

RESEARCH ARTICLE

Open data as an anticorruption tool? Using distributed cognition to understand breakdowns in the creation of transparency data

Tatiana M. Martinez¹  and Edgar A. Whitley² 

¹Regent's University London, London, UK

²Department of Management, London School of Economics and Political Science, London, UK

Corresponding author: Edgar A. Whitley; Email: E.A.Whitley@lse.ac.uk

Received: 02 December 2021; **Revised:** 19 December 2022; **Accepted:** 22 March 2023

Keywords: Anticorruption; data; data accuracy; distributed cognition; open data; transparency

Abstract

One of the drivers for pushing for open data as a form of corruption control stems from the belief that in making government operations more transparent, it would be possible to hold public officials accountable for how public resources are spent. These large datasets would then be open to the public for scrutiny and analysis, resulting in lower levels of corruption. Though data quality has been largely studied and many advancements have been made, it has not been extensively applied to open data, with some aspects of data quality receiving more attention than others. One key aspect however—accuracy—seems to have been overlooked. This gap resulted in our inquiry: how is accurate open data produced and how might breakdowns in this process introduce opportunities for corruption? We study a government agency situated within the Brazilian Federal Government in order to understand in what ways is accuracy compromised. Adopting a distributed cognition (DCog) theoretical framework, we found that the production of open data is not a neutral activity, instead it is a distributed process performed by individuals and artifacts. This distributed cognitive process creates opportunities for data to be concealed and misrepresented. Two models mapping data production were generated, the combination of which provided an insight into how cognitive processes are distributed, how data flow, are transformed, stored, and processed, and what instances provide opportunities for data inaccuracies and misrepresentations to occur. The results obtained have the potential to aid policymakers in improving data accuracy.

Policy Significance Statement

This research aims to analyze how data are produced to determine in what ways data accuracy can be compromised, thus undermining the effectiveness of open data in detecting levels of corruption in that environment. Looking at procurement processes undertaken by a Brazilian government agency, we conclude that human agents dictate how data are created, flow, and are stored despite the presence of technical systems that aimed at automating these processes. This allows for opportunities for data to be misrepresented and concealed, which in turn allows for corruption to continue to flourish. Policymakers can thus make use of the framework outlined in this study to both (a) improve the efficiency of procurement systems and (b) improve data accuracy so that open data can be adopted as a more effective anticorruption tool.

1. Introduction

Open data have been hailed by scholars and practitioners as a new and important way of enhancing levels of transparency and hence accountability within government (Bertot et al., 2010; Cranefield et al., 2014; Rose et al., 2015). The rationale underscoring this belief is that the more governments open themselves up to scrutiny by citizens, the better they will behave. The result has been an uptick of investments in transparency mechanisms including open data by governments across the world. Though there has been a great level of interest in how open government policies are positively correlated with high levels of accountability, less interest has been dedicated to understanding contexts where the deployment of open data did not result in more accountable governments.

One of the drivers for such shortcomings, and the focus here, stems from issues concerning data quality. Data quality encompasses different dimensions and their relative importance varies according to different scholars and perspectives (Wang and Strong, 1996; Scannapieco and Catarci, 2002; Bovee et al., 2003). While completeness and timeliness are frequently mentioned, other aspects such as accuracy (sometimes considered a standalone dimension, for others a subcategory of data integrity) are often overlooked. In Berners-Lee et al.'s (2006) framework for a 5-star open dataset, for example, this aspect is not mentioned. Instead, more attention is often paid as to whether the datasets produced conform to pre-established parameters (i.e., format). However, despite evidence that corrupt public officials will intentionally conceal and alter data to avoid detection, there is little research as to the impact of this on open datasets. Low quality open datasets may limit the ability of open data to curb corruption and hold public officials to account. Additionally, there has been limited study of how it might be possible to conceal or alter data in open datasets. The age-old adage of “garbage in, garbage out” means that assessments undertaken on the basis of inaccurate data could result in distorted analyses. Before exploring means of improving the accuracy of open data, an important objective is to better understand how the production of open data can be corrupted and this is the focus of this paper.

The paper studies the case of the Brazilian federal government, whose investments in open data have been considered a model to be followed (as evidenced by its leadership role in the Open Government Partnership Initiative). Moreover, Brazil is considered one of the most transparent countries in the world, ranking eighth in the Open Government Index in 2018. However, at the same time its political struggles and issues with corruption have been widely documented in recent years. As such, understanding the process of open data initiatives and management seems opportune toward highlighting what challenges and issues have undermined the Brazilian initiative toward curbing corruption and enhancing levels of accountability.

In order to study this process, this research conducted a case study, focusing on the open data efforts undertaken by one specific agency within the Brazilian government (referred to as Agency X) to respond to transparency legislation enacted in Brazil in 2009 and later expanded into the freedom of information act enacted in 2011. Agency X is a procurement agency, documenting its processes on information systems, the output of which is recorded and published on the Brazilian Transparency Portal, open to public scrutiny. This research therefore analyzed how Agency X's open data are created and managed and the opportunities for corruption that the processes enable.

This research adopted distributed cognition (DCog) as a theoretical framework. DCog draws attention to the cognitive distribution of work activities between individuals and artifacts as they interact within a particular environment, and how data flow within that context. DCog has been widely used in the human–computer interface literature.

Findings suggest cognitive processes are unevenly distributed between individuals and technical systems, with individuals leading the data flow and transformation, largely dictating how data are stored and processed. This has created opportunities for “breakdowns” in the data flow whereby the quality of data produced is compromised. Moreover, it allows for individuals to continue to engage in corruption thereby creating mechanisms to evade detection. Thus, investments in open data policies have failed to yield the expected results. This paper provides an in-depth account of how open data are created and limitations of it and is organized as follows: [Section 2](#) reviews background literature which frames this

research and help inform some of the lines of investigation pursued; [Section 3](#) presents more insight into the case study conducted and the methodology employed; [Section 4](#) presents a discussion of findings; and [Section 5](#) ends with concluding remarks.

2. Background

Though research has largely concentrated on analyzing the resulting datasets and evaluating it according to established dimensions for data quality (Davies, 2012; Peixoto, 2013; Lourenço, 2015; Dawes et al., 2016), less attention has been paid as to how data are produced and stored and the effects on data quality and its potential to curb corruption. This study thus aims to understand how open data are produced in practice so as to gain insights about its potential to effectively curb corruption. This section is dedicated toward reviewing the theory which underpins this study, reviewing research which has focused on uses of technology as an anticorruption deterrent, conceptualizing data quality, discussing how corruption is framed in this study, and reviewing DCoG which is the theoretical framework adopted here.

2.1. Technology and the fight against corruption

The anticorruption discourse gained traction in the mid-1990s, with a range of international organizations such as the United Nations, the International Monetary Fund, the World Bank, and Transparency International playing pivotal roles in raising awareness around the issues resulting from corrupt practices and lobby for measures which had the intent of curbing corruption (Bukovansky, 2006). This subsequently resulted in policies and recommendations on how to tackle it (Hopkin, 2002).

Among the recommendations that arose out of this movement was a push to reduce “information asymmetry” in order to strengthen oversight and control (Rothstein, 2011). This requires that governments make more information available to citizens, thereby creating a more open and transparent environment (Arellano-Gault and Lepore, 2011; Adam and Fazekas, 2021) and forcing governments to behave more honestly (Bertot et al., 2010).

Technological advances and the internet further amplified this discourse by creating platforms which allowed for wider dissemination of data (Heide and Villeneuve, 2020). As a result, the number of studies which sought to confirm the effectiveness of open data as a deterrent to corrupt practices grew. Indeed, part of a wider and more concerted investment effort in e-government, there is a significant body of research in this area which highlights the effectiveness of technology adoption as a deterrent for corruption (Bhatnagar, 2003; Shim and Eom, 2009). Regarding data transparency specifically, research strongly suggests a negative correlation between open data and corruption (Meijer, 2001; Bhatnagar, 2003; Pina et al., 2007; Andersen, 2009; Heald, 2012; Srivastava et al., 2016; Bauhr and Grimes, 2017; De Simone et al., 2017). This favorable outlook, coupled with perceived good governance ideals (Margetts, 2011) and legitimacy in the international stage (David-Barrett and Okamura, 2016), resulted in investments by governments worldwide in open platforms (Bertot et al., 2010). By the early 2000s, there were already over 14,000 agencies online worldwide (Norris, 2002). Likewise, in 1990 only two countries had Freedom of Information (FoIA) legislation regulating open government initiatives; by 2013, over 80 did (Calland and Bentley, 2013), and currently approximately 125 countries have enacted FoIA (IISD, 2019).

