Citizens as Inforgs:
The Challenges of Representation and Privacy in Data-Driven Decision Processes

Jeffrey Alan Johnson, Utah Valley University

Paper Presented at the
2015 western Political Science Association Meeting
Las Vegas, Nevada
April 4, 2015

Abstract

Pressures toward “data-driven” or “evidence-based” decisions are ubiquitous in contemporary public administration. Implicit in this view, however, is a view of data as an objective representation of observed reality. This paper challenges that approach and explores the ethical implications of a constructive theory of data. Creating data requires some process that narrows the many possible representations of a given state of the world to a single data state. This process is carried out within translation regimes: systems of technical rules and social practices that establish a one-to-one correspondence between a given state of the world and a data state. These regimes also translate the entities about which data is collected into “inforgs,” entities that exist solely as bundles of information. Inforgs significantly complicate the way that data-driven decision processes can be considered representative of citizens, requiring decisionmakers to create constructs that ultimately represent themselves rather than students. Standard approaches to protecting citizen privacy are considerably more problematic in translated data processes. When the individuals exist solely as inforgs as in a data-driven decision process, restrictions on the flow of information destroys or at least degrades the inforg itself, excluding the associated person from the process. I conclude by suggesting that a justice-centered approach to information can help manage these challenges.

Assistant Director, Institutional Effectiveness & Planning
Lecturer, College of Humanities and Social Sciences
800 West University Parkway • Orem, Utah 84058
+1.801.863.8993 • jeffrey.johnson@uvu.edu • @the_other_jeff

Citizens as ‘Inforgs’

1 INTRODUCTION

Data, it seems, is to be the savior of the 21st Century. Whether in business, government, or higher education, pressures toward “data-driven” or “evidence-based” decisions are ubiquitous, promising more insight, more efficiency, and better outcomes than was previously possible. Implicit in this view, however, is a scientifically realist view of data: data can save us because it is an objective representation of observed reality that can thus transcend politics to bring organizations to the correct decision. If this view of data is incorrect, the edifice that legitimizes data becomes far less stable.

This paper takes the latter approach to understanding data. It establishes the translation regime as a mechanism by which the social construction of data takes place, and suggests that translation regimes should be viewed as hybrid techno-social structures rather than purely technical ones. Creating data requires some process that narrows the many possible representations of a given state of the world to a single data state. This process is carried out within translation regimes: systems of technical rules and social practices that establish a one-to-one correspondence between a given state of the world and a data state. The technical structures of a relational database, such as tables, functions, business rules, and queries translate states of the world into data states based on standards established by social structures such as cultures, states, and organizations. These regimes also translate the entities about which data is collected into “inforgs,” entities that exist solely as bundles of information.

Inforgs behave quite differently that people within many of the structures that guide data use and data-driven decisionmaking, fundamentally changing the power dynamics of representation in decision process. I explore two structures related to representation in this paper. First, inforgs significantly complicate the way that data-driven decision processes can be considered representative of citizens. While a less data-driven process can support a range of models of representation, a data-driven process that translates citizens as inforgs requires decisionmakers to create constructs that ultimately represent themselves rather than citizens. Standard approaches to protecting citizen privacy are also considerably more problematic in translated data processes. Approaches to privacy typically rely on restricting the flow of information. A traditional approach views this as a protection of an individual. But when the individuals exist solely as inforgs as in a data-driven decision process, restrictions on the flow of information destroys or at least degrades the inforg itself, excluding the associated person from the process. I conclude by suggesting that a justice-centered approach to information can help manage these challenges.

2 METHODOLOGY

The concept of the translation regime developed below is rooted in an examination of the data system in place at Utah Valley University (UVU) while I worked as a Senior Research Analyst in its Institutional Research & Information office from 2009-2013,[1] a system representative of the enterprise data systems used throughout government in most advanced democracies. That experience involved extensive work in data extraction and limited database design and administration, primarily in the Banner ODS database. This is supplemented by discursive

[1] The analysis presented here is strictly my own, and should not be taken as representing in any way the policy of Utah Valley University or the views of its leadership.

analysis of the Structured Query Language (SQL) implementing the data systems and the data standards established by the federal Integrated Post-secondary Education Data System (IPEDS) and the Utah System of Higher Education (USHE) reporting processes, and occasionally by analysis of online interviews with eight UVU students regarding their own perceptions of their social identities conducted as a (highly unsatisfactory) pilot project for a larger study.

