS =(war OR wars AND govern* AND democra*) OR TS =(war OR wars AND govern* AND institution$) OR TS =(war OR wars AND govern* AND (civil OR civic)) OR TS =(war OR wars AND govern* AND power) OR TS =(war OR wars AND govern* AND citizen$) OR TS =(war OR wars AND govern* AND gender) OR TS =(war OR wars AND govern* AND nation*) OR TS =(war OR wars AND govern* AND interest$) OR TS =(war OR wars AND govern* AND labo$r) OR TS =(war

OR wars AND govern* AND opposition$) OR TS =(war OR wars AND (party OR parties) AND democra*) OR TS =(war OR wars AND (party OR parties) AND institution$) OR TS =(war OR wars AND (party OR parties) AND (civil OR civic)) OR TS =(war OR wars AND (party OR parties) AND power) OR TS =(war OR wars AND (party OR parties) AND citizen$) OR TS =(war OR wars AND (party OR parties) AND gender) OR TS =(war OR wars AND (party OR parties) AND nation*) OR TS =(war OR wars AND (party OR parties) AND interest$) OR TS =(war OR wars AND (party OR parties) AND labo$r) OR TS =(war OR wars AND (party OR parties) AND opposition$) OR TS =(war OR wars AND democra* AND institution$) OR TS =(war OR wars AND democra* AND (civil OR civic)) OR TS =(war OR wars AND democra* AND power) OR TS =(war OR wars AND democra* AND citizen$) OR TS =(war OR wars AND democra* AND gender) OR TS =(war OR wars AND democra* AND nation*) OR TS =(war OR wars AND democra* AND interest$) OR TS =(war OR wars AND democra* AND labo$r) OR TS =(war OR wars AND democra* AND opposition$) OR TS =(war OR wars AND institution$ AND (civil OR civic)) OR TS =(war OR wars AND institution$ AND power) OR TS =(war OR wars AND institution$ AND citizen$) OR TS =(war OR wars AND institution$ AND gender) OR TS =(war OR wars AND institution$ AND nation*) OR TS =(war OR wars AND institution$ AND interest$) OR TS =(war OR wars AND institution$ AND labo$r) OR TS =(war OR wars AND institution$ AND opposition$) OR TS =(war OR wars AND (civil OR civic) AND power) OR TS =(war OR wars AND (civil OR civic) AND citizen$) OR TS =(war OR wars AND (civil OR civic) AND gender) OR TS =(war OR wars AND (civil OR civic) AND nation*) OR TS =(war OR wars AND (civil OR civic) AND interest$) OR TS =(war OR wars AND (civil OR civic) AND labo$r) OR TS =(war OR wars AND (civil OR civic) AND opposition$) OR TS =(war OR wars AND power AND citizen$) OR TS =(war OR wars AND power AND gender) OR TS =(war OR wars AND power AND nation*) OR TS =(war OR wars AND power AND interest$) OR TS =(war OR wars AND power AND labo$r) OR TS =(war OR wars AND power AND opposition$) OR TS =(war OR wars AND citizen$ AND gender) OR TS =(war OR wars AND citizen$ AND nation*) OR TS =(war OR wars AND citizen$ AND interest$) OR TS =(war OR wars AND citizen$ AND labo$r) OR TS =(war OR wars AND citizen$ AND opposition$) OR TS =(war OR wars AND gender AND nation*) OR TS =(war OR wars AND gender AND interest$) OR TS =(war OR wars AND gender AND labo$r) OR TS =(war OR wars AND gender AND opposition$) OR TS =(war OR wars AND nation* AND interest$) OR TS =(war OR wars AND nation* AND labo$r) OR TS =(war OR wars AND nation* AND opposition$) OR TS =(war OR wars AND interest$ AND labo$r) OR TS =(war OR wars AND interest$ AND opposition$) OR TS =(war OR wars AND labo$r AND opposition$) OR TS =(govern* AND (party OR parties) AND democra*) OR TS =(govern* AND (party OR parties) AND institution$) OR TS =(govern* AND (party OR parties) AND (civil OR civic)) OR TS =(govern* AND (party OR parties) AND power) OR TS =(govern* AND (party OR parties) AND citizen$) OR TS =(govern* AND (party OR parties) AND gender) OR TS =(govern* AND (party OR parties) AND nation*) OR TS =(govern* AND (party OR parties) AND interest$) OR TS =(govern* AND (party OR parties) AND labo$r) OR TS =(govern* AND (party OR parties) AND opposition$) OR TS =(govern* AND democra* AND institution$) OR TS =(govern* AND democra* AND (civil OR civic)) OR TS =(govern* AND democra* AND power) OR TS =(govern* AND democra* AND citizen$) OR TS =(govern* AND democra* AND gender) OR TS =(govern* AND democra* AND nation*) OR TS =(govern* AND democra* AND interest$) OR TS =(govern* AND democra* AND labo$r) OR TS =(govern* AND democra* AND opposition$) OR TS =(govern* AND institution$ AND (civil OR civic)) OR TS =(govern* AND institution$ AND power) OR TS =(govern* AND institution$ AND citizen$) OR TS =(govern* AND institution$ AND gender) OR TS =(govern* AND institution$ AND nation*) OR TS =(govern* AND institution$ AND interest$) OR TS =(govern* AND institution$ AND labo$r) OR TS =(govern* AND institution$ AND opposition$) OR TS =(govern* AND (civil OR civic) AND power) OR TS =(govern* AND (civil OR civic) AND citizen$) OR TS =(govern* AND (civil OR civic) AND gender) OR TS =(govern* AND (civil OR civic) AND nation*) OR TS =(govern* AND (civil OR civic) AND interest$) OR TS =(govern* AND (civil OR civic) AND labo$r) OR TS =(govern* AND (civil OR civic) AND opposition$) OR TS =(govern* AND power AND citizen$) OR TS =(govern* AND power AND gender) OR TS =(govern* AND power AND nation*) OR TS =(govern* AND power AND interest$) OR TS =(govern* AND power AND labo$r) OR TS =(govern* AND power AND opposition$) OR TS =(govern* AND citizen$ AND gender) OR TS =(govern* AND citizen$ AND nation*) OR TS =(govern* AND citizen$ AND interest$) OR TS =(govern* AND citizen$ AND labo$r) OR TS =(govern* AND citizen$ AND opposition$) OR TS =(govern* AND gender AND nation*) OR TS =(govern* AND gender AND interest$) OR TS =(govern* AND gender AND labo$r) OR TS =(govern* AND gender AND opposition$) OR TS =(govern* AND nation* AND interest$) OR TS =(govern* AND nation* AND labo$r) OR TS =(govern* AND nation* AND opposition$) OR TS =(govern* AND interest$ AND labo$r) OR TS =(govern* AND interest$ AND opposition$) OR TS =(govern* AND labo$r AND opposition$) OR TS =((party OR parties) AND democra* AND institution$) OR TS =((party OR parties) AND democra* AND (civil OR civic)) OR TS =((party OR parties) AND democra* AND power) OR TS =((party OR parties) AND democra* AND citizen$) OR TS =((party OR parties) AND democra* AND gender) OR TS =((party OR parties) AND democra* AND nation*) OR TS =((party OR parties) AND democra* AND interest$) OR TS =((party OR parties) AND democra* AND labo$r) OR TS =((party OR parties) AND democra* AND opposition$) OR TS =((party OR parties) AND institution$ AND (civil OR civic)) OR TS =((party OR parties) AND institution$ AND power) OR TS =((party OR parties) AND institution$ AND citizen$) OR TS =((party OR parties) AND institution$ AND gender) OR TS =((party OR parties) AND institution$ AND nation*) OR TS =((party OR parties) AND institution$ AND interest$) OR TS =((party OR parties) AND institution$ AND labo$r) OR TS =((party OR parties) AND institution$ AND opposition$) OR TS =((party OR parties) AND (civil OR civic) AND power) OR TS =((party OR parties) AND (civil OR civic) AND citizen$) OR TS =((party OR parties) AND (civil OR civic) AND gender) OR TS =((party OR parties) AND (civil OR civic) AND nation*) OR TS =((party OR parties) AND (civil OR civic) AND interest$) OR TS =((party OR parties) AND (civil OR civic) AND labo$r) OR TS =((party OR parties) AND (civil OR civic) AND opposition$) OR TS =((party OR parties) AND power AND citizen$) OR TS =((party OR parties) AND power AND gender) OR TS =((party OR parties) AND power AND nation*) OR TS =((party OR parties) AND power AND interest$) OR TS =((party OR parties) AND power AND labo$r) OR TS =((party OR parties) AND power AND opposition$) OR TS =((party OR parties) AND citizen$ AND gender) OR TS =((party OR parties) AND citizen$ AND nation*) OR TS =((party OR parties) AND citizen$ AND interest$) OR TS =((party OR parties)