Despite the optimism surrounding open data as an anticorruption tool, there has been little research into the mechanisms that actually make it successful. There have been a few dissenting views regarding its effectiveness (Fox, 2007; Fox and Haight, 2010; Arellano-Gault and Lepore, 2011; Frølich, 2011; Gaventa and McGee, 2013); however, as observed by Bannister and Connolly (2011, p. 1), these are still “minority voices.” Moreover, as Adam and Fazekas (2021) note, “It is not yet clear under which conditions do ICTs facilitate rather than inhibit corruption” (2021, p. 2). The result has been that outliers to this proposition, such as Brazil and South Africa, both of which have consistently ranked high in the Open Budget Index but have struggled with grappling corruption, remain overlooked.

One of the reasons for this has been due to the type of studies conducted, which have been mostly quantitative in nature, none of which having documented negative outcomes as a result of investments in

technology by governments (Adam and Fazekas, 2021). There have been only a few qualitative cases which attest to negative outcomes, such as the study by Kossow and Dykes (2017) in Kenya, and cases across Eastern Europe presented by ReSPA (2013). These do not, however, look at uses of open data specifically, but instead document wider investments in technology. Moreover, though they provide evidence that corruption can continue to take place even after investments in technology have been made by governments, further details into how officials manage to go undetected are scarce.

These studies thus further corroborate the notion that technology alone is not enough to curb attempts at wrongdoing. With most of the literature pointing at the positive impact of digital technologies in the fight against corruption, the contextual factors that can limit the effectiveness of such tools have been an area which has been largely underexplored. Questions such as how government dynamics impact the deployment of such tools and in what ways anticorruption tools can be undermined by the existing context have thus far gone unaddressed.

Margetts (2011) notes that technology itself cannot be regarded as the means of institutional reform, but rather as a tool. Rose (2005) also observes that when thinking about the use of technology across government, national contexts and specificities cannot be ignored. Another point to be considered, according to Adam and Fazekas (2021, p. 4) is that technology can also create “vulnerabilities” and become the source of opportunities for new avenues of corruption. Among the vulnerabilities anticipated by the authors are hacking, the increased possibility of identifying public officials who may be amenable to receiving bribes, and reducing contact between officials and citizens, thus adding barriers to petty corruption. However the opportunities for data manipulation and concealment by officials, facilitated by the very systems designed to curb corruption, were not anticipated in their studies, anticipating that this is an area that merits further investigation.

In summary, there seems to be a high level of optimism concerning open data, and information technologies more generally, as a means for curbing corruption. However, there have also been some suggestion that technology may not be as transformative as initially anticipated, with a few cases demonstrating a limited impact. With regard to the use of open data specifically, as the Brazilian case denotes, making data more widely available will not on its own automatically translate into reduced corruption—it is not per se a transformative initiative (O’Neill, 2006).

2.2. *Defining corruption*

Historically, the concept of corruption has been subject to many different definitions (Gaskins, 2013). The most commonly adopted definition has been “the abuse of public power for private gain” (Bardhan, 1997; Heidenheim, 1988). This definition, despite shedding light on what corruption entails, does not specify intentionality. Across the literature, intentionality (and consequent intent to engage in acts of misconduct) can either be construed as rational or irrational actions (Palmer, 2012). For the purposes of this study, however, corruption is established as:

“behavior perpetrated by organizational officials (i.e. directors, managers, and/or employees) in the course of fulfilling their organizational roles that is judged by social control agents (i.e. prosecutors, regulatory agency officials, judges, journalists etc.) to be illegal, unethical, or socially irresponsible. Collective organizational wrongdoing involves the sustained coordination of multiple organizational participants.” (Palmer, 2008, p. 108)

Adopting such a definition moves us beyond the discussions of the psychological mechanisms underlying the intention to commit or not commit acts of corruption, but instead grounds it on the basis of how actions could be perceived by members of the organization, control agents, and members of the civic society (i.e., is it legal? Is it immoral or unethical? Is it irresponsible?). In this regard, Nye (1967) determines that corruption is usually behavior which “deviates” from formal attributions and responsibilities.

Consequently, intentionality, though important to classify and understand, becomes secondary. More important is comprehending whether the act committed can be perceived as corrupt, specially under the

watchful eye of control agents and citizens for whom transparency mechanisms have primarily been created. Regarding the production of open data specifically, the idea that it can be corrupted, thus causing a breakdown in the data flow and consequently compromising data quality, seems plausible.

There is some discussion in the literature as to whether increased discretionary power allowing public officials to deviate from formal rules may result in corruption or not. While Fazekas et al. (2021) argue that it may result in higher levels of corruption risks, Best et al. (2017) found that it may translate into efficiency gains. It seems that results are therefore mixed in this regard and are case dependent, “depending on where inefficiencies and corruption risks lie” (Cocciolo et al., 2021). Each context should be investigated on a case-by-case basis to determine whether deviating from formal procedures can be constituted as corruption. Any such deviations could result in breakdowns which might compromise data quality. Following Palmer’s (2008) definition, we are guided as to how such actions might be perceived by control agents in terms of legality and morality.

Another aspect of the definition established by Palmer (2008) that is worth highlighting is the fact that it clarifies that corruption is typically the product of a sustained and coordinated collective effort by members of the organization. This was particularly evident in the case of the data collected for the case study which is the object of this research (as will be reviewed in subsequent sections). Moreover, it is reflective of what is established in DCog, in that the focus is on the collective efforts of agents.

2.3. Data quality

There seems to be little consensus as to how exactly to determine the quality of data in general, and more specifically, open data (Vetrò et al., 2016). Data quality has been typically defined as “fitness for use,” though this definition is problematic since a dataset may be fit for purpose but low in terms of its quality, and vice versa (Mocnik et al., 2017). Thus, it seems more reasonable to evaluate the quality of datasets on the basis of their internal attributes (e.g. timeliness, completeness, etc.). Under this prism, data quality encompasses several dimensions (or internal attributes), with different scholars emphasizing different dimensions. In this regard, Bovee et al. (2003) propose a framework for evaluating data quality on the basis of four dimensions: integrity, accessibility, interpretability, and relevance. As they explain, to determine the quality of data we must (1) believe it to be free from defects (integrity) and is comprised of four subattributes: accuracy, completeness, consistency, and existence. In addition, we must (2) be able to obtain information that we might find useful (accessibility), (3) understand it and find meaning in it (interpretability), and (4) find it applicable to our domain and purpose of interest in a given context (relevance).

Most research on data quality as applied to open data has been concerned with the concept of completeness (Cunha et al., 2011), accessibility (Gant and Gant, 2002; Esteves and Joseph, 2008; Lourenço, 2015; Ruijter et al., 2017), and interpretability (Lausen et al., 2005; Palmirani et al., 2014). The subdimension of accuracy, however, which is integral to ensuring the integrity of data, has often been overlooked. One reason for this omission is the difficulty in measuring and establishing a benchmark (or framework) for accuracy, in addition to the belief that technology is transformative when targeting corruption (Kling et al., 2005; Fox and Haight, 2010; Adam and Fazekas, 2021). However, as Adam and Fazekas (2021) observe, current research largely ignores the different responses to the deployment of technologies aimed at curbing corruption which they suggest may even involve distorting the use of the tools and corrupting them. In the instance of use of open data as an anticorruption mechanism, findings arrived at here suggest that corrupt government officials may intentionally misrepresent data to conceal misconduct and evade detection. This poses a problem from a data analysis perspective, because if the data inputted is incorrect, any analyses will be rendered inaccurate and incapable of correctly detecting corruption. Ensuring accuracy and ascertaining the value of the analyses, rendered as a result of it, seem therefore imperative for the success of open data as an effective anticorruption measure (and establishing that effective investments in this area are made).

Due to the complexity of government, attempting to identify when data are misrepresented at a macro-level can be challenging. Thus, we suggest that it be looked at from a micro-level perspective of analysis,

closely monitoring how data are produced. By adopting a micro-level view, we then turn our focus to the public officials who are at the forefront of data creation, the dynamics between them and the technologies deployed to create open and transparent environments, looking for instances that may create opportunities for data quality to be compromised. These opportunities for data to be compromised are what we define as “breakdowns” in the information flow.

