
Disciplining Change, Displacing Frictions

Two Structural Dimensions of Digital Circulation

Across Land Registry Database Integration

Annalisa Pelizza

University of Twente

Abstract: Data acquire meaning through circulation. Yet most approaches to high-quality data aim to flatten this stratification of meanings. In government, data quality is achieved through integrated systems of authentic registers that reduce multiple trajectories to a single, official one. These systems can be conceived of as technologies to settle “data frictions”, controversies about which configurations of actors, agencies, sources and events produce more reliable data. Data frictions uncover two dimensions of data circulation: not only along the syntagmatic axis of alignment, but also along the paradigmatic axis of replacement. Drawing on empirical research investigating database integration at the Dutch land registry (Kadaster), this article aims to contribute to the theorization of digital circulation by recalling two semiotic dimensions along which circulation happens. It argues that even when complex infrastructures are implemented to discipline change, data frictions are not silenced, but displaced along the syntagmatic/paradigmatic axes.

Keywords: Data frictions; database integration; land registry; data circulation; syntagm; paradigm.

Corresponding author: Annalisa Pelizza, Science, Technology and Policy Studies, Faculty of Behavioural, Management and Social Sciences, University of Twente. P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: a.pelizza@utwente.nl.

I. Introduction

Contemporary accounts using the metaphor of digital circulation highlight the “obduracy” of data. “Big Data” applications, for example, conceive of data as stabilized resources that can be combined along integrated, streamlined networks. While it is widely accepted that the way data are matched can have profound sociotechnical consequences (Mayer-

Schönberger and Cukier 2013), the relationship between data and the states of the world from which they are drawn (the “facts”) is less often problematized. As Gitelman and Jackson (2013, 2-3) recall, “data are apparently before the fact: they are the starting point for what we know, who we are, and how we communicate. This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself”.

Media scholars seem to have less problems in recognizing the crafted nature of data. According to Rosenberg (2013), for example, while the term “fact” retains an ontological status, and “evidence” an epistemological one, the semantic function of the term “data” is eminently rhetorical. It indicates the pre-given elements in an argumentation, which do not undergo debate. This rhetorical function of data is not a contemporary trend, though. “Data” have been historically constructed as objective items by design. The first appearance of the term “data” in English, for example, goes back to the seventeenth century, when in mathematics, philosophy and theology it was used “to identify that category of facts and principles that were, by agreement, beyond argument” (Rosenberg 2013, 20). Similarly, in her seminal work on the history of the modern understanding of fact, Poovey (1998) wonders whether facts are incontrovertible data that demonstrate the truth, or theory-laden pieces of evidence aimed at persuasion. Only in the first case, she concludes, facts have been called “data”.

If data correspond by definition to logical assumptions, then the matter of concern for an emerging theory of data circulation is not so much pointing out their inherent bias, as much of the early discursive literature on digitization did (see among others Introna and Nissenbaum 2000). Rather, the question a sound theorization of the digital circulation of data should pass to address concerns the *material processes* through which data can reach the status of (and are kept as) such incontrovertible assumptions. How are data constructed in specific ways throughout the standards, protocols, paths wherein they circulate?

As of today, the constructed nature of data is shared understanding to a growing scholarship at the intersection of Science and Technology Studies (STS) and media studies (see among others Beer 2013; Beer and Burrows 2013; boyd and Crawford 2012; Kitchin 2016; Leonelli 2013; Vis 2013). As the editors of these two special issues rightly point out by drawing analogies with the early work of Appadurai, data acquire value and meaning through circulation (Balbi, Delfanti, and Magaudo 2016). This implies that a theory of data circulation cannot avoid considering the infrastructural technicalities that contribute with potentialities, constraints and path-dependencies to the “biographies and life trajectories” of data (Balbi, Delfanti, and Magaudo 2016, 8). The meaning of a set of data will thus emerge from the sedimentation of multiple passages, bringing trace of the standards, protocols, categorizations through which it reached the current form.

Yet most approaches to data try to conceal this stratification of meanings and values. Achieving high “data quality” has become a sensitive goal in order to implement applications. In many organizations, programs aimed at enhancing data quality provide the normative and technical infrastructure to pursue the flattening of data biographies and life trajectories. Normatively, high quality data are assumed as neutral and a-historical representations of states of the world, a Platonic substance to which actual data used in everyday practices should tend. Technically, sophisticated systems integrating databases at diverse levels are being developed in order to reduce multiple dataset versions to a single, most reliable one.

Achieving high-quality data is, if possible, even more strategic for governments, which conceive of data as the incontrovertible foundation on which policy is erected. As Agar (2003) has recalled, government bureaucracies were even built around the goal of ordering data circulation paths and optimizing data assets. This usually translates in integrating databases in hierarchical order. In contemporary processes of back-office integration, for example, quality of data can be enhanced through so called “systems of authentic registers” (de Vries 2012). Similar systems are meant to provide administrative agencies – in the fulfilment of their institutional duties – only with information from databases that have been labelled as the official (“authentic”) source for that type of information.

Similar integrated architectures aim to reduce multiple, heterogeneous data trajectories to a single one, by legislatively and technically defining which administrative database is leading, and which ones are the recipients. While their goal is to reduce misalignments between data generated in different ways and at different times, double entries and mismatches uncover the irreducible nature of data, even in law-regulated governmental information infrastructures. One can thus wonder how similar systems are implemented, so that data can be certified as “authentic”. Which alternative forms of knowledge, actors and versions of the past are silenced in the pursuit of high-quality data? What are the “costs” of assuring the objective standardization of data?

To use the words of Edwards (2010), systems of authentic registers aim to settle “data frictions”. While “data” is the name given to the outcome of procedures aimed at crystallizing change, attempts to codify change into univocal digital trajectories inevitably entail frictions between competing procedures. In the next sections we thus propose to conceive of data frictions as controversies about which configuration of actors, events and knowledge sources produce more reliable data. This understanding will be exemplified against the frictions arising when older procedures for government data production are replaced with newer, “more reliable” ones for high-quality data production.

The notion of “data friction” will also help us to uncover two dimensions of data circulation. While most studies on digital circulation tend to conceive of circulation as alignment of data, files, digital artefacts from a

point to another, they forget to consider the complementary movement of replacement. However, data and other digital artefacts do not only circulate *along* infrastructures, they also circulate *across* infrastructures. This point is expressed by the structuralist distinction between syntagmatic and paradigmatic dimensions of change, that will be introduced in section 2, as well.

This article's goal is to contribute to the theorization of digital circulation by recalling the two complementary dimensions along which circulation happens: not only along the syntagmatic axis of alignment, but also along the paradigmatic axis of replacement. It argues that, in order to circulate on one of the two dimensions, data have to "pay the price" on the other dimension. Less abstractly, the article shows that even when complex matching infrastructures are implemented to discipline change, digital circulation does not completely silence frictions, but displaces them along the chain of data circulation. Further new actions on the syntagmatic axis must thus be undertaken to handle frictions on the paradigmatic one.

In sections 3 and 4 this argument is exemplified against an empirical research investigating database integration at the Dutch land registry. The research was conducted between 2013 and 2015, and shows that silencing frictions between old and new procedures for enhancing personal data quality required to activate further new procedures and organizational units to handle uncertain cases. Section 5 will summarize the findings and discuss them in the light of the syntagmatic and paradigmatic distinction. Finally, the conclusions will focus on how the empirical findings presented and the concepts used can contribute to the theorization of digital circulation.