AND citizen$ AND labo$r) OR TS =((party OR parties) AND citizen$ AND opposition$) OR TS =((party OR parties) AND gender AND nation*) OR TS =((party OR parties) AND gender AND interest$) OR TS =((party OR parties) AND gender AND labo$r) OR TS =((party OR parties) AND gender AND opposition$) OR TS =((party OR parties) AND nation* AND interest$) OR TS =((party OR parties) AND nation* AND labo$r) OR TS =((party OR parties) AND nation* AND opposition$) OR TS =((party OR parties) AND interest$ AND labo$r) OR TS =((party OR parties) AND interest$ AND opposition$) OR TS =((party OR parties) AND labo$r AND opposition$) OR TS =(democra* AND institution$ AND (civil OR civic)) OR TS =(democra* AND institution$ AND power) OR TS =(democra* AND institution$ AND citizen$) OR TS =(democra* AND institution$ AND gender) OR TS =(democra* AND institution$ AND nation*) OR TS =(democra* AND institution$ AND interest$) OR TS =(democra* AND institution$ AND labo$r) OR TS =(democra* AND institution$ AND opposition$) OR TS =(democra* AND (civil OR civic) AND power) OR TS =(democra* AND (civil OR civic) AND citizen$) OR TS =(democra* AND (civil OR civic) AND gender) OR TS =(democra* AND (civil OR civic) AND nation*) OR TS =(democra* AND (civil OR civic) AND interest$) OR TS =(democra* AND (civil OR civic) AND labo$r) OR TS =(democra* AND (civil OR civic) AND opposition$) OR TS =(democra* AND power AND citizen$) OR TS =(democra* AND power AND gender) OR TS =(democra* AND power AND nation*) OR TS =(democra* AND power AND interest$) OR TS =(democra* AND power AND labo$r) OR TS =(democra* AND power AND opposition$) OR TS =(democra* AND citizen$ AND gender) OR TS =(democra* AND citizen$ AND nation*) OR TS =(democra* AND citizen$ AND interest$) OR TS =(democra* AND citizen$ AND labo$r) OR TS =(democra* AND citizen$ AND opposition$) OR TS =(democra* AND gender AND nation*) OR TS =(democra* AND gender AND interest$) OR TS =(democra* AND gender AND labo$r) OR TS =(democra* AND gender AND opposition$) OR TS =(democra* AND nation* AND interest$) OR TS =(democra* AND nation* AND labo$r) OR TS =(democra* AND nation* AND opposition$) OR TS =(democra* AND interest$ AND labo$r) OR TS =(democra* AND interest$ AND opposition$) OR TS =(democra* AND labo$r AND opposition$) OR TS =(institution$ AND (civil OR civic) AND power) OR TS =(institution$ AND (civil OR civic) AND citizen$) OR TS =(institution$ AND (civil OR civic) AND gender) OR TS =(institution$ AND (civil OR civic) AND nation*) OR TS =(institution$ AND (civil OR civic) AND interest$) OR TS =(institution$ AND (civil OR civic) AND labo$r) OR TS =(institution$ AND (civil OR civic) AND opposition$) OR TS =(institution$ AND power AND citizen$) OR TS =(institution$ AND power AND gender) OR TS =(institution$ AND power AND nation*) OR TS =(institution$ AND power AND interest$) OR TS =(institution$ AND power AND labo$r) OR TS =(institution$ AND power AND opposition$) OR TS =(institution$ AND citizen$ AND gender) OR TS =(institution$ AND citizen$ AND nation*) OR TS =(institution$ AND citizen$ AND interest$) OR TS =(institution$ AND citizen$ AND labo$r) OR TS =(institution$ AND citizen$ AND opposition$) OR TS =(institution$ AND gender AND nation*) OR TS =(institution$ AND gender AND interest$) OR TS =(institution$ AND gender AND labo$r) OR TS =(institution$ AND gender AND opposition$) OR TS =(institution$ AND nation* AND interest$) OR TS =(institution$ AND nation* AND labo$r) OR TS =(institution$ AND nation* AND opposition$) OR TS =(institution$ AND interest$ AND labo$r) OR TS =(institution$ AND interest$ AND opposition$) OR TS =(institution$ AND labo$r AND opposition$) OR TS =((civil OR civic) AND power AND citizen$) OR TS =((civil OR civic) AND power AND gender) OR TS =((civil OR civic) AND power AND nation*) OR TS =((civil OR civic) AND power AND interest$) OR TS =((civil OR civic) AND power AND labo$r) OR TS =((civil OR civic) AND power AND opposition$) OR TS =((civil OR civic) AND citizen$ AND gender) OR TS =((civil OR civic) AND citizen$ AND nation*) OR TS =((civil OR civic) AND citizen$ AND interest$) OR TS =((civil OR civic) AND citizen$ AND labo$r) OR TS =((civil OR civic) AND citizen$ AND opposition$) OR TS =((civil OR civic) AND gender AND nation*) OR TS =((civil OR civic) AND gender AND interest$) OR TS =((civil OR civic) AND gender AND labo$r) OR TS =((civil OR civic) AND gender AND opposition$) OR TS =((civil OR civic) AND nation* AND interest$) OR TS =((civil OR civic) AND nation* AND labo$r) OR TS =((civil OR civic) AND nation* AND opposition$) OR TS =((civil OR civic) AND interest$ AND labo$r) OR TS =((civil OR civic) AND interest$ AND opposition$) OR TS =((civil OR civic) AND labo$r AND opposition$) OR TS =(power AND citizen$ AND gender) OR TS =(power AND citizen$ AND nation*) OR TS =(power AND citizen$ AND interest$) OR TS =(power AND citizen$ AND labo$r) OR TS =(power AND citizen$ AND opposition$) OR TS =(power AND gender AND nation*) OR TS =(power AND gender AND interest$) OR TS =(power AND gender AND labo$r) OR TS =(power AND gender AND opposition$) OR TS =(power AND nation* AND interest$) OR TS =(power AND nation* AND labo$r) OR TS =(power AND nation* AND opposition$) OR TS =(power AND interest$ AND labo$r) OR TS =(power AND interest$ AND opposition$) OR TS =(power AND labo$r AND opposition$) OR TS =(citizen$ AND gender AND nation*) OR TS =(citizen$ AND gender AND interest$) OR TS =(citizen$ AND gender AND labo$r) OR TS =(citizen$ AND gender AND opposition$) OR TS =(citizen$ AND nation* AND interest$) OR TS =(citizen$ AND nation* AND labo$r) OR TS =(citizen$ AND nation* AND opposition$) OR TS =(citizen$ AND interest$ AND labo$r) OR TS =(citizen$ AND interest$ AND opposition$) OR TS =(citizen$ AND labo$r AND opposition$) OR TS =(gender AND nation* AND interest$) OR TS =(gender AND nation* AND labo$r) OR TS =(gender AND nation* AND opposition$) OR TS =(gender AND interest$ AND labo$r) OR TS =(gender AND interest$ AND opposition$) OR TS =(gender AND labo$r AND opposition$) OR TS =(nation* AND interest$ AND labo$r) OR TS =(nation* AND interest$ AND opposition$) OR TS =(nation* AND labo$r AND opposition$) OR TS =(interest$ AND labo$r AND opposition$)) OR (TS =((politic* OR policy OR policies) AND (state OR states) AND conflict$) OR TS =((politic* OR policy OR policies) AND (state OR states) AND (election$ OR elector*)) OR TS =((politic* OR policy OR policies) AND (state OR states) AND countr*) OR TS =((politic* OR policy OR policies) AND (state OR states) AND war OR wars) OR TS =((politic* OR policy OR policies) AND (state OR states) AND govern*) OR TS =((politic* OR policy OR policies) AND (state OR states) AND (party OR parties)) OR TS =((politic* OR policy OR policies) AND (state OR states) AND democra*) OR TS =((politic* OR policy OR policies) AND (state OR states) AND institution$) OR TS =((politic* OR policy OR policies) AND (state OR states) AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND (state OR states) AND power) OR TS =((politic* OR policy OR policies) AND (state OR states) AND citizen$) OR TS =((politic* OR policy OR policies) AND (state OR states) AND gender) OR TS =((politic* OR policy OR policies) AND (state OR states) AND nation*) OR TS =((politic* OR policy OR policies) AND (state OR states)