Sharp and Robinson’s (2006, p. 4) definition of breakdowns seems particularly well suited in this instance. In DCog terminology, which is the theoretical framework adopted in this study, they are considered “potential failures in communication or information flows that will impair the system’s performance or prevent the system from achieving its goals.” Thus, understanding the ways this takes place could lead to insights as to why corruption persists even after the deployment of open data mechanisms. Where corruption is present, the actions of individuals deviating from their regular work practices may constitute attempts to conceal their illicit activities and would therefore be the cause of breakdowns in the data flow.

3. Distributed Cognition

In order to study how government officials create data, this study adopted DCog as a theoretical framework. Devised by Edwin Hutchins in the 1990s, DCog emphasizes the role of cognition, defined as “to be those that are involved in memory, decision making, inference, reasoning, learning, and so on” (Hutchins, 2000, p. 1). For Hutchins and DCog, however, these cognitive processes are not constrained to an individual mind, but rather are extended and expanded conceptually to encompass both the social and cultural contexts involved in those processes.

DCog has been applied to a number of different settings: Hutchins (1995) pioneered the model by applying it to U.S. Navy’s navigation system, noting that the navigation involved numerous sailors working with multiple artifacts to enable the safe movement of the ships. Similarly, Rogers (1992) studied engineer practice and Halverson (1995) applied it to air traffic control. It is also particularly suited for workplace studies (Hollan et al., 2000). Such disparate settings reflect its wide-ranging application. Thus, though it has never been applied in the context of a governmental agency, its diverse ranging applicability, coupled with its interest in “work systems,” makes DCog a suitable approach to this study.

As a theoretical framework, DCog is focused on studying the cognitive distribution of work activities between individuals and artifacts as they interact in a particular environment. This “cognitive system” comprises both human agents and artifacts (technical systems, paper files, etc.). It is therefore a “distributed” cognitive system since knowledge is not restricted only to a human mind but rather expanded to include the role of artifacts as well. In other words, both human agents and artifacts work in tandem, to guide ships, manage air traffic control or, in our case, produce data. Each element of the distributed process deals with a part of the process, and each reinforces and restricts others in different ways. The focus in DCog is therefore in the interplay between agents, and how cognitive processes are shared between them. In particular, distributed cognitive theory establishes two important principles (Hutchins, 1995; Hollan et al., 2000):

- The unit of analysis should be the cognitive system. A cognitive system arises out of processes where functional relationships of different elements are present (Hollan et al., 2000). That is, a group of people and artifacts who do not interact do not constitute a cognitive system (Rogers and Ellis, 1994).
- Cognitive processes may involve various processes of coordination, such as between members of a social group and between people and artifacts. Additionally, most coordination has a temporal component that needs to be taken into consideration.

Such disparate settings reflect its wide-ranging application. Though it has never been applied in the context of a governmental agency, its diverse ranging applicability, coupled with its interest in “work systems,” makes DCog a suitable approach to this study. In DCog, cognition is distributed across agents

who actively engage in a shared space, working in tandem to produce an outcome (e.g. the sailors and instruments guiding a ship into a harbor (Hutchins, 1995). Thus, the focus is on understanding how cognitive processes are distributed across each of the components, individuals, and technical artifacts. This definition also represents the process of creating transparency data in the case study site, which consists of civil servants working with each other and the technical systems deployed, the output of which will be data publicized through a technical platform so that civil society can access it and interact with it. The interplay between human agents and artifacts within a governmental agency, whereby cognitive systems are shared and through which data propagate, thus constitutes the unit of analysis in this study.

Understanding how cognitive processes are distributed across each of the components is key in DCog. Doing so can provide meaningful insights as to how the system functions and how data are produced and will aid in identifying breakdowns in the process. Such breakdowns, which can be conceptualized as workarounds or deviations from norms, can directly impact the flow of data, the sequence of tasks, and quality of data produced (Sharp and Robinson, 2006; Galliers et al., 2007). These “breakdowns” become instances in which the system allows the researcher to clearly identify issues such as “what happens when the information flow breaks down or when alternative ways of handling the information flow emerge in the system” (Lindblom and Thorvald, 2017, p. 66).

One criticism directed toward the adoption of DCog relates to its abstractedness (Berndt et al., 2015, p. 432). To counter this, methods that facilitate the application of DCog in practice and aid in articulating DCog principles have been developed, such as the distributed cognition for teamwork (DiCoT) framework by Furniss and Blandford (2006), which was adopted here for data analysis.

DiCoT consists of a semi-structured way of applying the DCog theoretical framework to research and builds on DCog literature in order to build models which reflect different aspects of the “complex computational system.” Each model has DCog principles associated with it, which act as focal points for the researcher, guiding them through the analysis. The five models are physical layout, artifacts, information flow, social structures, and evolutionary. It is important to note that the DiCoT framework is flexible and does not require that all models be adopted for analysis (Furniss and Blandford, 2006). Hence, in this study, focus was paid to the artifact and information flow models to aid in determining breakdowns of data and the sources of it.

4. Case Study Site and Methods

DCog establishes that the unit of analysis should be the assemblage that produces the data. To this extent, Liu et al. (2008, p. 2) affirm that “the unit of analysis should not be a human individual, but a cognitive system composed of individuals and the artefacts they use to accomplish a task.” As is described below, these artifacts include both the information systems and paper-based (physical) files.

To study this phenomenon, a case study of a governmental agency situated within the Brazilian federal government was conducted, hereby referred to as Agency X (name disguised for confidentiality issues). Agency X was selected due to: (1) authors had good access to the organization; (2) being a work system whose main function is producing data that will later be made public on Brazil’s Transparency Portal; and (3) presenting signs that ongoing incidents of corruption were present (this determination was made while investigating suitable cases for the study). Through the course of this research, some of the corrupt practices identified at Agency X include fraudulent bidding processes in order to favor certain suppliers over others, kickbacks for “middlemen” acting as consultants and intermediating contact between Agency X and suppliers, contracting of services provided by civil servants who work at Agency X (a practice prohibited by law so as to avoid a conflict of interest), and skipping stages in the procurement process as established by law and formal procedure, resulting in negative financial planning and overspending.

According to Merriam (2002, p. 12), “there are three major sources of data for a qualitative study—interviews, observations, and documents.” In the instance of this case study, the data collection methods used were semi-structured interviews, participant observation, and document analysis.

A total of 45 interviews were conducted over two rounds: the first round, with 28 participants, took place between February and April of 2015. The second round, a follow-up round with 17 participants,

took place in February 2016. This second round had the aim to ensure no gaps in information gathering remained, and to identify whether any additional events which might be relevant to this study had taken place in the year since the first round of interviews had taken place.

All interviews were conducted with civil servants at Agency X. The majority of interviewees worked directly with the information system, as part of the procurement process; others were indirectly involved (upper management and technical support). All interviews were conducted in Portuguese and on-site at Agency X. Snowball sampling (Merriam, 1998) was adopted in order to identify those agency workers who should be interviewed.

In addition to interviews, time was spent on-site at Agency X for four afternoons during the period between February and April of 2016. This opportunity allowed for observation of interactions between those working in that environment, crucial for aiding in mapping the DiCoT models adopted for the analysis of data (artifact and information flow models). It also allowed for the opportunity to have access to the systems in question and directly observe how they operate. Complementing this stage of data collection, internal documents, which consisted of internal manuals and guidelines, Power Point presentations, screenshots of technical systems, and information reproduced from both their external website and intranet, were also analyzed. Documents were selected on the basis of the information they contained to further complement the other two sources of data collection used (interviews and observations) and ensure a full picture of the DiCoT models adopted.

Agency X's main function is to execute and manage procurement processes, fielding purchase requests of supplies and services from various government agencies which it supports (all within Ministry X, the organization to which it is subjugated). Its main responsibilities include the management of all activities related to procurement: conducting the bidding process, placing orders, supply chain management, budget execution, and payments. Agency X is a relatively small organization, comprising of under 100 civil servants. The procurement processes it handles are, for the most part, small in terms of monetary values (specially in comparison to processes handled and managed across the entire Federal Government sphere), which is reflective of the support role it holds within the wider ministerial organization.