2. Data Frictions, the Two Dimensions of Change and the Context of Analysis

"Data" is the name given to the outcome of efforts to discipline fleeting states of the world towards other ends. "Data need to be imagined as data to exist and function as such [...]. [*D*]ata are imagined and enunciated against the seamlessness of phenomena. We call them up out of an otherwise undifferentiated blur" (Gitelman and Jackson 2013, 3, authors' emphasis). This phenomenological stance describes the production of data from an undifferentiated experiential continuum. However, it does not say much about the life trajectories of data processing, distribution, re-use and combination. To what extent can also further steps of data circulation be described in a constructivist way?

Data circulation models usually rely upon the information flow metaphor (Castells 1996), which in turn can be traced back to the origins of telecommunication theory. According to Shannon and Weaver (1949), six

elements constitute any formal model of communication: an information source, a message, a transmitter, a channel, a receiver, a destination. The message is supposed to flow along the channel in a stabilized form and in a linear way. Thus one of the main assumptions of data circulation is that data are discrete full-blown and relatively stabilized units of information circulating throughout digital infrastructures serving as backbone.

This decoupling of content and channel underpinned by the flowing metaphor characterizes most approaches in the Information Society. Van Dijk (2012), for example, sees a lightly deterministic relationship between the networked model of data circulation and the emerging combination of organizational forms leading to network economy. Similarly, Mayer-Schönberger and Lazer (2007) point out the interaction between information flows and power redistribution in government organizational structures, but they do not consider how data themselves undergo transformations along this distribution. In similar conceptualizations of information as “flowing inside” infrastructures, data are conceived of as stabilized entities that can be successfully transported from point A to a point B once the right infrastructural elements are aligned.

However, while “flow” suggests the smooth movement of a liquid, “data do not flow like oil” (Borgman 2015). Nor do they circulate immutable from the database to the application server through multiple web services. Rather, data must be adapted, re-coded, standardized, harmonized. Far from being smooth-running activities, similar practices entail controversies about which configuration of actors, events and sources are generative of more reliable data. When, for example, a bank is required to share data with the tax office, tensions might arise about the format in which data should be codified: language, categorizations, update frequency, granularity of values, to name a few elements. In a similar case, decisions on which format should be adopted imply a pre-emptive decision on which organization and set of procedures can assure the highest quality of data.

In other words, data circulation requires constant attempts to discipline multiplicity and change. A similar disciplining effort inevitably entails tensions. Edwards (2010) has distinguished three types of tensions. “Computational friction” refers to time, human and energy resources needed in the processing of numbers: “[t]he terms ‘input’ and ‘output’ express the moments at which numbers pass from inside the computer to outside it, but many things happen to those numbers before they become input and after they become output. Every calculation requires time, energy, and human attention. These expenditures of energy and limited resources in the processing of numbers can be called *computational friction*” (Edwards 2010, 83, author’s emphasis).

Whereas computational friction opposes the transformation of data into knowledge and information, and thus reduces the amount of information that can be extracted from an input, “metadata friction” concerns obstacles in computation and data processing to make them commen-

surable and comparable. “The effort involved in finding existing metadata, digitizing them, and combining them with whatever metadata you already have might be termed ‘metadata friction’” (Edwards 2010, 318). Edwards and colleagues (2011) further extend the notion of metadata friction to describe conflicts arising among scientists using metadata to enhance interoperability among scientific data, tools and services.

For the purposes of this article, we are particularly interested in the third type of friction, what Edwards called “data friction”. Data friction “refers to the costs in time, energy, and attention required simply to collect, check, store, move, receive, and access data” (Edwards 2010, 84). Data friction is deeply influential for digital circulation, in that it constitutes a constraint to circulation: “[w]henever data travel — whether from one place on Earth to another, from one machine (or computer) to another, or from one medium (e.g. punch cards) to another (e.g. magnetic tape) — data friction impedes their movement” (Edwards 2010, 84).

Noteworthy for us here is the widely encompassing idea of data circulation suggested by Edwards. Data move along two dimensions: not only between different actors and organizations, but also between different media and materialities. Obtaining “data from many locations, consistent across both space and time [requires] a lengthy chain of operations, including observation, recording, collection, transmission, quality control, reconciliation, storage, cataloguing, and access” (Edwards 2010, 84). Any of these translations – be it from one actor to another, or between two different materialities – constitutes an opportunity for data loss or corruption, that is, it offers an interface for data friction.

To illustrate the fragility of this concatenation of processes, Edwards discusses the case of climatological data in the first half of the twentieth century, when meteorological computation witnessed the competition between incumbent analogue and incipient numerical approaches. Data friction was caused, for example, by the need to communicate measurements over telegraph, or between two operators, or to encode them in Morse code. Eventually, data friction widened the practice gap between data used by empirical forecasters and those needed by theoretical climatologists.

The concept of data friction is useful for a theory of data circulation in that it helps to recover two dimensions of circulation: not only between two spatial or organizational points, but also between different materialities. Whereas the materiality of data is a recognised starting point for much literature on data circulation (Dourish and Mazmanian 2013), the debate usually lingers on a vague analytical or even ontological distinction between the socio-cultural and the material (Bates, Lin and Goodale 2016). Differently, we propose that in order to grasp the transformations that data undergo when circulating, we need to appreciate the combination of the two relational, rather than ontological, movements of *alignment* and *replacement*. Section 2.1 introduces this structuralist distinction, and stresses the importance for a theory of data circulation to grasp

not only the alignment of existing infrastructural elements, but also the potentialities that might arise from their replacement.

A second gap in the debate concerns the fact that data frictions are not seen as rich analytical sites (Pelizza 2016a), useful to uncover data biographies, but as undesirable interferences to be silenced. Therefore, the question of how infrastructural processes are implemented in order to minimize data frictions is still an underrepresented field of concern. If data acquire their meaning through their circulation along diverse infrastructural paths, then data frictions do not only concern data as the final outcome, but also the competing procedures through which they are generated. Section 2.2 thus introduces systems of integrated databases as technologies aimed to handle frictions about which configurations of actors, events and sources produce more reliable data.

2.1 The Syntagmatic and Paradigmatic Dimension of Data Circulation

As seen, the notion of data friction helps to uncover two dimensions of data circulation: data move not only between different actors and organizations, but also between different media and materialities. The structuralist notions of syntagm and paradigm can explain this difference without recurring to essentialist distinctions. In summary, we can say that in the literature underpinned by the flowing metaphor data circulation is mainly seen as deploying along the syntagmatic axis, while the paradigmatic dimension is less acknowledged.

Recalling the system vs. process binary, linguist Luis Hjelmslev (1963) distinguished between syntagmatic and paradigmatic functions that articulate the relationship between grammatical, phonetical or semantic items. Relations that link items along the syntagmatic axis are said to be combinatory: they align item A “AND” item B. Items are co-present. For example, the items “plumber”, “faucet”, “fix”, “tape” are aligned along the syntagmatic axis in the sentence “the plumber fixes the faucet with tape”.

Conversely, relationships deployed along the paradigmatic axis are of the type “EITHER...OR”, that is, they are characterized by complementarity. Items in this case are never co-present, but can replace each other while respecting the consistency of the sentence. For example, “tape” and “spanner” can equally fit the slot X in the sentence “the plumber fixes the faucet with the X”: one can be replaced with the other, and they are both variables that can be selected from a shared “plumbing dictionary”. Figure 1 visualizes the two axes and the illustrative sentence.