AND interest$) OR TS =((politic* OR policy OR policies) AND (state OR states) AND labo$r) OR TS =((politic* OR policy OR policies) AND (state OR states) AND opposition$) OR TS =((politic* OR policy OR policies) AND conflict$ AND (election$ OR elector*)) OR TS =((politic* OR policy OR policies) AND conflict$ AND countr*) OR TS =((politic* OR policy OR policies) AND conflict$ AND war OR wars) OR TS =((politic* OR policy OR policies) AND conflict$ AND govern*) OR TS =((politic* OR policy OR policies) AND conflict$ AND (party OR parties)) OR TS =((politic* OR policy OR policies) AND conflict$ AND democra*) OR TS =((politic* OR policy OR policies) AND conflict$ AND institution$) OR TS =((politic* OR policy OR policies) AND conflict$ AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND conflict$ AND power) OR TS =((politic* OR policy OR policies) AND conflict$ AND citizen$) OR TS =((politic* OR policy OR policies) AND conflict$ AND gender) OR TS =((politic* OR policy OR policies) AND conflict$ AND nation*) OR TS =((politic* OR policy OR policies) AND conflict$ AND interest$) OR TS =((politic* OR policy OR policies) AND conflict$ AND labo$r) OR TS =((politic* OR policy OR policies) AND conflict$ AND opposition$) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND countr*) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND war OR wars) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND govern*) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND (party OR parties)) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND democra*) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND institution$) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND power) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND citizen$) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND gender) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND nation*) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND interest$) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND labo$r) OR TS =((politic* OR policy OR policies) AND (election$ OR elector*) AND opposition$) OR TS =((politic* OR policy OR policies) AND countr* AND war OR wars) OR TS =((politic* OR policy OR policies) AND countr* AND govern*) OR TS =((politic* OR policy OR policies) AND countr* AND (party OR parties)) OR TS =((politic* OR policy OR policies) AND countr* AND democra*) OR TS =((politic* OR policy OR policies) AND countr* AND institution$) OR TS =((politic* OR policy OR policies) AND countr* AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND countr* AND power) OR TS =((politic* OR policy OR policies) AND countr* AND citizen$) OR TS =((politic* OR policy OR policies) AND countr* AND gender) OR TS =((politic* OR policy OR policies) AND countr* AND nation*) OR TS =((politic* OR policy OR policies) AND countr* AND interest$) OR TS =((politic* OR policy OR policies) AND countr* AND labo$r) OR TS =((politic* OR policy OR policies) AND countr* AND opposition$) OR TS =((politic* OR policy OR policies) AND war OR wars AND govern*) OR TS =((politic* OR policy OR policies) AND war OR wars AND (party OR parties)) OR TS =((politic* OR policy OR policies) AND war OR wars AND democra*) OR TS =((politic* OR policy OR policies) AND war OR wars AND institution$) OR TS =((politic* OR policy OR policies) AND war OR wars AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND war OR wars AND power) OR TS =((politic* OR policy OR policies) AND war OR wars AND citizen$) OR TS =((politic* OR policy OR policies) AND war OR wars AND gender) OR TS =((politic* OR policy OR policies) AND war OR wars AND nation*) OR TS =((politic* OR policy OR policies) AND war OR wars AND interest$) OR TS =((politic* OR policy OR policies) AND war OR wars AND labo$r) OR TS =((politic* OR policy OR policies) AND war OR wars AND opposition$) OR TS =((politic* OR policy OR policies) AND govern* AND (party OR parties)) OR TS =((politic* OR policy OR policies) AND govern* AND democra*) OR TS =((politic* OR policy OR policies) AND govern* AND institution$) OR TS =((politic* OR policy OR policies) AND govern* AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND govern* AND power) OR TS =((politic* OR policy OR policies) AND govern* AND citizen$) OR TS =((politic* OR policy OR policies) AND govern* AND gender) OR TS =((politic* OR policy OR policies) AND govern* AND nation*) OR TS =((politic* OR policy OR policies) AND govern* AND interest$) OR TS =((politic* OR policy OR policies) AND govern* AND labo$r) OR TS =((politic* OR policy OR policies) AND govern* AND opposition$) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND democra*) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND institution$) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND power) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND citizen$) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND gender) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND nation*) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND interest$) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND labo$r) OR TS =((politic* OR policy OR policies) AND (party OR parties) AND opposition$) OR TS =((politic* OR policy OR policies) AND democra* AND institution$) OR TS =((politic* OR policy OR policies) AND democra* AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND democra* AND power) OR TS =((politic* OR policy OR policies) AND democra* AND citizen$) OR TS =((politic* OR policy OR policies) AND democra* AND gender) OR TS =((politic* OR policy OR policies) AND democra* AND nation*) OR TS =((politic* OR policy OR policies) AND democra* AND interest$) OR TS =((politic* OR policy OR policies) AND democra* AND labo$r) OR TS =((politic* OR policy OR policies) AND democra* AND opposition$) OR TS =((politic* OR policy OR policies) AND institution$ AND (civil OR civic)) OR TS =((politic* OR policy OR policies) AND institution$ AND power) OR TS =((politic* OR policy OR policies) AND institution$ AND citizen$) OR TS =((politic* OR policy OR policies) AND institution$ AND gender) OR TS =((politic* OR policy OR policies) AND institution$ AND nation*) OR TS =((politic* OR policy OR policies) AND institution$ AND interest$) OR TS =((politic* OR policy OR policies) AND institution$ AND labo$r) OR TS =((politic* OR policy OR policies) AND institution$ AND opposition$) OR TS =((politic* OR policy OR policies) AND (civil OR civic) AND power) OR TS =((politic* OR policy OR policies) AND (civil OR civic) AND citizen$) OR TS =((politic* OR policy OR policies) AND (civil OR civic) AND gender) OR TS =((politic* OR policy OR policies) AND (civil OR civic) AND nation*) OR TS =((politic* OR policy OR policies) AND (civil OR civic) AND interest$) OR TS =((politic* OR policy OR policies) AND (civil OR civic) AND labo$r) OR TS =((politic* OR policy OR policies) AND (civil OR civic) AND opposition$) OR TS =((politic* OR policy OR policies) AND power AND citizen$) OR TS =((politic* OR policy OR policies)