All of its activities are recorded on SIAFI (a technical system owned and managed by Brazil's Secretary of National Treasury [STN] and deployed across all federal government agencies in 2015), which feeds data into the Brazilian Transparency Portal (which holds all public spending data). As such, SIAFI is the main management tool adopted by STN and is at the forefront of data creation, all of its data being published and open to public scrutiny. In addition to SIAFI, Agency X also utilizes another internal technical system (referred to here as "IntSys"), implemented in 2011 to attend to changes to the legislative framework around freedom of information¹.

5. DCog in Agency X

The focus of this study is on data creation and what processes facilitate this. In doing so, it then becomes possible to identify breakdowns in the information flows since these have the potential of undermining the system's effectiveness. To identify potential breakdowns, it is necessary to first understand how the system *should* work, then comparing it to how it *actually* works.

The procurement process at Agency X can be broken down into eight different stages and involves six different groups of people, across two different departments: purchasing and finance. As can be seen in Figure 1, the procurement process begins when a requester ("Z") has identified a need and places a request through the internal system (IntSys; Stage I). This requester will be located at either Agency X or one of the agencies interlinked through the IntSys within Ministry X. Stage II will then consist of team member "A" fielding this request and redirecting it to "B," a team member who will be responsible for that specific product category.

¹ In 2011, Law No. 12,527, dubbed the Access to Information Law, was enacted expanding requirements of what information government agencies were required to publish and make available to the public via the Transparency Portal.

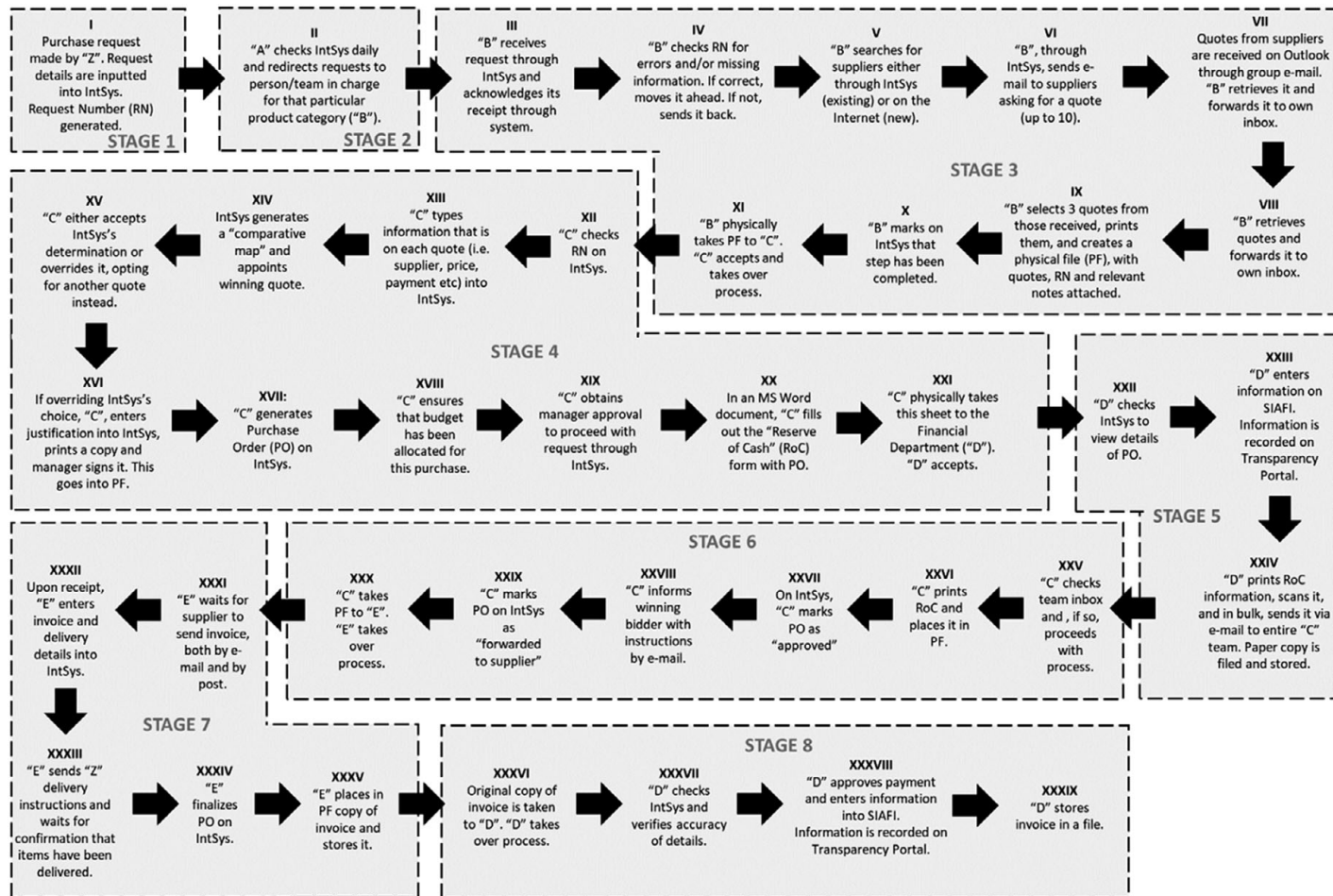


Figure 1. Description of the procurement process at Agency X.

Team member “B” will then select suppliers to contact and request quotes from (Stage III). Up to 10 suppliers may be contacted, and the process will be moved along to team member “C” once at least three quotes have been received (the bidding process—part 1). Stage IV then consists of “C” choosing the winning bid (bidding process—part 2) and handing this process to the financial department (“D”) so that they may ensure that budget has been allocated and reserve that amount for that particular purchase (“Reserve of Cash,” Stage V). Team member “D” will then hand the process back to “C” so that they may inform the supplier who has provided the winning bid and obtain their acknowledgement (Stage VI). Once that has been received, the process will then move along to team member “E” who will ensure that supplies are delivered and will process the invoice (Stage VII) so that the financial department can once again take over the process in order to remit payment (Stage VIII). The artifacts and data flows in this distributed process are described in more detail below (see also [Figures 1 and 2](#)).

As highlighted by DCog as well as involving a number of different roles, the procurement process is enabled via several artifacts. Although IntSys provides the official computerized record of the procurement, some activities will be handled via e-mail and telephone. In addition, a physical file is created for each procurement containing copies of documents and exchanges. This physical file moves across teams as the procurement process progresses. Finally, SIAFI is used by the financial department to record the Reserve of Cash and payment instances and relevant data pertaining to it. The SIAFI data are then made available via the Transparency Portal.

Through the DiCoT approach, data were analyzed in a structured way and coded using DCog principles identified by Furniss and Blandford (2006). As noted above, the focus was on two of their five models: artifact and information flow. On the basis of these two models, a picture emerges as to how the distributed cognitive system operates at Agency X, when and how breakdowns in process can take place, and the effect this bears on data quality.

5.1. Artifact model

Furniss and Blandford (2010, p. 2) state that “the artefact model concerns itself with the artefacts, representations, and tools that are used to store, transform and communicate information.” In the case of Agency X the main artifacts are used which make up the DCog and aid in the process of data creation are IT-based artifacts and paper-based ones (see [Table 1](#) for an overview).

The main artifact is the internal system (IntSys), used by Agency X to process the purchase request. The initial request placed by “Z” is recorded in IntSys and is the trigger for the rest of the procurement process. All stages of the procurement process are registered here. The other main technical system, SIAFI, is only used at two stages of the procurement process: Reserve of Cash and Payment. Users in the procurement process do not have access to this system since it is only used by finance department personnel.

The physical file is another artifact. It serves as the trigger between stages of the process since it is the act of the file being passed on to the next agent that alerts that person that the process has evolved to their station and alerts them to the need to act on it. IntSys does not have the functionality built into it to alert agents; only the physical file has that innate property. Thus, its role in propagating data, aiding in communication, and coordination of tasks is significant. Additionally, as it reflects the data that are stored in the system, it also acts as a back-up source of data on the procurement and is what Hutchins (1995) would refer to as the principle of distributed memory, that is, an artifact that serves the purpose of redundancy.

Data in the physical file, in fact, go beyond what is contained on the IntSys. For example, all e-mails exchanged in reference to the procurement process should be printed out and stored in this file; the IntSys on the other hand will not have this level of detailed data since there is nowhere on the system to attach or store this kind of data.