Given its heuristic potential, this distinction has been applied to much diverse fields of analysis. For example, Jakobson (1960) has further developed it in the field of literary theory, by identifying the mechanism that underpins poetry texts with the co-presence along the syntagmatic axis of items whose relationship usually lies on the paradigmatic dimension of replacement.

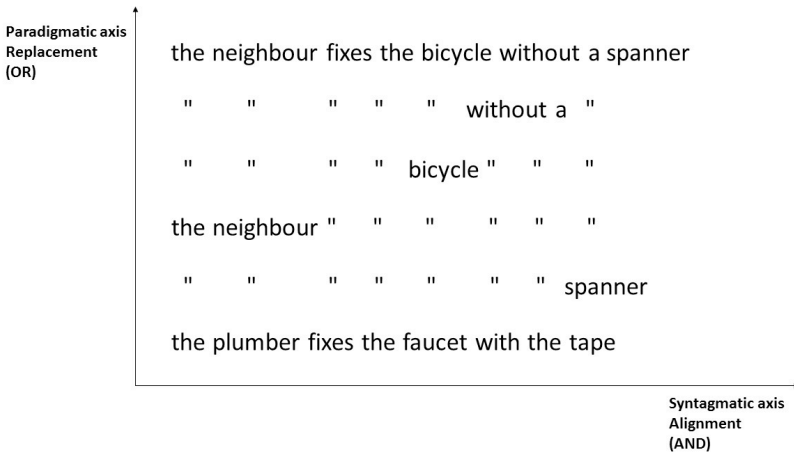


Fig. 1 – Syntagmatic axis of alignment and paradigmatic axis of replacement.

The paradigmatic and syntagmatic distinction has turned out useful also in translational studies of technology, in order to distinguish different types of displacements or “shiftings”. Notably, Latour (1992) resorted to this distinction to describe the dynamics of a steel bar used to keep children in a safe sitting position while travelling on the rear seats of a car. According to Latour’s analysis, the steel bar translates the parents’ verbal orders to a disobedient son into an extra-somatic disciplining artefact. This translation from one kind of materiality (i.e., voice and sound) to another (i.e., steel) is labelled “shifting down” and happens along the paradigmatic axis of replacement. Conversely, the syntagmatic AND dimension ties together the actions actually deployed to reach the final goal of forcing children to a safe position (Latour 1992).

What Latour’s analysis of the steel bar shows is that “it is impossible to move in the AND direction without paying the price on the OR dimension, that is renegotiating the sociotechnical assemblage” (Latour 1992, 172, caption to Figure 10.5). In a similar way, we hypothesize that a sound theory of data circulation needs to consider not only the alignment of infrastructural elements that allow data to flow, but also the new paths, procedures and potentialities that might arise from their replacement; and the mutual influence of alignment and replacement, their “price”.

A close argument is implicitly given by Star and Lampland when they suggest that “slippage [...] between a standard and its realization in action becomes a crucial unit of analysis for the study of standardization and quantification” (Star and Lampland 2009, 15). Bowker and Star further develop this point when they recall that “the slippage between classifications and standards on the one hand, and the contingencies of practice on the other, form core problematics both in the sociology of science

and in studies of use and design in information science. A rich body of work has grown up in both fields that documents the clever ways people organize and reorganize when local circumstances of their activities do not match the prescribed categories or standards” (Bowker and Star 1999, 293)

Despite their frequent references to linguistics, Bowker and Star explain this tension by recurring to the formal vs. informal binary. In an attempt to achieve a higher degree of abstraction and thus generalizability, we suggest that the slippage between standards and practices can be conceived of as a form of friction between different configurations of elements that struggle to replace each other. Therefore, the question arises about the price this replacement requires along the syntagmatic axis. The Dutch Kadaster case below discussed shows that it is impossible to renegotiate the sociotechnical assemblage of data circulation without stretching the length of the circulation path, the number of steps tied together in a linear sequence. That is, contrarily to but coherently with Latour’s example, it shows that it is not possible to move in the OR direction without paying the price on the AND dimension. Before passing to the discussion of the empirical case, we nevertheless need to better introduce governmental interoperable databases as a key field of analysis.

2.2 Integrated Registers as Technologies for Data Friction Settlement

In 2008 Geoffrey Bowker observed that the replacement of past memories by formal information processing is one of the characterizing activities of institutions. “If we completely know a system in the present, and we know its rules of change (how a given input leads to a given output) then we don’t need to bring to mind anything about the past” (Bowker 2008, 4). This obliteration of the past is carried on through standardization of institutional procedures whose goal is to discipline change. This is one way in which institutions can flatten data biographies and life trajectories.

The standardization of procedures is an explicit goal in governmental contexts, in which policy-making requires achieving a level of objectification and standardization of data that allows universal application, what is commonly labelled as “data quality”. Such standardization is often pursued by developing complex integrated database infrastructures (European Commission 2013; OECD 2011; United Nations 2012). The first principle of the Whitehouse’s “Digital Government Strategy”, for example, states that “to harness [federal information] value to the fullest extent possible, we must adopt an information-centric approach to digital services *by securely architecting for interoperability*” (US Federal Chief Information Officer 2012, our emphasis). The rationale is that high-quality data can be obtained by replacing multiple, heterogeneous procedures for

data collection, update and distribution with a single one, thus reducing misalignment between differently generated data.

This is also the rationale underpinning the so called “systems of authentic registers” (Digitale Overheid nd). Systems of authentic registers are interoperable, hierarchically organized information systems in which government databases named as official (“authentic”) sources – e.g., registers of persons, companies, cars – are re-used by other government agencies in the pursuing of their institutional duties. For example, welfare agencies might be allowed to use only personal data taken from a national civil register labelled as “authentic”. Similarly, land registries and cadastres might be the only authorities entitled to provide geographical data to other agencies. These systems have the explicit goal of enhancing the quality of data by reducing all kinds of variations to a standardized version of the data, and their functioning is usually regulated by law.

Similar government systems are expected to comply with five provisions of an unwritten contract with citizens. A “government should: (1) not ask for what is already known; (2) offer quick and good services; (3) not allow itself to be misled; (4) know its facts; and (5) feed the community with confidence” (De Vries 2012, 9). The first principle concerns the circulation of data internal to the government apparatus, that should be preferred to continuously asking citizens for their data. The second and third principles refer to efficiency and correctness, respectively. The fourth to self-reflexivity, and the fifth to the outcome of complying with these principles, in terms of trust building. Consequently, systems of authentic registers are expected to benefit citizens and business (i.e., reduction of administrative burden, tailored solutions), government finances (i.e., cost reduced thanks to streamlining of data management), and government administration (i.e., higher efficiency in rules enforcement).

What is most important to our analysis, authentic registers are underpinned by a distinctive idea about what “data quality” is. While data quality is an umbrella term highly dependent on the intended use of data, and can thus only be defined according to variables related to situated applications (Pipino, Lee and Wang 2002), internal data consistency is a crucial parameter of data quality in authentic registers and government systems at large. “Quality”, in this context, can be understood as the reduction of misalignments, “errors” and inconsistencies among data produced through diverse procedures, agencies and infrastructures (technical executive at Kadaster, see section 4 below). That is, quality refers to the reduction of multiplicity to unity, of multiple data trajectories to a single one.