AND power AND gender) OR TS =((politic* OR policy OR policies) AND power AND nation*) OR TS =((politic* OR policy OR policies) AND power AND interest$) OR TS =((politic* OR policy OR policies) AND power AND labo$r) OR TS =((politic* OR policy OR policies) AND power AND opposition$) OR TS =((politic* OR policy OR policies) AND citizen$ AND gender) OR TS =((politic* OR policy OR policies) AND citizen$ AND nation*) OR TS =((politic* OR policy OR policies) AND citizen$ AND interest$) OR TS =((politic* OR policy OR policies) AND citizen$ AND labo$r) OR TS =((politic* OR policy OR policies) AND citizen$ AND opposition$) OR TS =((politic* OR policy OR policies) AND gender AND nation*) OR TS =((politic* OR policy OR policies) AND gender AND interest$) OR TS =((politic* OR policy OR policies) AND gender AND labo$r) OR TS =((politic* OR policy OR policies) AND gender AND opposition$) OR TS =((politic* OR policy OR policies) AND nation* AND interest$) OR TS =((politic* OR policy OR policies) AND nation* AND labo$r) OR TS =((politic* OR policy OR policies) AND nation* AND opposition$) OR TS =((politic* OR policy OR policies) AND interest$ AND labo$r) OR TS =((politic* OR policy OR policies) AND interest$ AND opposition$) OR TS =((politic* OR policy OR policies) AND labo$r AND opposition$) OR TS =((state OR states) AND conflict$ AND (election$ OR elector*)) OR TS =((state OR states) AND conflict$ AND countr*) OR TS =((state OR states) AND conflict$ AND war OR wars) OR TS =((state OR states) AND conflict$ AND govern*) OR TS =((state OR states) AND conflict$ AND (party OR parties)) OR TS =((state OR states) AND conflict$ AND democra*) OR TS =((state OR states) AND conflict$ AND institution$) OR TS =((state OR states) AND conflict$ AND (civil OR civic)) OR TS =((state OR states) AND conflict$ AND power) OR TS =((state OR states) AND conflict$ AND citizen$) OR TS =((state OR states) AND conflict$ AND gender) OR TS =((state OR states) AND conflict$ AND nation*) OR TS =((state OR states) AND conflict$ AND interest$) OR TS =((state OR states) AND conflict$ AND labo$r) OR TS =((state OR states) AND conflict$ AND opposition$) OR TS =((state OR states) AND (election$ OR elector*) AND countr*) OR TS =((state OR states) AND (election$ OR elector*) AND war OR wars) OR TS =((state OR states) AND (election$ OR elector*) AND govern*) OR TS =((state OR states) AND (election$ OR elector*) AND (party OR parties)) OR TS =((state OR states) AND (election$ OR elector*) AND democra*) OR TS =((state OR states) AND (election$ OR elector*) AND institution$) OR TS =((state OR states) AND (election$ OR elector*) AND (civil OR civic)) OR TS =((state OR states) AND (election$ OR elector*) AND power) OR TS =((state OR states) AND (election$ OR elector*) AND citizen$) OR TS =((state OR states) AND (election$ OR elector*) AND gender) OR TS =((state OR states) AND (election$ OR elector*) AND nation*) OR TS =((state OR states) AND (election$ OR elector*) AND interest$) OR TS =((state OR states) AND (election$ OR elector*) AND labo$r) OR TS =((state OR states) AND (election$ OR elector*) AND opposition$) OR TS =((state OR states) AND countr* AND war OR wars) OR TS =((state OR states) AND countr* AND govern*) OR TS =((state OR states) AND countr* AND (party OR parties)) OR TS =((state OR states) AND countr* AND democra*) OR TS =((state OR states) AND countr* AND institution$) OR TS =((state OR states) AND countr* AND (civil OR civic)) OR TS =((state OR states) AND countr* AND power) OR TS =((state OR states) AND countr* AND citizen$) OR TS =((state OR states) AND countr* AND gender) OR TS =((state OR states) AND countr* AND nation*) OR TS =((state OR states) AND countr* AND interest$) OR TS =((state OR states) AND countr* AND labo$r) OR TS =((state OR states) AND countr* AND opposition$) OR TS =((state OR states) AND war OR wars AND govern*) OR TS =((state OR states) AND war OR wars AND (party OR parties)) OR TS =((state OR states) AND war OR wars AND democra*) OR TS =((state OR states) AND war OR wars AND institution$) OR TS =((state OR states) AND war OR wars AND (civil OR civic)) OR TS =((state OR states) AND war OR wars AND power) OR TS =((state OR states) AND war OR wars AND citizen$) OR TS =((state OR states) AND war OR wars AND gender) OR TS =((state OR states) AND war OR wars AND nation*) OR TS =((state OR states) AND war OR wars AND interest$) OR TS =((state OR states) AND war OR wars AND labo$r) OR TS =((state OR states) AND war OR wars AND opposition$) OR TS =((state OR states) AND govern* AND (party OR parties)) OR TS =((state OR states) AND govern* AND democra*) OR TS =((state OR states) AND govern* AND institution$) OR TS =((state OR states) AND govern* AND (civil OR civic)) OR TS =((state OR states) AND govern* AND power) OR TS =((state OR states) AND govern* AND citizen$) OR TS =((state OR states) AND govern* AND gender) OR TS =((state OR states) AND govern* AND nation*) OR TS =((state OR states) AND govern* AND interest$) OR TS =((state OR states) AND govern* AND labo$r) OR TS =((state OR states) AND govern* AND opposition$) OR TS =((state OR states) AND (party OR parties) AND democra*) OR TS =((state OR states) AND (party OR parties) AND institution$) OR TS =((state OR states) AND (party OR parties) AND (civil OR civic)) OR TS =((state OR states) AND (party OR parties) AND power) OR TS =((state OR states) AND (party OR parties) AND citizen$) OR TS =((state OR states) AND (party OR parties) AND gender) OR TS =((state OR states) AND (party OR parties) AND nation*) OR TS =((state OR states) AND (party OR parties) AND interest$) OR TS =((state OR states) AND (party OR parties) AND labo$r) OR TS =((state OR states) AND (party OR parties) AND opposition$) OR TS =((state OR states) AND democra* AND institution$) OR TS =((state OR states) AND democra* AND (civil OR civic)) OR TS =((state OR states) AND democra* AND power) OR TS =((state OR states) AND democra* AND citizen$) OR TS =((state OR states) AND democra* AND gender) OR TS =((state OR states) AND democra* AND nation*) OR TS =((state OR states) AND democra* AND interest$) OR TS =((state OR states) AND democra* AND labo$r) OR TS =((state OR states) AND democra* AND opposition$) OR TS =((state OR states) AND institution$ AND (civil OR civic)) OR TS =((state OR states) AND institution$ AND power) OR TS =((state OR states) AND institution$ AND citizen$) OR TS =((state OR states) AND institution$ AND gender) OR TS =((state OR states) AND institution$ AND nation*) OR TS =((state OR states) AND institution$ AND interest$) OR TS =((state OR states) AND institution$ AND labo$r) OR TS =((state OR states) AND institution$ AND opposition$) OR TS =((state OR states) AND (civil OR civic) AND power) OR TS =((state OR states) AND (civil OR civic) AND citizen$) OR TS =((state OR states) AND (civil OR civic) AND gender) OR TS =((state OR states) AND (civil OR civic) AND nation*) OR TS =((state OR states) AND (civil OR civic) AND interest$) OR TS =((state OR states) AND (civil OR civic) AND labo$r) OR TS =((state OR states) AND (civil OR civic) AND opposition$) OR TS =((state OR states) AND power AND citizen$) OR TS =((state OR states) AND power AND gender) OR TS =((state OR states) AND power AND nation*) OR TS =((state OR states) AND power AND interest$) OR TS =((state OR states) AND power AND labo$r) OR TS =((state OR states) AND power AND opposition$)