This was noted in the interviews:

Int19: For us, it starts with the technology. The request come via the IntSys. But the IntSys works more like a register. The process actually we realize the process is “walking” when the folder

Table 1. Summary of artifacts

	Artifacts	Description
IT-based artifacts	Internal system (IntSys)	IntSys constitutes the main technical system used by those involved in the procurement process. It registers several instances of information that are pertinent to this process. At each completion stage of the procurement process, this is recorded on IntSys. This serves two purposes: ensuring the process is moving forward and record who owns the process at each particular stage
	SIAFI	SIAFI is the main interface between the procurement process and the Transparency Portal. It serves the purpose of centralizing all budgetary execution information for entire federal government. It is managed solely by the financial department
	MS Word	Used to record supporting documentation, such as the “Reserve of Cash” form
	E-mail	Used to communicate with suppliers and, occasionally, with “Z” (purchase requester)
	Telephone	Used to communicate with suppliers and “Z” (purchase requester)
	Printer	Used to print documentation that needs to be inserted into paper files
Paper-based artifacts	Paper files	Paper files should reflect the process that is recorded on the system. Also serves the purpose of triggering the next stage in the procurement process, once it is delivered to the next person in line
	Paper notes	Used to set reminders or any relevant notes that should be added to paper file
	Invoice	Invoice is required in order to finalize the procurement process on IntSys. It will then be used by financial department to make payment and record information on SIAFI

moves. IntSys is more of a register. It doesn't lead the process. It's the folder that tells us that we have to do something about it. It's obvious, if someone wants to check processes on the IntSys, they can. But we can only ever do anything about it when we receive the paper.

Thus, the physical file reflects a greater source of history of the processes undertaken and is integral to the completion of the procurement. It does not, due to its format, automatically communicate with the other artifacts. Importantly, what gets stored in it (or not) is led by the people involved and not the processes by which Agency X is bound. This creates opportunity for agents to circumvent steps in the process and omit or conceal data which in turn results in breakdowns.

Regarding the role of SIAFI, it is important to note that this is the only system that communicates with the Transparency Portal. Its use, however, as noted previously, is restricted to the finance department personnel. Thus, the only stages in the procurement process which are communicated with the external audience (i.e., civil society) are the two stages handled by this department, who do not have a view of the entire procurement process other than what is inputted into IntSys; they do not handle the physical file.

This is important to note since due to constraints of the way IntSys is built and its innate properties, this means that not all procurement information is necessarily contained there. This potentially raises the question as to how well is the system redesigned to capture all data, and how transparent it really is in terms of communicating with members of the public. This also helps us map how data resulting from these processes are stored and communicated and where opportunities for breakdowns may occur. For this to be clearer, we combine this model with the information flow model, to be reviewed in the next section.

5.2. Information flow model

Berndt et al. (2015, p. 433) affirm that the information flow model in DCog concerns itself with “how information moves and is transformed in the system (and any obstacles to effective information flow) —both formally and informally; whether any agent acts as an information hub ... and how interruptions are managed.” In other words, it describes the data flows between the agents of a system, the different communication channels utilized by them, and key flow properties. As such, the main communication channels should be identified and discussed. Table 2 presents an overview of the different agents involved in the procurement process and what their roles are as identified as a result of the data gathered.

Communication between these different agents takes place through a range of different means, either via the IntSys, via e-mail or telephone, or face-to-face. The communication flows between agents are depicted in Figure 2. As can be observed, communication between agents takes place via the IntSys in points 1, 2, and 3 only. From that moment onwards, communication then takes place via either e-mail, telephone, or face-to-face, possibly because of the enhanced communication “bandwidth” they provide. Although the procurement process will be registered on IntSys throughout each of the procurement stages,

Table 2. *Summary of agents involved in the procurement process*

Agent	Role
“Z”	Responsible for placing request on IntSys. Will be the agent responsible for triggering the entire procurement process
“A”	Responsible for checking requests on IntSys and redirecting it to the person/team in charge for that particular product category (“B”) through IntSys
“B”	Responsible for identifying suppliers who can meet Z’s requirement. Will make contact with suppliers and request quotes. “B” will also be the one to start a physical file which will move across agents of the process
“C”	Will select the winning bid either by using IntSys to appoint winner or manually doing so. Will also be the one to handle the first point of contact with the financial team (“Reserve of Cash”)
“D”	Financial team who will be involved twice in the procurement process: (a) in the Reserve of Cash stage, which will allow the purchase to be confirmed; (b) in the payment stage, when process has been finalized. Responsible for entering information on SIAFI, the system which communicates with the Transparency Portal
“E”	Responsible for finalizing the procurement process. Will confirm delivery and process invoice, before handing off process to the finance team for payment
Managers	Every team (Z, A, B, C, D, E) has a member of upper management overseeing their activities. They may intervene in the process at any time
Suppliers	Suppliers are initially contacted at Stage III of the procurement process. One supplier will be selected at Stage IV when they will be expected to acknowledge request and supply demand. They will later be in contact with “E” during Stage VII when they will be expected to invoice Agency X for their services

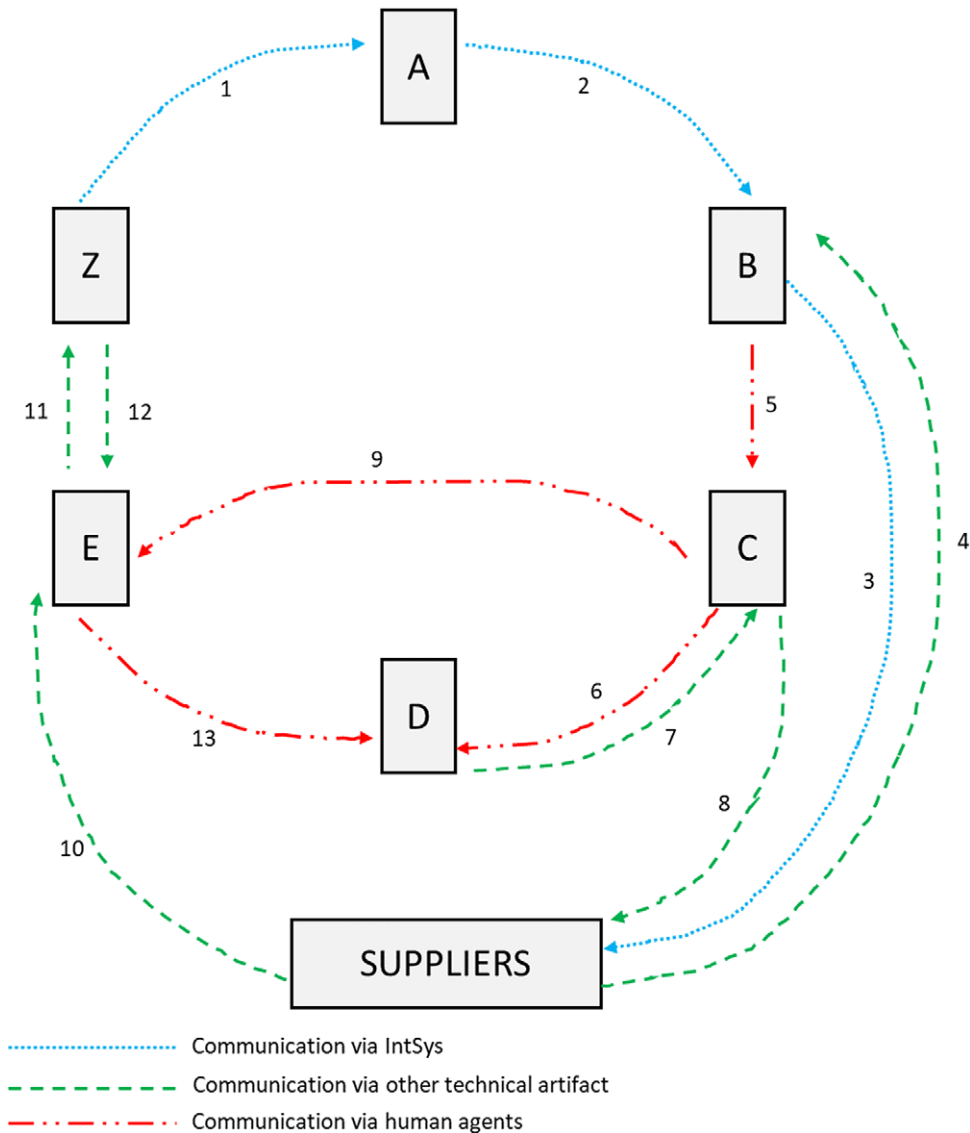


Figure 2. Communication flows.

the actual communication and coordination between the different participants take place via other means and, internally, between teams B, C, D, and E, it will be face to face.