Being aimed to enhance data quality under this specific understanding, we thus propose to consider systems of authentic registers as technologies to settle data frictions as defined by Edwards (2010). These systems are designed to minimize data misalignments, that is, discrepancies between data produced through different procedures. Therefore, they act as technologies aimed at establishing what should count as “data”: which

configurations of actors, sources, agencies and events produce more reliable data.

By adopting a similar approach to standardization practices, new questions can be formulated. For example, what is the cost of similar friction settlements? Does any trace remain in data circulation of forced standardization of data and procedures? Is the past obliterated once and for all, or might it unpredictably pop up at later stages? In what follows we describe a case in which complex matching infrastructures are implemented to discipline change, enhance the quality of data and reduce data frictions. However, the case turns out to show that similar infrastructures do not completely silence frictions, but displace them along the chain of data circulation. This displacement requires, then, that further actions be undertaken to handle frictions; actions that “pay for” the replacement of old procedures with new ones with the alignment of new procedural steps.

3. Case and Method. The Dutch Kadaster: A Pioneer in Database Integration

The research on which this paper is based was conducted in the context of the “Translating Institutions” project funded by the European Commission’s FP7 Marie Curie Actions. Data collection and analysis were carried on from September 2013 to August 2015 at the Dutch land registry: “Kadaster”.

The Dutch Kadaster is a non-departmental public body, operating under the political responsibility of the Minister of Infrastructure and the Environment. It combines the functions of cadastre, land registry and mapping agency. It collects and maintains administrative spatial data about properties and the rights associated to them. In particular, it maintains the national cadastral map and register, the register of buildings and addresses, the topographical map database, and also the ship and infrastructures registers.

Being originally a government agency, Kadaster was turned into an autonomous body in the early 1990s. From that moment on it had to finance itself autonomously. To this end, it redefined its organization by reducing local offices, and redesigned its service provision policy. On one hand, it univocally identified its customer groups (i.e., civil-law notaries, local authorities, businesses, financial institutions and private individuals); on the other hand, it gave impetus to the creation of new services and programs oriented to those groups. A pre-condition to develop new services was the interoperability of Kadaster and other agencies’ databases.

In the Translating Institutions research we focused on the very first of these integration programs, the one aimed at matching personal data

stored and maintained internally at Kadaster in its person records database (the *Kadastrale Personen Registratie* – KPR) with personal data from the municipal person register (*Gementelijke Basis Administratie Personen* – GBA, now *Basis Registratie Personen* – BRP). KPR contains the first names, family names, birthplace and other data of any owner of a property registered at the Dutch land registry (*Basis Registratie Kadaster* – BRK). On the other hand, GBA is the main source for personal data, a national database kept updated by the joint effort of all Dutch municipalities.

Since April 2007 GBA has indeed been classified by law as the “authentic” register for personal data. Since then the use of GBA-originated data has become mandatory for government agencies (and private actors accountable for public functions, e.g., notaries) in the pursuing of their institutional duties (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties – BZK 2013). In operative terms, this means that whenever the tax authority, Kadaster, or any other government agency need personal data in order to conduct their statutory activities, they are obliged to use data provided by the GBA database.

However, this law was only the point of arrival of a much longer process started in early 1990s. When at that time the Kadaster undertook the endeavour to integrate its KPR database with data from the GBA national register, the system of authentic registers did not exist yet. Conversely, Kadaster was the pioneer in developing database integration across government agencies at a time when few experiments in register integration were carried on. Twenty years later, a national system of authentic registers backed by law was issued, a new organization at Kadaster was set up, and – what is key to this article – a new definition of which procedures can be considered generating “reliable data” was established.

In the Translating Institutions research we reconstructed the trajectory that brought from the very first scattered experiments in database integration to this complex legal, technical and organizational scenario. To do so, we firstly analysed over thirty documents, including laws and decrees, technical and organizational papers, design concepts and schemas, web pages and system screenshots, in-house publications. Secondly, between 2014 and 2015 we conducted nine semi-structured and narrative interviews with technical (4), legal (3) and administrative (2) profiles at Kadaster, both at officer (2) and executive (7) levels. Interviews were audio-visually recorded, and pictures of schemas, drawings, screenshots used during interviews to illustrate the technical details were taken.

Reading of documents allowed the researcher to familiarize with the organizational context and the technical case in particular. It also allowed her to reconstruct the technical functioning of the integration process at different temporal stages. Finally, grey literature reading suggested the main topics to be addressed during the semi-structured interviews.

The interview format started with an unstructured narrative moment, in which interviewees could autonomously produce a self-generating story

of the GBA-KPR integration in the context of Kadaster. This solution was partly requested by interviewees themselves, who prepared self-presentations of the Kadaster and of the integration program, and partly responded to the need to preliminarily establish a trustful relationship between an outsider interviewer and interviewees who tended to strongly identify themselves with the organization. The initial central topic of narration – the changes brought about by the KPR-GBA integration along a 20-year time span – was relevant to both the informants and the researcher (Bauer 1996).

After the narrative interview, a semi-structured interview was conducted by the researcher, addressing the following set of topics: changes that using the KPR-GBA-integrated system has brought to the working routine of Kadaster personnel (both for the specific profile of the interviewee and other profiles), dismantling of old or introduction of new profiles required by the integration, main problems faced by the organization when implementing the integration, main trans-organizational problems (e.g., between Kadaster, notaries and municipalities), difficulties in framing and delegating problematic tasks and eventual solutions (if any), actors' preliminary knowledge and learning processes, opportunities for empowerment, specific technical details about functioning.

Analysis of the material collected was conducted firstly by identifying the main breakdowns (without *a priori* distinguishing among technical, legislative or organizational) identified by interviewees themselves. Then, all actors (humans or artefacts) involved in those problematic situations were mapped, together with the forms of knowledge they were endowed with. Finally, changes in actors/tasks patterns were identified. This last activity brought to a tripartite categorization of actors: existing actors who lost their functions and did not acquire new ones, existing actors who lost their functions and acquired new ones, new actors who were introduced thanks to their knowledge. This article reports only partial results of this thorough investigation, inasmuch as they are relevant to its main argument.

4. Cadastral and Personal Data Integration as Replacement of Bureaucratic Procedures

The integration of cadastral and personal data at the Dutch Kadaster constitutes the archetype of database integration programs as technologies for data friction settlement. In the early 1990s the use of GBA-originated personal data was identified as a possible solution to a set of problems concerning the quality of data stored in Kadaster's own person records database (KPR), and used in its internal processes. In the words of a Kadaster executive, "the integration of KPR with data from GBA has simplified things a lot. Before, we had a lot of problems with double en-

tries. You could have different names for the same person. With the GBA, procedures were simplified a lot and we got a higher quality of data” (technical executive).

For years before the KPR-GBA integration, the procedure to record new property deeds – and consequently vendors’ and owners’ data – in KPR had been based on the principle that the notarial deed was the leading official source. The deed was compiled by notaries using data taken for IDs and sent in both paper and digital format to one of the fifteen Kadaster production teams based at as many local offices. Each local office corresponded to one of the fifteen administrative areas in which the Dutch territory was subdivided. When recording a deed in the Kadaster land registry (BRK), production teams used to query in KPR the new owner’s name, as it appeared in the deed itself, without further verifications. Since deeds were established by law as the leading source, verifications conducted by the notary at the moment of the stipulation were considered sufficient. Figure 2 visually represents this earlier procedure.