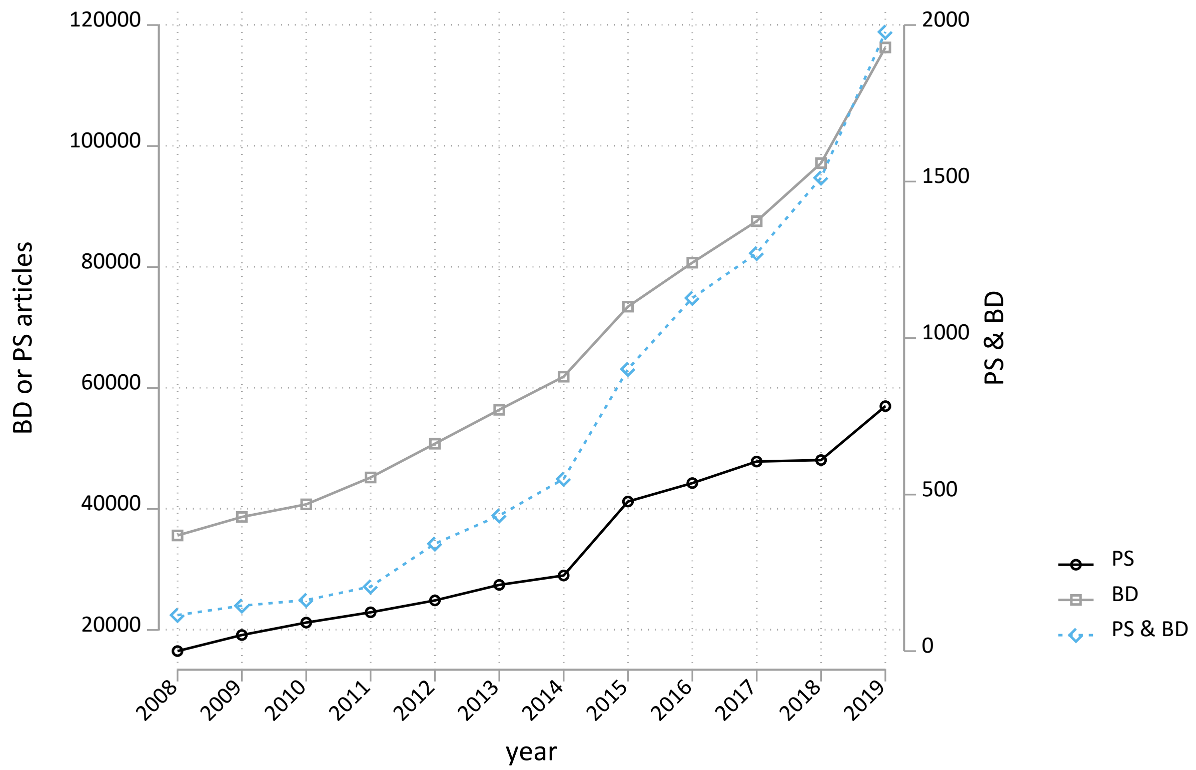
OR TS =((state OR states) AND citizen$ AND gender) OR TS =((state OR states) AND citizen$ AND nation*) OR TS =((state OR states) AND citizen$ AND interest$) OR TS =((state OR states) AND citizen$ AND labo$r) OR TS =((state OR states) AND citizen$ AND opposition$) OR TS =((state OR states) AND gender AND nation*) OR TS =((state OR states) AND gender AND interest$) OR TS =((state OR states) AND gender AND labo$r) OR TS =((state OR states) AND gender AND opposition$) OR TS =((state OR states) AND nation* AND interest$) OR TS =((state OR states) AND nation* AND labo$r) OR TS =((state OR states) AND nation* AND opposition$) OR TS =((state OR states) AND interest$ AND labo$r) OR TS =((state OR states) AND interest$ AND opposition$) OR TS =((state OR states) AND labo$r AND opposition$) OR TS =(conflict$ AND (election$ OR elector*) AND countr*) OR TS =(conflict$ AND (election$ OR elector*) AND war OR wars) OR TS =(conflict$ AND (election$ OR elector*) AND govern*) OR TS =(conflict$ AND (election$ OR elector*) AND (party OR parties)) OR TS =(conflict$ AND (election$ OR elector*) AND democra*) OR TS =(conflict$ AND (election$ OR elector*) AND institution$) OR TS =(conflict$ AND (election$ OR elector*) AND (civil OR civic)) OR TS =(conflict$ AND (election$ OR elector*) AND power) OR TS =(conflict$ AND (election$ OR elector*) AND citizen$) OR TS =(conflict$ AND (election$ OR elector*) AND gender) OR TS =(conflict$ AND (election$ OR elector*) AND nation*) OR TS =(conflict$ AND (election$ OR elector*) AND interest$) OR TS =(conflict$ AND (election$ OR elector*) AND labo$r) OR TS =(conflict$ AND (election$ OR elector*) AND opposition$) OR TS =(conflict$ AND countr* AND war OR wars) OR TS =(conflict$ AND countr* AND govern*) OR TS =(conflict$ AND countr* AND (party OR parties)) OR TS =(conflict$ AND countr* AND democra*) OR TS =(conflict$ AND countr* AND institution$) OR TS =(conflict$ AND countr* AND (civil OR civic)) OR TS =(conflict$ AND countr* AND power) OR TS =(conflict$ AND countr* AND citizen$) OR TS =(conflict$ AND countr* AND gender) OR TS =(conflict$ AND countr* AND nation*) OR TS =(conflict$ AND countr* AND interest$) OR TS =(conflict$ AND countr* AND labo$r) OR TS =(conflict$ AND countr* AND opposition$) OR TS =(conflict$ AND war OR wars AND govern*) OR TS =(conflict$ AND war OR wars AND (party OR parties)) OR TS =(conflict$ AND war OR wars AND democra*) OR TS =(conflict$ AND war OR wars AND institution$) OR TS =(conflict$ AND war OR wars AND (civil OR civic)) OR TS =(conflict$ AND war OR wars AND power) OR TS =(conflict$ AND war OR wars AND citizen$) OR TS =(conflict$ AND war OR wars AND gender) OR TS =(conflict$ AND war OR wars AND nation*) OR TS =(conflict$ AND war OR wars AND interest$) OR TS =(conflict$ AND war OR wars AND labo$r) OR TS =(conflict$ AND war OR wars AND opposition$) OR TS =(conflict$ AND govern* AND (party OR parties)) OR TS =(conflict$ AND govern* AND democra*) OR TS =(conflict$ AND govern* AND institution$) OR TS =(conflict$ AND govern* AND (civil OR civic)) OR TS =(conflict$ AND govern* AND power) OR TS =(conflict$ AND govern* AND citizen$) OR TS =(conflict$ AND govern* AND gender) OR TS =(conflict$ AND govern* AND nation*) OR TS =(conflict$ AND govern* AND interest$) OR TS =(conflict$ AND govern* AND labo$r) OR TS =(conflict$ AND govern* AND opposition$) OR TS =(conflict$ AND (party OR parties) AND democra*) OR TS =(conflict$ AND (party OR parties) AND institution$) OR TS =(conflict$ AND (party OR parties) AND (civil OR civic)) OR TS =(conflict$ AND (party OR parties) AND power) OR TS =(conflict$ AND (party OR parties) AND citizen$) OR TS =(conflict$ AND (party OR parties) AND gender) OR TS =(conflict$ AND (party OR parties) AND nation*) OR TS =(conflict$ AND (party OR parties) AND interest$) OR TS =(conflict$ AND (party OR parties) AND labo$r) OR TS =(conflict$ AND (party OR parties) AND opposition$) OR TS =(conflict$ AND democra* AND institution$) OR TS =(conflict$ AND democra* AND (civil OR civic)) OR TS =(conflict$ AND democra* AND power) OR TS =(conflict$ AND democra* AND citizen$) OR TS =(conflict$ AND democra* AND gender) OR TS =(conflict$ AND democra* AND nation*) OR TS =(conflict$ AND democra* AND interest$) OR TS =(conflict$ AND democra* AND labo$r) OR TS =(conflict$ AND democra* AND opposition$) OR TS =(conflict$ AND institution$ AND (civil OR civic)) OR TS =(conflict$ AND institution$ AND power) OR TS =(conflict$ AND institution$ AND citizen$) OR TS =(conflict$ AND institution$ AND gender) OR TS =(conflict$ AND institution$ AND nation*) OR TS =(conflict$ AND institution$ AND interest$) OR TS =(conflict$ AND institution$ AND labo$r) OR TS =(conflict$ AND institution$ AND opposition$) OR TS =(conflict$ AND (civil OR civic) AND power) OR TS =(conflict$ AND (civil OR civic) AND citizen$) OR TS =(conflict$ AND (civil OR civic) AND gender) OR TS =(conflict$ AND (civil OR civic) AND nation*) OR TS =(conflict$ AND (civil OR civic) AND interest$) OR TS =(conflict$ AND (civil OR civic) AND labo$r) OR TS =(conflict$ AND (civil OR civic) AND opposition$) OR TS =(conflict$ AND power AND citizen$) OR TS =(conflict$ AND power AND gender) OR TS =(conflict$ AND power AND nation*) OR TS =(conflict$ AND power AND interest$) OR TS =(conflict$ AND power AND labo$r) OR TS =(conflict$ AND power AND opposition$) OR TS =(conflict$ AND citizen$ AND gender) OR TS =(conflict$ AND citizen$ AND nation*) OR TS =(conflict$ AND citizen$ AND interest$) OR TS =(conflict$ AND citizen$ AND labo$r) OR TS =(conflict$ AND citizen$ AND opposition$) OR TS =(conflict$ AND gender AND nation*) OR TS =(conflict$ AND gender AND interest$) OR TS =(conflict$ AND gender AND labo$r) OR TS =(conflict$ AND gender AND opposition$) OR TS =(conflict$ AND nation* AND interest$) OR TS =(conflict$ AND nation* AND labo$r) OR TS =(conflict$ AND nation* AND opposition$) OR TS =(conflict$ AND interest$ AND labo$r) OR TS =(conflict$ AND interest$ AND opposition$) OR TS =(conflict$ AND labo$r AND opposition$) OR TS =((election$ OR elector*) AND countr* AND war OR wars) OR TS =((election$ OR elector*) AND countr* AND govern*) OR TS =((election$ OR elector*) AND countr* AND (party OR parties)) OR TS =((election$ OR elector*) AND countr* AND democra*) OR TS =((election$ OR elector*) AND countr* AND institution$) OR TS =((election$ OR elector*) AND countr* AND (civil OR civic)) OR TS =((election$ OR elector*) AND countr* AND power) OR TS =((election$ OR elector*) AND countr* AND citizen$) OR TS =((election$ OR elector*) AND countr* AND gender) OR TS =((election$ OR elector*) AND countr* AND nation*) OR TS =((election$ OR elector*) AND countr* AND interest$) OR TS =((election$ OR elector*) AND countr* AND labo$r) OR TS =((election$ OR elector*) AND countr* AND opposition$) OR TS =((election$ OR elector*) AND war OR wars AND govern*) OR TS =((election$ OR elector*) AND war OR wars AND (party OR parties)) OR TS =((election$ OR elector*) AND war OR wars AND democra*) OR TS =((election$ OR elector*) AND war OR wars AND institution$) OR TS =((election$ OR elector*) AND war OR wars AND (civil OR civic)) OR TS =((election$ OR elector*) AND war OR wars AND power) OR TS =((election$ OR elector*) AND war OR wars AND citizen$) OR TS =((election$ OR elector*) AND war OR wars AND gender) OR TS