This reflects the fact that the process of data creation for the transparency portal is led by the human components of this system. It also reflects the fact that many of the interactions inherent to this process will not be formally recorded. On the contrary, it can be determined that many of the exchanges and coordination among teams occur in the pocket of informality, an action that is encouraged by the fact that stages in the process are only triggered when an artifact—the physical file—is physically moved from one team to another by the human agents (who can therefore be conceived as the triggers of data flows).

How data flow through a system is of particular interest in DCog and what can be observed at Agency X is that rarely do data flow exclusively due to the action of technical artifacts. Informal communication

can on the one hand have a positive role by facilitating processes. On the other hand, it can also have a perverse effect by allowing for processes and storing of data to be circumvented. As was noted in interviews:

Int19v2: “It does happen that we pay in advance. Without the invoice (...). We print the cover sheet off the IntSys and we take it to the finance department. And then later we’ll go back and register everything when everything does get delivered. Then it’s like... my manager will call up the finance manager. And they’ll sort it out between them. I’ll just fill out the details and take the paper to them.

Note that, as reported, though the steps in the procurement process were not observed, this was later registered as it had been, which compromises data integrity and is a clear example of a breakdown.

The high degree of informality also allows for human agents to dictate how the procurement process should be followed. For example, when asked about the process of selecting suppliers, one interviewee reported that:

Int18: The rule is the IntSys makes a few suggestions. The rest is up to each person, her way of working, what she finds on the Internet. We have to ask quotes to up to 10 suppliers. We then send it to the next stage when we have received at least three quotes back. So really, in fact, it’s up to each one’s criteria.

When combined, both the artifact and the information flow models provide clear instances of how breakdowns in the process can occur. The overreliance on physical artifacts, the lack of communication between artifacts and the resulting high levels of informality, results in opportunities for processes to be circumvented.

6. Findings

Data processing in this distributed cognitive system reveals itself through the propagation of representational states across different artifacts. In other words, the input of a purchase request into the IntSys triggers the entire system into action, and data will then be processed as it moves through the entire cognitive system through the different media, that is, the technical systems, other artifacts such as paper, and finally, through the people themselves. In analyzing how data are produced at Agency X and how data move through this system through the lenses of the two DiCoT models (artifact and information flow) revealed the first and important characteristic: the fundamental role individuals play in the cognitive system.

As the data analyzed showed, though the data stored and registered in the technical system are constrained by its inherent design (modularity) and its imbued properties, it is ultimately up to the human agents to move the data forward and dictate what data get stored and, significantly, how it gets stored into technical systems. Thus, IntSys serves the main purpose of registering steps of the procurement process, but it does not lead the process; it does not even have automatic reminders and triggers built into the system. This means that everything is manually processed, and individuals have to keep going back into IntSys to register data, input data that are not automatically populated into it. Moreover, since technical artifacts do not communicate directly with one another, humans act as both hub and behavioral triggers in the sequential control of action.

The core interactions are therefore between individuals, through language, revealing the high prevalence of informal communication that takes place at Agency X. This means that individuals act as mediating structures, communicating directly with one another, and potentially bypassing technical systems.

Data move and transform through the system through the action of individuals. With the exception of data that are automatically fed from SIAFI to the Transparency Portal, nothing moves automatically without the direct involvement of a human agent. Thus, the main behavioral trigger within this system is the people themselves, who will notify the next team in the chain of tasks that the process has evolved via

the delivery of the physical file to them. Paper, however, is a non-intelligent artifact. Alone, it accomplishes very little in relation to transparency, thereby once again reinforcing the dominant role individuals play in the information system. Thus, human agents act as more than hubs and behavioral triggers; they are also the largest retainers of knowledge as well.

In summary, cognitive processes are unevenly distributed between technical and human agents. Human agents are the ones leading the data flow, determining how data are created and stored. The technical component enforces certain processing rules, dictating how data are registered on its systems (according to its imbued properties), but not what gets stored, or even when. This temporal aspect is completely controlled by participants.

The uneven distribution of cognitive processes between the components of the system is noteworthy since it allows for instances of breakdowns in the process to occur. As discussed in [Section 2.3](#), breakdowns are significant since they undermine the system's effectiveness or prevent it from achieving its desired goal. In the case of this cognitive system, the goal is to produce open data which can be monitored and is subject to scrutiny by members of civil society, making this a more transparent and accountable system, which in turn would serve as a deterrent for public officials to engage in corruption. Previously, we posited that one of the reasons for choosing Agency X as the *locus* for this case study was due to the fact that ongoing cases of corruption had been identified there. As noted, some of the instances observed were fraudulent bidding processes that favored certain suppliers over others, kickbacks for "middlemen," and contracting of services provided by civil servants who work at Agency X (a practice which is illegal).

In [Section 2.2](#), we defined that corruption is characterized by a sustained and coordinated efforts by multiple individuals; it is therefore the result of a collective action. As [Figure 1](#) demonstrates, the procurement process at Agency X comprises several different stages and involves multiple teams. This could lead one to erroneously conclude that corruption could not take place here since corruption is often assumed to be an obscure act, shrouded in secrecy. However, what was observed at Agency X was the opposite of that and confirms Palmer's (2008) interpretation of how corruption takes place.

As DCog suggests, cognition is not constrained to one individual mind but is rather distributed between different agents, both human and technical. Thus, it could only be through collective effort that corruption would be able to take place here. The instances of corruption at Agency X were made possible due to the opportunities for breakdowns in the process that exist in the system. Though the deployment of technical systems envisioned a more transparent environment which transform this system and make it more accountable and hence less corrupt, that was not the case in practice. As Adam and Fazekas (2021) observed, this is an instance whereby the technology itself facilitates corruption. With its limited imbued properties, IntSys allows for stages in the process of data creation to be concealed. The high levels of informality observed at Agency X, coupled with the leading role human agents play, resulted in individuals collectively working together to undermine the effectiveness of this system to detect acts of corruption. Thus, compromised and incorrect data were stored in the technical system, so that its outputs did not portray reality.

As a result of combining the artifact and information flow models, examples of breakdown identified at Agency X were:

- The creation of a fake need, resulting in a series of fabricated pieces of data being registered and stored on the system. An example is that of a "double contract," that is, two contracts with the same supplier, for the same product category, were signed, with only slightly different terms. Both carried a minimum order amount that had to be paid yearly. The fact that both contracts were in fact for the same product was concealed when data were entered into the system.
- Certain suppliers being intentionally favored over others, meaning that the data registered on the system did not correspond to reality. The request will however retain the appearance of a regular bidding process on the technical system. This could either be to favor suppliers who offered kickbacks or civil servants who were supplying their own services across Agency X. Instances that would reveal that the bidding process was being misused were either misrepresented or concealed.

- Suppliers may be intentionally favored over others, but under the false pretence that it falls into certain legislative exceptions for bidding processes. Data in these instances were misrepresented.
- Changing documents to conceal irregular requests. Examples include altered invoices that conceal the exact contents of the purchase. The actual invoice is not registered on the technical artifacts, only on paper. Data in this instance are thus misrepresented and partially concealed.
- Skipping steps in the procurement process in the name of speediness or to favor certain suppliers, such as payment without an invoice. The process is then conducted “out of the system” (as described in interviews), and only later registered, after it has been completed. Therefore, data registered on the system did not accurately represent reality.

The instances identified above can be largely attributed to unethical decision making and were the result of a collective and concerted effort facilitated by how the process is designed. This is in line with what was found by Fazekas et al. (2021) whereby increased discretionary power results in increased corruption risks. That is not to say that all individuals benefitted from the acts of corruption committed at Agency X, but there was certainly a collective awareness. It must also be noted that instances of accidental breakdowns also took place and were observed. For example, when individuals mistakenly failed to register on IntSys when a stage in the procurement process had been completed.

Regardless, in both instances, be it intentional or accidental, data quality is compromised and the potential for civil society to act and scrutinize, the basis for which open data would be effective in curbing government corruption is limited in scope. The findings obtained in this study confirmed that this work system, responsible for creating data, can indeed be framed as a distributed cognitive system, making DCog a suitable theoretical framework for this type of study. Framing it in this manner and adopting DiCoT allows us to map the distribution between agents and identify the opportunities for breakdowns which may take place.