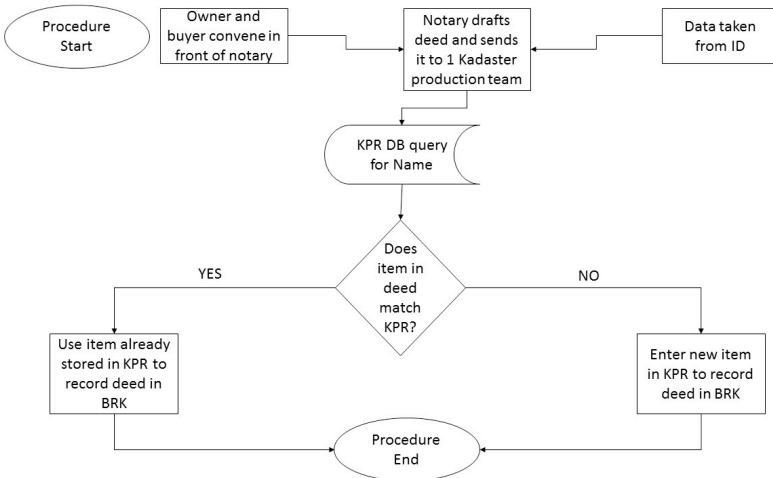


Figure 2 – Former integration procedure. Notarial deed is leading.

According to Kadaster personnel, this procedure could give rise to data misalignments or double entries. For example, if a person had already bought a property and her data had been entered in KPR at the moment of the first acquisition – but in the meanwhile her address, name or gender changed – in the case she subsequently bought another property, a different identity was entered in KPR. In order to avoid this and other misalignments among data collected at different points in time, the KPR-

GBA integration made available the personal data recorded by municipalities in the GBA register.

As Figure 3 shows, with the new system in a night batch all the names newly entered in KPR were sought for in the GBA database, according to a series of pre-defined search sets. If the KPR name matched a name in GBA, a “recipient indication” was entered in correspondence with that item in the GBA database. From that moment on, every spontaneous mutation in an item for which a recipient indication had been placed in GBA was automatically forwarded to KPR in a “push” mode. In case data modifications concerning one person occurred in the GBA, the recipient indication thus automatically generated an information flow towards the KPR, overwriting the previous data for that person.

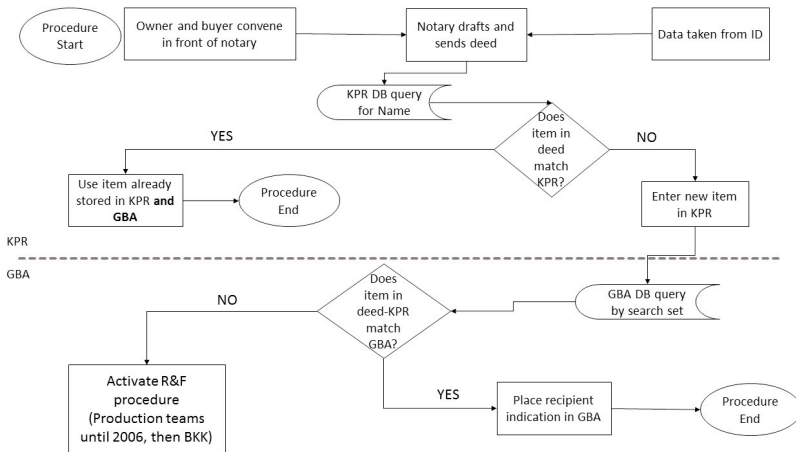


Figure 3 – New integration procedure, automatized.

As visualized in Figure 4, the mechanism of recipient indications established a permanent link between the KPR and GBA databases. This connection also made impossible for Kadaster local production team operators to further modify existing personal data in KPR. As a matter of fact, once a recipient indication was entered in GBA, a locking mechanism was triggered, so that Kadaster operators could not modify existing data anymore. By so doing, the KPR-GBA integration questioned the primacy of the deed as data source. While operators at local production teams continued to use deed-originated data when a new personal item was to be entered in KPR, those data could be overwritten as soon as a link with GBA was established, without for the operators to be possible to restore the original data from the deed. In other words, while with the

earlier procedure the notary deed was the leading source of personal data, with the new integrated system the leading source became *de facto* the GBA database.

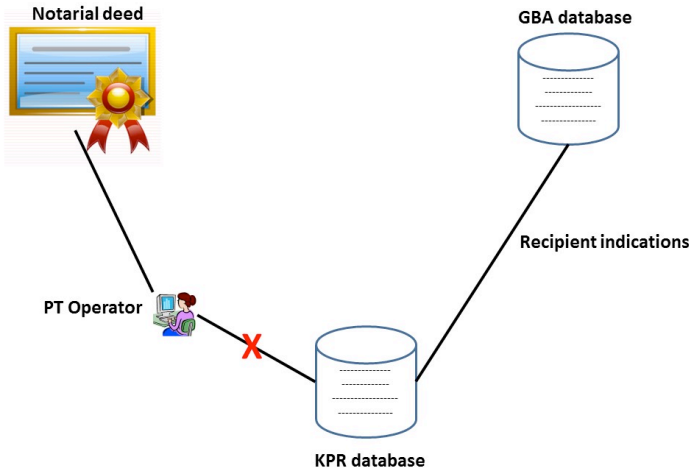


Figure 4 – Recipient indications mechanism. Local production team operators cannot modify data anymore.

The new system however did not only entail the replacement of information sources. Firstly, each of the two sources used to record different events: property passages formalized by notaries in deeds and processed by Kadaster local production teams, in the first regime; changes in name, address, gender, or deaths formalized by municipalities operators and updated in the national GBA node, in the second regime. Secondly, the two regimes involve different *actors and agencies*, endowed with different roles. In the earlier procedure the notarial class traditionally in charge of drafting official documents had a key certifying role, and operators at Kadaster local offices acted as intermediaries in charge of translating this certification into the KPR database. Differently, with the integrated system municipalities officers assumed a certifying role also for Kadaster, and human intermediaries were replaced by recipient indications and a “push” circulation mechanism. Eventually these replacements in actor/role patterns brought to a re-organization at Kadaster, with local production offices passing from fifteen to three (Pelizza 2016b).

Thirdly and summarizing, the new integrated system replaced existing bureaucratic procedures and established new data circulation paths. What indeed was at stake with the KPR-GBA integration was the identification of which configurations of sources, events, actors and agencies were expected to generate the most reliable data. In this regard, the KPR-GBA system of authentic registers is a clear example of controversy-handling technology aimed to settle data frictions. Being optimized to delete traces of the past in the form of non-updated data reported in deeds, and to assure constantly up-to-date data flows, the new integrated procedure was considered to be productive of the best data. The substitution of the earlier procedure with the new integrated one thus constituted at the same time a replacement of the process – i.e., the bureaucratic procedure – and of the outcome of that process – i.e., data.

4.1 The Feedback and Rectification Procedure: From Temporal to Procedural Frictions

However, not even the new integrated system was free from potential conflicts. Frictions arose between the two diverse procedures, which *de facto* continued to compete in establishing which events, actors and data circulation paths were to be conceived as producing reliable data. As one administrative officer put it,

“those connections to us, the duty to use those data, complicate of course the process. We have to use those data in the registration, but the deed also has the right from the notary, so he [i.e., the notary] has also used those key registries, but when he does not do it right, then the connection from the deed to this database – and the connection from all other key registers to us – is not matching. That’s it, the problem.” (Kadaster officer)

Similar frictions are dormant as long as the two procedures overlap – that is, as long as the search set returns univocal items and recipient indications can be placed in GBA (see Figure 3, case “yes” in last branching). It is only when the search set fails to return any result (Figure 3, case “no” in last branching), that mismatches between the two procedures becomes visible. In this case, further mechanisms had to be activated to silence data conflicts.