=((election$ OR elector*) AND war OR wars AND nation*) OR TS =((election$ OR elector*) AND war OR wars AND interest$) OR TS =((election$ OR elector*) AND war OR wars AND labo$r) OR TS =((election$ OR elector*) AND war OR wars AND opposition$) OR TS =((election$ OR elector*) AND govern* AND (party OR parties)) OR TS =((election$ OR elector*) AND govern* AND democra*) OR TS =((election$ OR elector*) AND govern* AND institution$) OR TS =((election$ OR elector*) AND govern* AND (civil OR civic)) OR TS =((election$ OR elector*) AND govern* AND power) OR TS =((election$ OR elector*) AND govern* AND citizen$) OR TS =((election$ OR elector*) AND govern* AND gender) OR TS =((election$ OR elector*) AND govern* AND nation*) OR TS =((election$ OR elector*) AND govern* AND interest$) OR TS =((election$ OR elector*) AND govern* AND labo$r) OR TS =((election$ OR elector*) AND govern* AND opposition$) OR TS =((election$ OR elector*) AND (party OR parties) AND democra*) OR TS =((election$ OR elector*) AND (party OR parties) AND institution$) OR TS =((election$ OR elector*) AND (party OR parties) AND (civil OR civic)) OR TS =((election$ OR elector*) AND (party OR parties) AND power) OR TS =((election$ OR elector*) AND (party OR parties) AND citizen$) OR TS =((election$ OR elector*) AND (party OR parties) AND gender) OR TS =((election$ OR elector*) AND (party OR parties) AND nation*) OR TS =((election$ OR elector*) AND (party OR parties) AND interest$) OR TS =((election$ OR elector*) AND (party OR parties) AND labo$r) OR TS =((election$ OR elector*) AND (party OR parties) AND opposition$)) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article). **Indexes**=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, **ESCI Timespan**=2008-2019

B. Operationalization of big data

B.1 Distribution of raw search results over time

Note that BDPS hits are represented on the Y axis to the *right*, while PS (non-BD) records and BD (non-PS records) are represented on the Y axis to the *left*.



B.2. Examples of abstracts of *irrelevant* and *excluded* records

Terms from the **political science** and **big data** search strings are highlighted to illustrate how these records were captured by the search.

Abstract coded as *irrelevant*
Renewable energies are attracting considerable interest as an appropriate alternative for the fossil fuel based sources of energy due to increasing concerns about global warming and environmental issues. In this regard, this paper takes a new look at designing a sustainable Switchgrass-based bioenergy supply chain network (BSCN) incorporating **conflicting** economic, environmental and **social** objectives. We propose a novel, multi-objective, mixed integer linear programming model as a decision-making (DM) tool. To deal with sustainability factors, the proposed model is solved using two-stage **algorithms** consisting of augmented e-constraint and TOPSIS. This approach paves the way to determine the suitable strategical and tactical decisions and manage bioenergy supply chain performance with respect to the preference of decision makers for appropriate trade-off among of the sustainability factors. Computational analysis is carried out to indicate the validity of the proposed model by using a real case study in Iran. The results demonstrate that achieving a desirable level of **social** and environmental preservation conducted a circa 15% increase in economic objective function at the end of planning horizon. The results also show that high investment cost is a precondition to improving BSCN. It behooves the **government** to prioritize their plans and compartmentalize their budgets and spending in more constructive and effective ways related to bioenergy supply chain planning.

Reference: Rabbani, M., Saravi, N. A., Farrokhi-Asl, H., Lim, S. F. W., & Tahaei, Z. (2018). Developing a sustainable supply chain optimization model for switchgrass-based bioenergy production: A case study. *Journal of Cleaner Production*, *200*, 827-843.

Abstract from entry coded as *excluded*
Political parties and candidates have not been immune to the changes that the Internet and **social media** have introduced in **electoral** campaigns. Yet, as the use of digital media by **political** elites is becoming a norm in the United States, in Europe, the decision to develop an online presence depends on the cross-**nation**al differences regarding candidates' constraints and incentives. European Parliament elections present an exceptional comparative opportunity to measure this potential diversity. Using an original database on the online presence of more than 5000 candidates competing under the label of incumbent parties in 2014, we demonstrate that there are two relevant groups of nonadopters, and that candidates' online campaign intensity varies significantly depending on incumbency and the ballot structure in their countries.

Reference: Lorenzo Rodriguez, J., & Garmendia Madariaga, A. (2016). Going public against institutional constraints? Analyzing the online presence intensity of 2014 European Parliament election candidates. *European Union Politics*, *17*(2), 303-323.

B.3. Coding BDPS

As noted in the main manuscript, we used a single research assistant to classify the records from the BDPS search as big data (to be included in the dataset) or not (excluded from subsequent work; see Figure 2 in the main manuscript. The research assistant was informed about the different steps we had undertaken to develop a coding protocol, i.e., the pilot coding and the subsequent work to reach agreement on all coding decisions.

As a general rule, we begin with the following definition and description of Big Data, where the *repurposing of existing data for scientific use* is the key component:

> Big Data is data that is repurposed, in a variety of different ways. This repurposing provides social scientists with access to (e.g.) sensor data, satellite and measurement data, social media exhaust and geotags, that can be "borrowed" from their original mission and repurposed quickly and cheaply for subsequent social scientific analysis. Such a definition taps into the new tools that are being used to repurpose (e.g., scraping, machine learning, etc. that allows us to use "digital footprints" in real time, over a number of platforms), as well as reflect the enormous scale (hence "Big") that this repurposing provides (in terms of volume, velocity and variety).

Over the course of coding our abstracts, several tools and techniques for handling datasets of this size were identified, among them:

- Automatic content analysis;
- Scraping/web crawling;
- Network analysis;
- Sentiment analysis and topic modelling; and
- Machine learning, both supervised and un-supervised.

How the coding was done

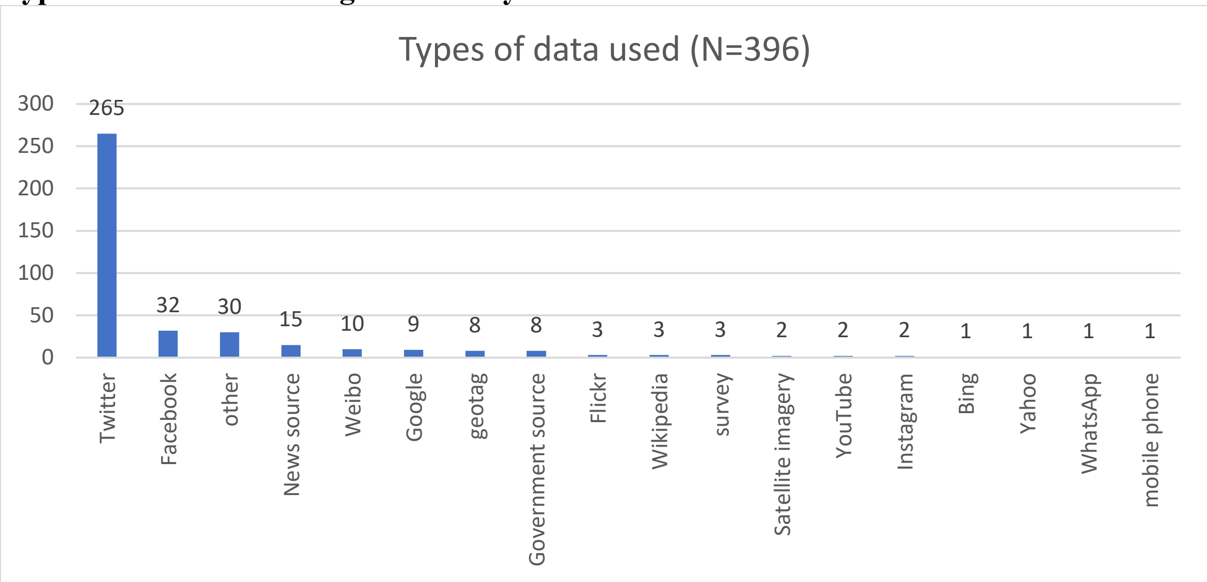
When coding the manuscripts, our RA distinguished among four categories: *Include* (the article is both PS and BD), *doubt* (needs further reading), *exclude* (PS, but not BD) and *junk* (totally unrelated to our project).

After an initial stage of coding, the articles deemed *include* and *doubt* where further inspected. In order to determine if an article falls within the realm of Big Data, the coder had to assess the data in the article (i.e., move beyond the abstract).

First, the defining theoretical assessment was whether the data had been collected for research purposes or repurposed. The methods chapter of the articles usually gave an answer to this question (for instance, a number of articles used survey data without this being clear from the abstract alone).