7. Conclusion

This research has analyzed how a particular form of data is produced to evaluate the possibilities for its quality to be compromised, thus undermining the anticorruption capabilities. Findings suggest that how this system is designed facilitates the perpetuation of corruption, creating opportunities for participants to circumvent detection by concealing and misrepresenting data produced. As discussed, human agents lead data creation and flow through this system, not the technical artifacts deployed with the intent of enhancing levels of accountability.

Important to note that the technical systems deployed and implemented—SIAFI and IntSys—were in response to legislative reforms enacted by the Brazilian federal government which had the goal of expanding transparency of data and making government officials more accountable, thereby reducing level of corruption. However, as can be observed from the data collected, this ideal was not achieved. The role human agents play in this distributed cognitive system results in breakdowns in the data flow, both intentional and accidental. In both instances, data accuracy, a sub-attribute of data integrity, a dimension key to ensuring the quality of data (as previously reviewed) is compromised rendering analyses of such data inaccurate. This hinders actions that can be taken by civil society at large in their pursuit to hold government to account and compromises the potential of data transparency to act a mechanism for anticorruption.

Given that data inaccuracies prove difficult to detect when macro-analyzing the data, studying these processes at a micro-level, as was done here, is useful toward detecting when breakdowns in data creation occur and may provide a path toward establishing a benchmark for ensuring the quality of data and open data as a meaningful mechanism in curbing corruption. The fact that this process can be framed as a distributed cognitive system suggests that this could serve as a useful framework for future studies. The application of DCog to governmental settings is novel but has proven effective in painting a clear picture in how the system functions and where breakdowns may occur.

Typically, the main obstacle toward adopting DCog as a framework is its excessive level of abstractness (Berndt et al., 2015). However, adopting the DiCoT model as a methodological tool facilitates this, guiding researchers' attention, aiding in the gathering and analysis of data, and allowing others to replicate the study more easily across other settings. Further research in this regard is suggested to extend and validate the use of DCog across other national contexts. Even in cases where no cases of corruption are suspected, adopting DCog as a framework might be useful to aid in determining where technology deployments are working as intended. This could be especially useful given the limited number of qualitative studies in this area (as previously discussed).

It is important that lessons be learned from this case. This case suggests that corruption is the result of a collective effort. The higher the level of informality within the system, the easier it is for individuals to coordinate mechanisms for circumvention. Findings suggest the need for reducing human touch points and more fully automating processes to minimize breakdowns in data creation and flow. This can prove to be challenging since it would require fully integrating the different technical artifacts (namely physical files and e-mails) into the technical systems and redefining how technical systems should operate (e.g., rearranging the distribution of tasks and building triggers into the system). Though there is some research to suggest that there are cases where increased discretionary power could lead to efficiency gains, in the case of Agency X, this instead resulted in more opportunities for acts of corruption to be concealed.

This study stopped short at analyzing the intent behind why individuals engage in corruption. Further investigation regarding intentionality might provide further insight into this phenomenon and could also aid in elucidating why, differently from expected, corruption is not covert and is instead the result of a collective effort. Though there has been some research into the psychology behind corrupt intent (e.g. Palmer, 2008), more research is required to fully comprehend it. It could also further elucidate why corruption continues to take place even when systems designed to more easily detect it are implemented.

It could be questioned how transparent systems really are given that civil society is only given insight into a few instances of the data creation process since not all steps of the process are publicly registered on the Transparency Portal. Further investigation into this might prove useful. Policymakers should take from this example lessons on how to build more effective and transparent procurement systems that effectively reduce opportunities for breakdowns. As noted previously, there has been a great push for increased investments in open data as a form of combatting corruption; ensuring therefore that investments effectively translate into meaningful results is paramount.

Funding statement. This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior)

Competing interest. The authors declare no competing interests exist.

Author contribution. All authors approved the final submitted draft.

Data availability statement. Authors confirm that analysis for all data generated during this study are included in <https://etheses.lse.ac.uk/3882/>

References

- Adam I and Fazekas M** (2021) Are emerging technologies helping win the fight against corruption? A review of the state of evidence. *Information Economics and Policy* 57, 100950.
- Andersen TB** (2009) E-government as an anti-corruption strategy. *Information Economics and Policy* 21(3), 201–210.
- Arellano-Gault D and Lepore W** (2011) Transparency reforms in the public sector: Beyond the new economics of organization. *Organization Studies* 32(8), 1029–1050.
- Bannister F and Connolly R** (2011) The trouble with transparency: A critical review of openness in e-government. *Policy & Internet* 3(1), 158–187.
- Bardhan P** (1997) Corruption and development: A review of issues. *Journal of Economic Literature* 35(3), 1320–1346.

- Bauhr M and Grimes M** (2017) Transparency to curb corruption? Concepts, measures and empirical merit, crime. *Law and Social Change* 68(4), 431–458.
- Berndt E, Furniss D and Blandford A** (2015) Learning contextual inquiry and distributed cognition: A case study on technology use in anaesthesia, *cognition. Technology & Work* 17(3), 431–449.
- Berners-Lee T, Hall W, Hendler JA, O'Hara K, Shadbolt N and Weitzner DJ** (2006) A framework for web science. *Foundations and Trends in Web Science* 1(1), 1–130.
- Bertot JC, Jaeger PT and Grimes JM** (2010) Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly* 27(3), 264–271.
- Best MC, Hjort J and Szakonyi D** (2017) *Individuals and Organizations as Sources of State Effectiveness*, No. w23350. Cambridge, MA: National Bureau of Economic Research. Available at <http://www.nber.org/papers/w23350.pdf> (Accessed 11 April 2023).
- Bhatnagar S** (2003) E-government and access to information. In Transparency International (ed), *Global Corruption Report*. Washington, D.C.: Transparency International, pp. 24–28.
- Bovee M, Srivastava RP and Mak B** (2003) A conceptual framework and belief-function approach to assessing overall information quality. *International Journal of Intelligent Systems* 18(1), 51–74.
- Bukovansky M** (2006) The hollowness of anti-corruption discourse. *Review of International Political Economy* 13(2), 181–209.
- Calland R and Bentley K** (2013) The impact and effectiveness of transparency and accountability initiatives: Freedom of information. *Development Policy Review* 31, s69–s87.
- Cocciolo S, Samaddar S and Okiya S** (2021) What we've been reading about rules and discretion in public procurement, *World Bank*. Available at <https://blogs.worldbank.org/governance/what-weve-been-reading-about-rules-and-discretion-public-procurement> (Accessed 11 April 2023).
- Cranefield J, Robertson O and Oliver G** (2014) Value in the mash: Exploring the benefits, barriers and enablers of open data apps. In *Presented at the European Conference on Information Systems*, Vol. 16. Israel: Association for Information Systems.
- Cunha MAVCda, Frega JR and Lemos IS** (2011) Portais de serviços publicos de informacao ao cidadao no Brasil: Uma descricao do perfil do visitante. *Revista Eletrônica de Sistemas de Informação* 10(1), 1–19.
- David-Barrett E and Okamura K** (2016) Norm diffusion and reputation: The rise of the extractive industries transparency initiative. *Governance* 29(2), 227–246.
- Davies T** (2012) Supporting open data use through active engagement. *Presented at the Using Open Data Workshop*, Brussels, Belgium: W3C. Available at https://www.w3.org/2012/06/pmod/pmod2012_submission_5.pdf (Accessed 11 April 2023).
- Dawes SS, Vidasova L and Parkhimovich O** (2016) Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly* 33(1), 15–27.
- De Simone E, Gaeta GL and Mourão PR** (2017) The impact of fiscal transparency on corruption: An empirical analysis based on longitudinal data, *the B.E. Journal of Economic Analysis & Policy* 17(4), 1–17.
- Esteves J and Joseph RC** (2008) A comprehensive framework for the assessment of eGovernment projects. *Government Information Quarterly* 25(1), 118–132.
- Fazekas M, Nishchal S and Søreide T** (2021) Public procurement under and after emergencies. In Bandiera O, Bosio E and Spagnolo G (eds), *Procurement in Focus Rules, Discretion, and Emergencies*. London, UK: CEPR Press, pp. 33–42. Available at <https://www.govtransparency.eu/wp-content/uploads/2022/01/Procurement-in-Focus.pdf> (Accessed 11 April 2023).
- Fox J** (2007) The uncertain relationship between transparency and accountability. *Development in Practice* 17(4–5), 663–671.
- Fox J and Haight L** (2010) Transparency reforms: Theory and practice. In Peschard J and Selee A (eds), *Mexico's Democratic Challenges: Politics, Government, and Society*. Stanford: Stanford University Press, pp. 135–161.
- Frølich N** (2011) Multi-layered accountability. Performance-based funding of universities. *Public Administration* 89(3), 840–859.
- Furniss D and Blandford A** (2006) Understanding emergency medical dispatch in terms of distributed cognition: A case study. *Ergonomics* 49(12–13), 1174–1203.
- Furniss D and Blandford A** (2010) DiCoT modeling: From analysis to design. In *Presented at the ACM Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery.
- Galliers J, Wilson S and Fone J** (2007) A method for determining information flow breakdown in clinical systems. *International Journal of Medical Informatics* 76, S113–S121.
- Gant JP and Gant DB** (2002) Web portal functionality and state government e-service. In *Presented at the 35th Annual Hawaii International Conference on System Sciences*. Hawaii: IEEE, pp. 1627–1636.
- Gaskins L** (2013) The Effect of Information and Communications Technology (ICT) Diffusion on Corruption and Transparency (A Global Study). Texas A&M University. Available at <https://rio.tamui.edu/etds/19> (Accessed 11 April 2023).
- Gaventa J and McGee R** (2013) The impact of transparency and accountability initiatives. *Development Policy Review* 31, s3–s28.
- Halverson CA** (1995) Inside the cognitive workplace: New technology and air traffic control. Dissertation, Cognitive Science Department, University of California, San Diego.
- Heald D** (2012) Why is transparency about public expenditure so elusive? *International Review of Administrative Sciences* 78(1), 30–49.
- Heide M and Villeneuve J-P** (2020) From secrecy privilege to information management: A comparative analysis of classification reforms. *Government Information Quarterly* 37(4), 101500.
- Heidenheim AJ** (ed) (1988) *Political Corruption: A Handbook*, 2nd edn. London: Routledge.