The mechanism of recipient indications indeed lists a set of cases in which data derived from the deed and entered in KPR do not match data in GBA (see Table 1). In these cases the indication cannot be placed in GBA, and the data user is obliged by law to give a feedback to the data provider. This duty was established as a way to increase the quality of data. According to the legislation, if a government agency doubts the correctness of data, it is obliged to notify this doubt to the data provider, the one managing the authentic register (BZK 2013, article 1.7/2).

- Cases impeding the placement of recipient indication in GBA:
- a – When no or foreign addresses were indicated in the KPR, so that it was impossible to establish in which municipal GBA the search had to be run
 - b – when no person was found in GBA matching the search set provided by KPR
 - c – when more than one person were found in GBA, that match the search set provided by KPR
 - d – when an item found in GBA was subsequently rejected by KPR (which uses stricter control sets than GBA)

Table 1 – Cases in which no recipient indication can be added to GBA.

In order to comply with this provisions, Kadaster introduced a new feedback and rectification (F&R) procedure. This was initially delegated to the Kadaster's local production teams, and then to the Objections, Complaints and Quality Measurement team (*Bezwaren, Klachten en Kwaliteitsteam* – BKK). Team operators firstly checked that no misspellings had been introduced when transcribing the name from the deed to KPR. If this was not the case, then operators submitted a correction request to the notary, in order to exclude that the notary had introduced a misspelling in the deed at an earlier stage. In case the notary confirmed the accurate transcription of personal data onto the deed, a reasonable doubt about the correctness of data stored in GBA could be raised. It was only at this stage that the rectification procedure towards the GBA was activated. While describing the whole procedure would require a considerable amount of detail, for the goals of this article it is sufficient to highlight its extreme complexity, depending on the heterogeneity of cases.

Three aspects are important to note as far as the F&R procedure is concerned. Firstly, feedback and rectification constituted a brand-new procedure, *triggered by a previously non-existent form of knowledge*. As a Kadaster executive recalled in an interview, before the KPR-GBA integration, it was very unlikely that notaries or Kadaster operators identified discrepancies between data in the deeds and in GBA, simply because there was no formalized comparative procedure. Conversely, with the KPR-GBA integrated system the Kadaster is notified whenever it is not possible to place a recipient indication in GBA. This newly generated knowledge makes it mandatory by law for Kadaster to report back to the data provider, that is, to add a further F&R procedure.

Secondly, the F&R procedure was implemented in order to process diversions from the standard procedure, that is, the matching of the old and new procedures. The standard procedure was designed to automatically perform tasks without requiring human intervention. However, when deviations like those listed in Table 1 took place, human mediation

(either production teams or BKK operators) was called back to action, in order to allocate the exception to one of the existing categories, and re-establish order. In this sense, human mediation performs an interpretive, ordering agency, and it is only thanks to it that the standard can be re-established. In other words, while aiming to avoid human intermediation, the standard can exist as such only because its existence is guaranteed by human agency.

Thirdly, the integrated system does only streamline processes as long as data match. It is sufficient that notaries and civil servants at municipalities spell names differently (a very common case in the Dutch dynamic linguistic context), for conflicts between different data circulation paths to emerge. Similar discrepancies are indeed considered “errors” overflowing the standard procedure, not ordinary variations. Therefore, additional procedural steps have to be added in order to compose the mismatch.

Similar “errors” reveal that data variation is the norm rather than the exception. While high data quality is assumed by the promoters of the KPR-GBA integration as a normative goal, friction settlement constitutes the result of ceaseless attempts to erase traces of the past. In this sense, this case substantiates Bowker’s initial insight that formal information processing works towards the replacement of memory (Bowker 2008). The KPR-GBA integration also suggests that the obliteration of the past is carried on at a cost, notably, a procedural cost: a further F&R procedure has to be activated in order to silence data conflicts.

In order to fully understand the implications of this result, we should consider that in the KPR-GBA integration the origin of data misalignment can be attributed to both variations across procedural steps (like the ones reported in Table 1) and across time. A Kadaster officer and an executive have stressed the interrelation between the two kinds of variation:

[Officer] what also makes it complicate are the mistakes from the past.

[Executive] Yes, that is also an aspect. In history some mistakes were made, that have an effect on these days. In the register there are errors from the past, from years ago.

[Officer] Twenty years ago! Then you have to look it up, go back in history and look for the problem, and sometimes you have to change the deed that is behind that. So it is possible that the notary has to change deeds from years ago, to make a new deed, some application on it, because there were mistakes in the past. We are all repairing it. (Administrative officer and administrative executive)

In these words variations take place between data recorded in the past and current states of the worlds. As such, they bring about path dependencies that explain the mismatching of data between different procedural steps. To align similar variations in the present and in the past, further rectification procedures have to be undertaken. In the last part of this article we suggest that similar rectification procedures constitute a price

that pays in the currency of longer procedures the cost of replacing past multiplicities with a single, standardized present.

5. Syntagmatic and Paradigmatic Dynamics

In summary, the KPR-GBA integration illustrates a case in which the implementation of a system of authentic registers entailed a series of frictions about which configurations of data sources, actors and agencies, events can produce more reliable data. The new integrated procedure replaced the leading source (the GBA database instead of notarial deeds), the actors and agencies (municipalities instead of notaries and local team operators), and the events (any change in personal data, rather than property passage) that are involved in the production of high-quality data. As a consequence, new data circulation paths replaced earlier procedures, and the final data eventually turned out to be rather different from data previously produced through the earlier procedure.

This aspect became evident once the two types of data started to differ, and recipient indications could not be placed. Similar misalignments or frictions required the re-introduction of human interpretation and additional procedural steps in order to compose the mismatch. Notably, database integration produced a brand-new form of knowledge highlighting discrepancies between data in the deeds and in GBA. Before the KPR-GBA integration, that kind of knowledge did not exist, because there was no formalized comparative procedure. Conversely, with the system of authentic registers the Kadaster is obliged to deal with the new kind of knowledge and to activate a new feedback and rectification procedure. In this respect, the KPR-GBA system of authentic registers is a clear example of technology aimed to settle data frictions.

The KPR-GBA integration showed that it is impossible to completely silence frictions, and that friction settlement comes “at a cost”. Here the cost is paid in the currency of a new F&R procedure, that stretches the length of the path for high-quality data production. This lengthening “pays for” the replacement of earlier procedures with the new one. In this sense, the F&R procedure displaces frictions along the chain of data circulation.

Recovering Latour’s analysis of the steel bar (see section 2.1), a parallel can be drawn. While Latour’s example shows that it is impossible to move in the AND direction without paying the price on the OR dimension, the KPR-GBA integration deploys a similar but inverted movement. It shows that it is not possible to move in the OR direction without paying the price on the AND dimension. It is not possible to renegotiate the sociotechnical assemblage of data circulation without stretching the length of the circulation path, the number of steps tied together in a linear sequence.