The coder then identified the type of data being used. In most cases the data can be described as "social media exhaust", meaning that researchers are tapping into the firehose of different social media platforms in order to produce large datasets. Although social media data seems to be most popular among researchers, other and more creative sources of data are also being used, for instance data obtained from mobile phone operators (Blumenstock et al. 2015), Flickr photos (Levin, Ali & Crandall, 2018), Google search logs (Street et al. 2015), public procurement records (Fazekas, Tóth and King, 2016), news articles (Gatterman, 2018) or geo-tag data (Ma et al. 2018). The result was a comprehensive list of different data types used, as depicted in the following histogram.

Types of Data Used in Big Data Analyses



Note: This data was derived from the 355 valid cases. The total number of coded data types exceeded this (N=396), as many of the studies relied on more than one type of data.

In order to use the data, two additional sources of information were collected for each of our valid cases: the nature of the data used, and the size of N (as far as this was possible). In most cases, obtaining this information meant reading the methods chapter (and in some cases, the whole article). This process resulted in a list of 355 articles judged to be valid cases, i.e., PS+BD.

The problem of "small" Big Data

On closer inspection, we found that many of our articles used social media data in the form of more contextual analyses of the digital behaviour of a certain group of individuals, for instance members of the U.S Congress (Peterson, 2012) or members of the German Bundestag (Geber and Scherer, 2015). When establishing the number of observations (N) for these articles it became clear that these cases are more akin to traditional political science, as in most cases the data were "small" enough to employ traditional techniques or manual coding, even though the data collected are "footprints of digital behaviour." These articles fall somewhere in between traditional political science and big data research, as they use a similar method for data extraction as articles with a much higher N, but they limit this extraction to certain number of individuals (for example, politicians) or institutions (for example, newspapers). In the end, we decided not to include these articles as a part of our valid cases.

Why so much Twitter?

Twitter data is remarkably over-represented in our dataset (see Figure above). According to Vargo et al. (2014), researchers tend to use Twitter data for a number of reasons. First, accessibility: Twitter posts are public and relatively short (max 140 characters) which eases the analytical work (e.g. text cleaning). Second, Twitter API are easier for researchers to access then other social media platforms. Third, Twitter allows for researchers to examine both private individuals, public figures, media outlets and other agents as they all operate simultaneously within the same environment. We are aware of the dangers associated with Twitter data (see, e.g., Boyd and Crawford, 2012), but our intent is to map recent developments, not to judge them.

B.4. Coding BDPS using machine learning

Two RAs, undergoing MScs in information technology, were hired to develop an algorithm that would identify BDPS articles based on the abstract. This exercise was done with the truncated dataset (2008-2018) coded for “political science topic” and “using big data”, before the 2019 data was available.

The nature of this task can be seen as a standard binary classification problem (Alpaydin, 2014). This means that we could choose from a number of pretrained models. The RAs ended up using a Bert-base-cased (Devlin et al., 2018) approach for its state-of-the-art results when working on smaller data sets (Howe et al., 2019). Even though the Bert-Large pretrained model performed better, technical limitations kept us from using it.

Our dataset was then preprocessed to fit the input nodes of the Bert network and was split into a training (90%) and evaluation set (10%). The Matthews Correlation Coefficient (MCC) is used as our metric of quality to reduce bias from relying on a larger set of invalids (Boughorbel et al., 2017). The best evaluation results produced a Matthews Correlation Coefficient of 0.507. It is clear that our data included a higher number of false negatives relative to the true positives (see table below). On the other hand, true negatives scored much higher than false positives.

Machine Learning Results

Statistic	1 Epoch	10 Epochs	20 Epochs	100 Epochs	Highest MCC Value (Epoch 24)
MCC	0.372	0.423	0.456	0.260	0.507
True Positive	11	12	11	4	11
False Negative	18	17	18	25	18
True Negative	430	432	437	441	440
False Positive	14	12	7	3	4

With an MCC higher than 0 the model has a higher rate of predicting correctly (than random guessing). After the best result (at Epoch 24), the machine started overfitting, which resulted in a lower MCC (see Dietterich, 1995). All in all, we concluded that our dataset included too few valid cases to provide a solid foundation for a machine learning exercise.

C. The baseline group

The table below shows the result from the “PS not BD” search in Web of Science by publication year, from which we drew a quasi-random sample, stratified by year.

C.1 Records in Web of Science from which baseline group was sampled

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
N	1,978	2,311	2,730	2,734	2,970	3,057	3,338	4,545	4,779	4,870	5,101	5,601	44,014

D. Coding transparency

Two research assistants coded the two datasets (BDPS and baseline group) for transparency. Both datasets were randomly split into two groups, so that each RA coded 50% of each dataset.

D.1. JETS signatories

American Journal of Political Science
American Political Science Review
American Politics Research
British Journal of Political Science
Comparative Political Studies
Conflict Management and Peace Science
Cooperation and Conflict
European Journal of Political Research
European Political Science
European Union Politics
International Interactions
International Security
Journal of Conflict Resolution
Journal of Elections, Public Opinion, and Parties
Journal of Experimental Political Science
Journal of European Public Policy
Journal of Peace Research
Journal of Theoretical Politics
Quarterly Journal of Political Science
Party Politics
Political Analysis
Political Behavior
Political Communication
Political Science Research and Methods
Research and Politics
Rivista Italiana di Scienza Politica
State Politics and Policy Quarterly
The Journal of Politics
The Political Methodologist

Source: The DA-RT statement, available at https://www.dartstatement.org/2014-journal-editors-statement-jets

D.2. Training of research assistants

The initial training consisted of an explanation of the codebook that they were to use. The research team also went through a set of pre-coded observations with different values on the variables. The RAs then coded a set of observations individually and we compared their results. Any discrepancies were discussed and solved. After the training period, the RAs met once a week to resolve doubts and difficult cases.

D.3. Codebook

The following variables were coded:

- *Empirical*: Whether the article employs an empirical research design or not (yes/no)
- *Quantqual*: If empirical: the nature of the research design (qualitative, quantitative, or mixed)
- *Dataset*: Whether the dataset used to produce the results reported in the article was found to be available or not, i.e., published by author(s) or journal (yes/no). In theory, a value of "no" does not exclude the possibility that the dataset is available online somewhere, but it could not be found with reasonable effort (as described in the four-step procedure in the main manuscript).
- *Code:* Whether the code to produce the results reported in the article was found to be available or not, i.e., published by author(s) or journal (yes/no).
- *On request*: Replication material is stated to be available on request.
- *Inarticle*: Replication material is stated by article author(s) to be available on a given location, including the journal's and/or author's homepage.

D.4. Reliability

Despite efforts to ensure consistent coding, including close cooperation between the RAs, the final datasets indicate that one RA has a stricter understanding of what constitutes "available" replication material. In other words, the two RAs have significantly different transparency scores.

Mean transparency scores, by dataset and RAs

	Overall	BDPS	Baseline group
RA 1	.275	0.116	0.322
RA 2	.198	0.042	0.231
Difference	*.077**	*0.073**	*0.081**
N	1,519	338	1,181

Note: * p(diff>0) < 0.001. Missingness is due to some observations being coded by both RAs or the RAs together with the authors.

While not ideal, we do not see this as a serious problem for our analysis. First, since both datasets were split into two equal parts and randomly assigned to each RA, the dataset is still balanced across 'treatment' (BDPS) and control group, and the lower reliability should not confound any real difference in transparency between the two datasets.