- Hollan J, Hutchins E and Kirsh D** (2000) Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction* 7(2), 174–196.
- Hopkin J** (2002) States, markets and corruption: A review of some recent literature. *Review of International Political Economy* 9(3), 574–590.
- Hutchins E** (1995) *Cognition in the Wild*. Cambridge, MA: The MIT Press.
- Hutchins E** (2000) Ecological cognition and cognitive ecology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44(22), 566–569.
- IISD** (2019) UNESCO Finds 125 Countries Provide for Access to Information, *IISD*. Available at <https://sdg.iisd.org/news/unesco-finds-125-countries-provide-for-access-to-information/> (Accessed 11 April 2023).
- Kling R, Rosenbaum H and Sawyer S** (2005) *Understanding and Communicating Social Informatics: A Framework for Studying and Teaching the Human Contexts of Information and Communication Technologies*. Medford, N.J: Information Today Inc.
- Kossov N and Dykes V** (2017) Embracing digitalisation: How to use ICTs to strengthen anti-corruption, *GIZ*. Available at <https://www.giz.de/de/downloads/giz2018-eng ICT-to-strengthen-Anti-Corruption.pdf> (Accessed 11 April 2023).
- Lausen H, Ding Y, Stollberg M, Fensel D, Lara Hernández R and Han S** (2005) Semantic web portals: State-of-the-art survey. *Journal of Knowledge Management* 9(5), 40–49.
- Lindblom J and Thorvald P** (2017) Manufacturing in the wild – Viewing human-based assembly through the lens of distributed cognition. *Production & Manufacturing Research* 5(1), 57–80.
- Liu Z, Nersessian NJ and Stasko JT** (2008) Distributed cognition as a theoretical framework for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 14(6), 1173–1180.
- Lourenço RP** (2015) An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly* 32(3), 323–332.
- Margetts H** (2011) The internet and transparency. *The Political Quarterly* 82(4), 518–521.
- Meijer A** (2001) Accountability in an information age: Opportunities and risks for records management. *Archival Science* 1(4), 361–372.
- Merriam SB** (1998) *Qualitative Research and Case Study Applications in Education*, 2nd Edn. San Francisco: Jossey-Bass Publishers.
- Merriam SB** (2002) Introduction to qualitative research. In Merriam SB (ed), *Qualitative Research in Practice: Examples for Discussion and Analysis*. San Francisco: Jossey-Bass Publishers.
- Mocnik F-B, Fan H and Zipf A** (2017) Data quality and fitness for purpose, *Presented at the 20th AGILE Conference on Geographic Information Science*, Wageningen, Netherlands.
- Norris P** (2002) Editorial: The bridging and bonding role of online communities. *Press/Politics* 7(3), 3–13.
- Nye JS** (1967) Corruption and political development: A cost-benefit analysis, *The American Political Science Review* 61(2), 417–427.
- O’Neill O** (2006) Transparency and the ethics of communication. In Hood C and Heald D (eds), *Transparency: The Key to Better Governance? Proceedings of the British Academy*. Oxford: Oxford University Press.
- Palmer D** (2008) Extending the process model of collective corruption. *Research in Organizational Behavior* 28, 107–135.
- Palmer D** (2012) *Normal Organizational Wrongdoing: A Critical Analysis of Theories of Misconduct in and by Organizations*. Oxford: Oxford University Press.
- Palmirani M, Martoni M and Girardi D** (2014) Open government data beyond transparency. In *Presented at the International Conference on Electronic Government and the Information Systems Perspective*. Cham: Springer, pp. 275–291.
- Peixoto T** (2013) The uncertain relationship between open data and accountability: A response to Yu and Robinson’s “the new ambiguity of open government. *UCLA Law Review* 60, 200–213.
- Pina V, Torres L and Acerete B** (2007) Are ICTs promoting government accountability? A comparative analysis of e-governance developments in 19 OECD countries. *Critical Perspectives on Accounting* 18(5), 583–602.
- ReSPA** (2013) Abuse of Information Technology (IT) for corruption, *Regional School of Public Administration*. Available at <https://respaweb.eu/download/doc/Abuse+of+Information+Technology+%28IT%29+for+Corruption.pdf/d867df158b864e8c843c15e5eece5016.pdf> (Accessed 11 April 2023).
- Rogers Y** (1992) Ghosts in the network: Distributed troubleshooting in a shared working environment. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, December 1. New York, NY: Association for Computing Machinery, pp. 346–355.
- Rogers Y and Ellis J** (1994) Distributed cognition: An alternative framework for analysing and explaining collaborative working. *Journal of Information Technology* 9(2), 119–128.
- Rose R** (2005) A global diffusion model of e-governance. *Journal of Public Policy* 25(1), 5–27.
- Rose J, Persson JS, Heeager LT and Irani Z** (2015) Managing e-government: Value positions and relationships. *Information Systems Journal* 25(5), 531–571.
- Rothstein B** (2011) *The Quality of Government Corruption, Social Trust, and Inequality in International Perspective*. Chicago: University of Chicago Press.
- Ruijter E, Grimmelikhuijsen S and Meijer A** (2017) Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly* 34(1), 45–52.
- Scannapieco M and Catarci T** (2002) Data Quality under the Computer Science perspective, *Archives & Computers* 2(1), 1–15.

- Sharp H and Robinson H** (2006) A distributed cognition account of mature XP teams. In Abrahamsson P, Marchesi M and Succi G (eds), *Extreme Programming and Agile Processes in Software Engineering*. Heidelberg: Springer, pp. 1–10.
- Shim DC and Eom TH** (2009) Anticorruption effects of information communication and technology (ICT) and social capital. *International Review of Administrative Sciences* 75(1), 99–116.
- Srivastava SC, Teo TSH and Devaraj S** (2016) You Can't bribe a computer: Dealing with the societal challenge of corruption through ICT. *MIS Quarterly* 40(2), 511–526.
- Vetrò A, Canova L, Torchiano M, Minotas CO, Iemma R and Morando F** (2016) Open data quality measurement framework: Definition and application to open government data. *Government Information Quarterly* 33(2), 325–337.
- Wang RY and Strong DM** (1996) Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems* 12(4), 5–33.