Figures 2 and 3 indeed show data circulation along the syntagmatic axis: a linear sequence of actions that use, recombine, query, match, enter, store, link and rectify data with the aim of obtaining a unique, indisputable, standardized item. However, this is not the only dimension along which change happens in this case. Figure 4 synthetically visualizes the replacement of an older procedure, a configuration of actors (i.e., notaries, Kadaster operators), roles (i.e., certifying, modifying) and events as data sources (i.e., properties buying and selling) with a newer procedure made of diverse events as sources of data, actors, roles. As in the previous example moving along the paradigmatic dimension gave the possibility of substituting “tape” with “spanner”, so in the KPR-GBA case the replacement of notaries as data certifier with municipalities officers, of production team operators with the mechanism of recipient indications, of property registration with changes in name, address, gender and death constitutes a movement along the paradigmatic axis (see Figure 5).

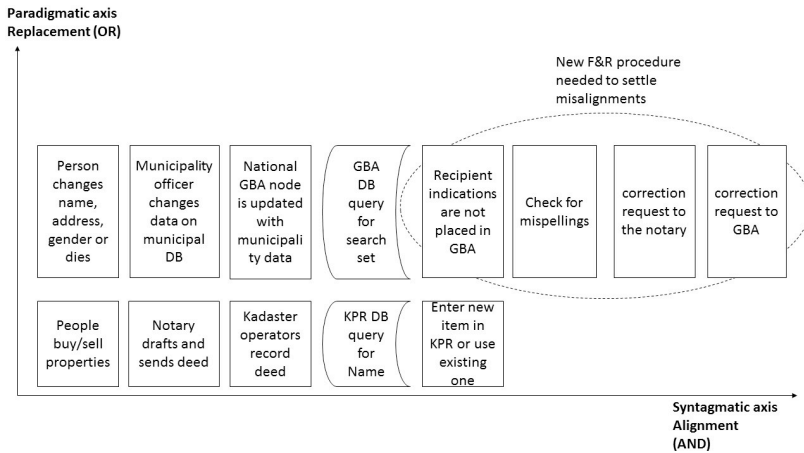


Fig. 5 – KPR-GBA integration along the syntagmatic/paradigmatic axes.

However, this replacement on the paradigmatic dimension requires adding a new, previously inexistent step on the syntagmatic axis. As above recalled, the feedback and rectification constituted a brand-new procedure required by the possibility to track misalignments in data between deeds/KPR and GBA. The replacement of older actors, roles and sources with the new formalized comparative procedure made it mandatory for Kadaster to report back to the data provider, that is, to add a further F&R procedure.

This is what we mean by saying that it is impossible to move in the OR direction without paying the price on the AND dimension, that the obliteration of the past is carried on at a procedural cost. A new F&R proce-

cedure has to be activated in order to silence data conflicts that only with the new integrated procedure come into existence. In this sense, frictions are “displaced” from the paradigmatic onto the syntagmatic axis. Figure 5 visually reports this interplay between the syntagmatic and the paradigmatic dimensions.

6. Conclusions

The adoption of a semiotic, structuralist category of analysis represents an original theoretical contribution, adding a novel instrument to the STS toolkit for the study of data circulation. While the circulation metaphor usually takes into account the syntagmatic distribution of data along a streamlined channel, the KPR-GBA system of authentic registers highlights that a thorough understanding of digital circulation cannot avoid considering also the paradigmatic dimension of replacement. The KPR-GBA case has shown that the length of the circulation path is not independent from a complementary movement of replacement, that is, from which actors, sources, events are chosen to be assembled in the sociotechnical system of data circulation. In this sense, data frictions are “displaced”, translated from the paradigmatic onto the syntagmatic axis: the price of replacement of old procedures with new ones is paid with a lengthened procedure.

All in all, reading the Kadaster findings in the light of the syntagmatic/paradigmatic binary contributes to the debate on data circulation in three respects. First, this research offers evidence to Bowker’s (2008) insight that formal information processing eventually aims to flatten memory into a perpetual present. However, differently from Bowker and Star (1999), it does not explain this tension by recurring to the formal vs. informal binary. Rather, this research attempts to achieve a higher degree of abstraction by uncovering the structural mechanism through which formal/informal dynamics are articulated. Namely, it suggests that the slippage between standards and practices can be conceived of as a form of friction between different procedures – different configurations of actors, sources and event – that struggle to replace each other.

Furthermore, this replacement of old procedures with new ones – but also of the multiplicity of the past with a standardized and unified present – comes at a cost. Replacement requires a price in terms of stretching the length of the circulation path. New technical procedures for data circulation do not generate only more reliable, high-quality data, but also entail an expansion of the infrastructure and the roles, techniques and institutions supporting it. Interpreting this mechanism in terms of syntagm and paradigm allows overcoming the formal vs. informal distinction in favour of a more structural understanding that might be generalizable to very different cases.

Secondly, this paper enriches the discussion on data frictions by further defining them as controversies about which configurations of actors, sources and events are expected to produce more reliable data; and systems of integrated databases as technologies aimed to handle such frictions. Edwards (2010) himself acknowledges that data move along two dimensions: not only between different actors and organizations, but also between different media and materialities. However, Edwards does not explicitly theorize the difference between these two dimensions. Furthermore, once accommodated Edwards' frictions do not seem to leave visible traces.

Differently, by following frictions along the syntagmatic and paradigmatic axes, this research on the one hand conceptualizes the two different kinds of changes as happening either along the AND or the OR dimension. On the other hand, it shows that systems of authentic registers aimed at enhancing data quality do not completely silence frictions, but displace them along the chain of data circulation. Both these conclusions enrich the debate on data friction by suggesting a not so new, but forgotten tool of analysis.

Thirdly, this research contributes to the broader STS debate on data circulation by recalling two dimensions of circulation. It stresses that circulation does not only require that elements are aligned in the right order – as it is implicit in the flowing metaphor, but also that a specific configuration of elements replace other ones in a meaningful way. Furthermore, when this happens, the whole socio-technical assemblage need to be re-arranged, since new forms of knowledge, procedures, actors and agencies are introduced. Since this re-arrangement implies the “payment of a price” either along the syntagmatic or the paradigmatic axis, we propose to replace the long-standing flowing metaphor that data circulation has inherited from early telecommunication theory, with a *payment metaphor*.

In summary, this article shows that – as in the biophysical world nothing is created and nothing is lost, but everything is transformed – so in data circulation the past, frictions, misalignment cannot be fully erased by deploying integrated systems that enhance data quality. Rather, data resist being reduced to standardized items. This resistance cannot be eliminated, but is rather translated, diluted into longer procedures, displaced along the syntagmatic axis. A longer procedure that includes new rectification steps pays the price of efforts to discipline change that cannot be fully silenced. This is a key analytical and theoretical insight for a theory of data circulation.

Acknowledgements

This work has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration

under Grant Agreement no. 326660, Marie Curie Actions Intra-European Fellowship for Career Development. This article however reflects only the author's views and the EU is not liable for any use that may be made of the information contained therein.

The author is thankful to *Tecnoscienza's* editors and the two blind reviewers. She expresses her highest gratitude to Peter Laarakker (University of Twente, Kadaster), Gerard Leenders, Dinie Maas, Theo Splithof and the interviewees at Kadaster for their enthusiastic interest in this project.