E. Descriptive statistics and analysis

E.1. Publication outlets for BDPS articles, by PS – non-PS journal

non-PS journals		PS journals	
Journal name	**N**	**Journal name**	**N**
Social Science Computer Review	21	*Journal of Information Technology & P..*	12
Plos One	20	*American Political Science Review*	3
Information Communication & Society	14	*International Journal of Press-Politics*	3
New Media & Society	13	*Policy and Internet*	3
Social Media + Society	8	*Political Analysis*	3
American Behavioral Scientist	6	*Political Communication*	3
Epj Data Science	6	*Electoral Studies*	2
Government Information Quarterly	6	*German Politics*	2
Social Network Analysis and Mining	6	*Party Politics*	2
Sustainability	5	*Regulation & Governance*	2
Communication & Society-Spain	4	*Studies in Conflict & Terrorism*	2
Media and Communication	4	*American Journal of Political Science*	1
Technological Forecasting and Social ..	4	*American Politics Research*	1
Big Data & Society	3	*Annals of the American Academy of Pol..*	1
Expert Systems with Applications	3	*Australian Journal of Political Science*	1
International Communication Gazette	3	*British Journal of Political Science*	1
International Journal of Communication	3	*Comparative Politics*	1
Isprs International Journal of Geo-In..	3	*Democratization*	1
Media War and Conflict	3	*Journal of Conflict Resolution*	1
Science	3	*Journal of Public Policy*	1
Social Networks	3	*Local Government Studies*	1
Applied Geography	2	*Perspectives on Politics*	1
Asian Journal of Communication	2	*Policy Studies Journal*	1
Aslib Journal of Information Management	2	*Political Psychology*	1
Big Data	2	*Political Science Research and Methods*	1
Computers in Human Behavior	2	*Presidential Studies Quarterly*	1
European Journal of Communication	2	*Public Opinion Quarterly*	1
Global Environmental Change-Human and..	2	*Terrorism and Political Violence*	1
Global Media and Communication	2		
Ieee Access	2		
Industrial Management & Data Systems	2		
Information Research-an International..	2		
Information Systems Frontiers	2		
International Journal of Advanced Com..	2		
International Journal of Geographical..	2		
International Journal of Market Resea..	2		
Jcom-Journal of Science Communication	2		
Journal of Communication	2		
Journal of Language and Politics	2		
Journalism Practice	2		
Nuclear Engineering and Technology	2		
Online Information Review	2		
Partecipazione E Conflitto	2		
Physica a-Statistical Mechanics and I..	2		

Policy Studies	2
Proceedings of the National Academy o..	2
Psychological Science	2
Sage Open	2
Telematics and Informatics	2
Tripodos	2

Note: non-PS journals that appear with one article each: *Acm Transactions on the Web, African Journal of Science Technology.. , African Security Review, American Sociological Review, Applied Sciences-Basel , Asian Journal of Political Science, Asian Studies Review, Canadian Foreign Policy, Canadian Public Administration-Admini.., Cartography and Geographic Informatio.., China Quarterly, Climatic Change, Communication & Sport, Communication Research and Practice, Communication Review, Complexity, Computer, Computer Journal, Computer Supported Cooperative Work-t.., Comunicar, Conflict Management and Peace Science, Contemporary Social Science, Convergence-the International Journal.., Critical Perspectives on Accounting, Data, Democracy & Security, Drug and Alcohol Dependence, Drustvena Istrazivanja, Educational Policy, Election Law Journal, Energy Policy Engineering, Applications of Artificia.., Environmental Communication-a Journal.., Environmental Science & Policy, Ethnic and Racial Studies, European Journal on Criminal Policy a.., European Societies, French Politics, German Politics and Society, Heliyon, Ieee Transactions on Computational So.., Ieee Transactions on Emerging Topics .., Ieee Transactions on Multimedia, Information Technologies & Internatio.., Information Technology & People, International Affairs, International Arab Journal of Informa.., International Journal of Cyber Crimin.., International Journal of E-Politics, International Journal of Environmenta.., International Journal of Forecasting, International Journal of Medical Info.., International Journal of Pervasive Co.., International Journal of Politics Cul.., Internet Research, Italian Political Science Review, Javnost-the Public, Journal of Balkan and near Eastern St.., Journal of Choice Modelling, Journal of Cleaner Production, Journal of Comparative Economics, Journal of Computer-Mediated Communic.., Journal of Construction Engineering a.., Journal of Contemporary China, Journal of Elections Public Opinion a.., Journal of Ethnic and Migration Studies, Journal of Experimental Psychology-Ge.., Journal of Information Science, Journal of Law and Courts, Journal of Marketing Research, Journal of Organizational and End Use.., Journal of Social Marketing, Journal of Universal Computer Science, Journal of the American Society for I.., Korea Observer, Mass Communication and Society, Media Culture & Society, Media International Australia, Mehran University Research Journal of.., Networks & Spatial Economics, New Journal of Physics, Nordicom Review, Profesional De La información, Professional Geographer, Program-Electronic Library and Inform.., Psychological Reports, Public Administration Review, Quality & Quantity, Revista De Cercetare Si Interventie S.., Revista De Comunicacion De La Seeci, Revista Internacional De Relaciones P.., Sadhana-Academy Proceedings in Engine.., Scientific Reports, Sexuality Research and Social Policy, Sociological Methodology, Sociological Methods & Research, Sociological Science, Southern Communication Journal, Transportation in Developing Economies, West European Politics, World Development,* and *Yale Law Journal.*

E.2. Descriptive statistics

Variable	**Mean**			**Min**	**Max**
	Overall	*BDPS*	*Baseline group*		
Transparency	0.233	0.076	0.280	0	1
Code	0.237	0.093	0.280	0	1
Dataset	0.275	0.124	0.319	0	1
RA id	0.500	0.488	0.503	0	1
Political science	-	0.152	-	0	1
JETS	0.286	0.054	0.355	0	1
N	1,555	355	1,200		

Note: Transparency (the dependent variable in all analyses) takes the value of 1 if both *code* and *dataset* are equal to 1. RA id is an id variable for the two research assistants who coded transparency. *Political science* is not coded for the baseline group, as they by definition are published in political science journals.

F. References

Alpaydin, Ethem (2014) "Introduction to Machine Learning." Third edition of *Adaptive Computation and Machine Learning.* (Cambridge, Mass: MIT Press).

Amina, B. and T. Azim (2019) "SCANCPECLENS: A Framework for Automatic Lexicon Generation and Sentiment Analysis of Micro Blogging Data on China Pakistan Economic Corridor." *Ieee Access:* 133876-133887

Blumenstock, J., Cadamuro, G., & On, R. (2015) Predicting poverty and wealth from mobile phone metadata. *Science 350*(6264), 1073-1076.

Boughorbel, Sabri Fethi Jarray, and Mohammed El-Anbari (2017) Optimal classifier for imbalanced data using Matthews correlation coefficient metric. *PloS one*, 12(6).

Boyd, D., & Crawford, K. (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Demirel-Pegg, T. and Moskowitz, J. (2009) "US Aid Allocation: The Nexus of Human Rights, Democracy, and Development." Journal of Peace Research 46(2): 181-198.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018) BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805.

Dietterich, Tom (1995) Overfitting and undercomputing in machine learning. *ACM computing surveys* (CSUR), 27(3): 326-327.

Fazekas, M., Tóth, I. J., & King, L. P. (2016). An objective corruption risk index using public procurement data. *European Journal on Criminal Policy and Research*, *22*(3), 369-397.

Gattermann, K. (2018) Mediated personalization of executive European Union politics: Examining patterns in the broadsheet coverage of the European Commission, 1992–2016. *The international journal of press/politics*, *23*(3), 345-366.

Geber, S., & Scherer, H. (2015) My voter, my party, and me: American and German parliamentarians on Facebook. *Journal of Information Technology & Politics*, *12*(4), 360-377.

Giles; Michael W. and James C. Garrand (2007) "Ranking Political Science Journals: Reputational and Citational Approaches." *PS* 40(4): 741-51.

Howe, Jerrold Soh Tsin, Lim How Khang, and Ian Ernst Chai (2019) Legal area classification: A comparative study of text classifiers on Singapore supreme court judgments. *CoRR*, abs/1904.06470.

Jensen, Magnus Rom and Jonathon W. Moses (2020) "The State of Political Science, 2020." . *European Political Science* (2020) 20(1): 14-33. https://doi.org/10.1057/s41304-020-00297-4

Levin, N., Ali, S., & Crandall, D. (2018) Utilizing remote sensing and big data to quantify conflict Intensity: The Arab Spring as a case study. *Applied geography*, *94*, 1-17.

Ma, T., Lu, R., Zhao, N., & Shaw, S. L. (2018). An estimate of rural exodus in China using location-aware data. *PLoS one*, *13*(7).

Monroe, B. L. (2013) The five Vs of big data political science introduction to the virtual issue on big data in political science political analysis. *Political Analysis*, *21*(5), 1-9.
Peterson, R. D. (2012). To tweet or not to tweet: Exploring the determinants of early adoption of Twitter by House members in the 111th Congress. *The Social Science Journal*, *49*(4), 430-438.
Street, A., Murray, T. A., Blitzer, J., & Patel, R. S. (2015). Estimating voter registration deadline effects with web search data. *Political Analysis*, *23*(2), 225-241.
Vargo, C. J., Guo, L., McCombs, M., & Shaw, D. L. (2014). Network issue agendas on Twitter during the 2012 US presidential election. *Journal of Communication*, *64*(2), 296-316.
Weaver, I. S., Williams, H., Cioroianu, I., Williams, M., Coan, T., & Banducci, S. (2018). Dynamic social media affiliations among UK politicians. *Social networks*, *54:*132-144.