References

- Agar, J. (2003) *The Government Machine: A Revolutionary History of the Computer*, Cambridge, MA, The MIT Press.
- Balbi, G., Delfanti, A. and Magaudda, P. (2016) *Digital Circulation: Media, Materiality, Infrastructures. An Introduction*, in "Tecnoscienza. Italian Journal of Science & Technology Studies", 7 (1), pp. 7-15.
- Bates, J., Lin, Y.W. and Goodale, P. (2016) *Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows*, in "Big Data & Society", 2, pp. 1-12.
- Bauer, M. (1996) *The Narrative Interview. Comments on a Technique for Qualitative Data Collection*, in *Methodological Series 1*. London School of Economics and Political Science, Methodology Institute, pp. 1-17.
- Beer, D. (2013) *Popular Culture and New Media*, London, Palgrave Macmillan.
- Beer, D. and Burrows R. (2013) *Popular Culture, Digital Archives and the New Social Life of Data*, in "Theory, Culture & Society", 30 (4), pp. 47-71.
- Borgman, C. (2015) *Big Data, Little Data, No Data*, Cambridge, MA, MIT Press.
- Bowker, G. (2008) *Memory Practices in the Sciences*, Cambridge, MA, MIT Press.
- Bowker, G. C. and Star, S. L. (1999) *Sorting Things Out: Classification and Its Consequences*, Cambridge, MA, MIT Press.
- boyd, D. and Crawford, K. (2012) *Critical Questions for Big Data*, in "Information, Communication & Society", 15 (5), pp. 662-679.
- Castells, M. (1996) *The Rise of the Network Society: The Information Age: Economy, Society, and Culture. Vol. 1.*, Oxford, Blackwell Publishers.
- De Vries, M. (2012) *Funding of a system of key registers in a psi-economics and contemporary perspective: The Dutch experience in a Danish context*, The Hague – Copenhagen, Danish Ministry for Housing, Urban and Rural Affairs.
- Digitale Overheid [Dutch Digital National Government] (nd) *Stelsel van Basisregistratie* [System of Authentic Registers], in <http://www.digitaleoverheid.nl/onderwerpen/stelselinformatiepunt/stelselvan-basisregistraties>, (retrieved December 19, 2015 - only in Dutch).

- Dourish, P. and Mazmanian, M. (2013) *Media as Material: Information Representations as Material Foundations for Organizational Practice*, in P.R. Carlile, D. Nicolini, A. Langley and H. Tsoukas (eds.), *How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies*, Oxford, Oxford University Press, pp. 92-118.
- Edwards, P.N. (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge, MA, MIT Press.
- Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C. and Borgman, C.L. (2011) *Science Friction: Data, Metadata, and Collaboration*, in "Social Studies of Science", 41 (5), pp. 667-690.
- European Commission (2013) Directive 2013/37/EU of European Parliament and Council of 26 June amending Directive 2003/98/EC on the re-use of public sector information, in "Official Journal of the European Union", eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF, (retrieved August 9, 2014).
- Gitelman, L. and Jackson, V. (2013), *Introduction*, in L. Gitelman (ed.), "Raw data" is an Oxymoron, Cambridge, MA, The MIT Press, pp. 1-14.
- Hjelmslev, L. (1963) *Prolegomena to a Theory of Language*, Madison, University of Wisconsin Press.
- Introna, L.D. and Nissenbaum, H. (2000) *Shaping the Web: Why the Politics of Search Engines Matters*, in "The Information Society", 16 (3), pp. 169-185.
- Jakobson, R. (1960) *Closing Statement: Linguistics and Poetics*, in T. Sebeok (ed.) *Style in Language*, New York and London, MIT Press, pp. 350-77.
- Kitchin, R. (2016) *Thinking critically about and researching algorithms*, in *Information, Communication & Society*, published online ahead of print on 25 Feb 2016 (<http://www.tandfonline.com/doi/abs/10.1080/1369118X.2016.1154087-?journalCode=rics20>).
- Latour, B. (1992) *Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts*, in W.E. Bijker and J. Law. (eds.), *Shaping Technology /Building Society: Studies in Sociotechnical Change*, Cambridge, MA, MIT Press, pp. 225-258.
- Leonelli, S. (2013) *Global Data for Local Science: Assessing the Scale of Data Infrastructures in Biological and Biomedical Research*, in "BioSocieties", 8 (4), pp. 449-465.
- Mayer-Schönberger, V. and Lazer, D. (2007) *From Electronic Government to Information Government*, in V. Mayer-Schönberger and D. Lazer (eds.), *Governance and Information Technology: From Electronic Government to Information Government*, Cambridge, MA, MIT Press, pp. 1-14.
- Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data. A Revolution That Will Transform How We Live, Work, and Think*, New York, Houghton Mifflin Harcourt.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) [Dutch Ministry of Internal Affairs and Crown Relations] (2013) *Wet van 3 juli 2013 houdende*

- de nieuwe regels voor een basisregistratie personen (Wet Basisregistratie Personen)* [Law of 3 July 2013 containing new rules for an authentic register of persons (Law Authentic Register of Persons)], Staatsblad 2013, 315, available at <http://bit.ly/1lZsGJd>, (retrieved December 19, 2015).
- OECD (2011) *Communiqué on Principles for Internet Policy-Making*, in “*Brochure on OECD High Level Meeting – The Internet Economy: Generating Innovation and Growth*”, <http://www.oecd.org/internet/innovation/48289796.pdf>, (retrieved August 9, 2014).
- Pelizza, A. (2016a) *Developing the Vectorial Glance: Infrastructural Inversion for the New Agenda on Governmental Information Systems*, in “*Science, Technology and Human Values*”, 41 (2), pp. 298-321.
- Pelizza, A. (2016b) *Materializing Autarchy. Counter-evidence of vendor lock-in in cadastral information systems*, in *Proceedings of the 24th World Congress of the International Political Science Association (IPSA)*, “*Reshuffling Government*” Panel, Poznan (PL), 23-28 July.
- Pipino, L.L., Lee, Y.W. and Wang, R.Y. (2002) *Data Quality Assessment*, in “*Communications of the ACM*”, 45 (4), pp. 211-218.
- Poovey, M. (1998) *A History of the Modern Fact. Problems of Knowledge in the Sciences of Wealth and Society*, Chicago, IL, University of Chicago Press.
- Rosenberg, D. (2013) *Data before the Fact*, in L. Gitelman (ed.), “*Raw data*” is an *Oxymoron*, Cambridge, MA, The MIT Press.
- Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Information*, Urbana and Chicago, IL, University of Illinois Press.
- Star, S.L. and Lampland, M. (2009) *Reckoning with Standards*, in M. Lampland and S.L. Star (eds.), *Standards and Their Stories. How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca and London, Cornell University Press, pp. 3-24.
- United Nations (2012) *E-Government Survey 2012: E-Government for the people*, New York: United Nations, in <http://unpan1.un.org/intradoc/groups/public/documents/un/unpan048065.pdf>, (retrieved August 12, 2014).
- US Federal Chief Information Officer (2012) *Digital Government: Building a 21st Century Platform to Better Serve the American People*, Washington, DC, Whitehouse, <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>, (retrieved August 12, 2016).
- Van Dijk, J. (2012) *The Network Society*, London, Sage Publications.
- Vis, F. (2013) *A Critical Reflection on Big Data: Considering APIs, Researchers and Tools as Data Makers*, in “*First Monday*”, 18 (10), <http://firstmonday.org/ojs/index.php/fm/article/view/4878/3755> (accessed August 4, 2016).