UVU's data backbone is the Ellucian Banner relational database running on an Oracle 10g database server.[2] Banner consists of normalized set several thousand data tables managing student and administrative data and optimized for Online Transactional Processing (OLTP), locally referred to as "Prod" (a reference to it as the production database). The bulk of institutional data analysis is performed using the Banner Operational Data Store (ODS), which consists of a denormalized set of fewer but much larger tables optimized for Online Analytical Processing (OLAP). The data contained in the ODS is either identical to or derived from that in Prod. Both databases are extensively customized for UVU. Prod also connects to several other data systems, including the Wolverine Track advising information system, Ellucian Student Success CRM, and the Canvas learning management system.

Most government reporting comes from three customized relational tables. One table, referred to locally as STUDENT,[3] contains information that is constant about individual students

[2] As a full review of database structure and operation is beyond the scope of practicality in a paper, tedious for those already familiar with them, and redundant give the many excellent sources available, this discussion presumes a basic, non-technical understanding of databases. I have aimed to provide enough background to understand the points in my argument in ways that do not overly burden those unfamiliar with databases with technical knowledge but are still recognizable to technical specialists. I apologize to readers of both sorts to the extent that I haven't succeeded in that.

[3] Table and field names will be indicated in capital letters, with TABLE_NAMES in Roman typeface and *FIELD_NAMES* in italics. Specific table names have been replaced with generic descriptive names to maintain data security. These descriptive names are often correspond with similar tables and fields included in a standard Banner installation that may exist but are generally not

across courses within a term such as demographics, contact information, or overall academic characteristics. The second table, COURSE, contains information that is constant across all students in a section for a term. The final table, STUDENT_COURSE, contains information specific to a student within a specific course such as course grade or (since some courses can award variable credit) credits attempted. Using appropriate joins, STUDENT, COURSE, and STUDENT_COURSE can provide most of the information that the institution would need to understand its students and academic offerings. For example, joining STUDENT and STUDENT_COURSE would allow the institution to determine the distribution of courses taken by major and gender. STUDENT_COURSE would identify the courses taken by each student; STUDENT would provide the major and gender information. Each table is a “live” data table, showing data as it exists currently for all terms (including any transactions that affect data for a term after the term has ended, such as retroactive withdrawals from courses). A set of “freeze tables” contain data snapshots allowing time-series analysis throughout a term, and include freezes for the official census and end-of-term reporting dates.

These frozen data from the official reporting dates is used principally for state and federal government reporting. But there is a strong expectation that data reported by the institution for non-government purposes, including that used to make and justify decisions, will be consistent with the government reporting data. For example, between 2010 and 2012, UVU created a web-based data dashboard to provide more specific information on retention and graduation rates than was reported to IPEDS. It nonetheless relied on IPEDS definitions of retention and graduation rates, demographic categories, and reporting cohorts. The cohort definition is especially important, as the IPEDS cohort includes only first-time, full-time, degree-seeking undergraduates entering in fall, who made up only 32% of new UVU students in the

---

used at UVU. Field names have also been changed where the name in the table is sufficiently obscure to make understanding difficult for the reader.

2012-2013 academic year. Because of the expectation that locally-used data will be consistent with government reporting data, the translation regime in place at UVU is defined disproportionately by the rules that govern the three customized government reporting tables.

While certainly the experience with this system is less systematic as a data collection technique than would be ideal, given that the objective of this paper is to establish a theoretical framework for understanding data as a type of social artifact that influences the achievement of social justice, it does not seem unreasonable to interpret that experience using frames and techniques common to emergent methods in social science. The approach used here shares especially some (but not all) features with constructivist grounded theory (Charmaz, 2008). This approach is especially appropriate for the study of information systems on three grounds that are especially relevant to the study of information justice:

> [F]irst, it was useful for areas where no previous theory existed; second, it incorporated the complexities of the organizational context into the understanding of the phenomena; and third, that [grounded theory method] was uniquely fitted to studying process and change. (Urquhart, 2007, p. 341)

There are clear parallels between grounded theory and the work presented here. I use an abductive approach to building theory from experience in which both methods of inquiry and substantive findings are emergent rather than predetermined as I move between experience and theory, testing the concepts developed previously for consistency with further iterations of inquiry. My approach also works at a distance from existing literature on other problems in information systems and technological ethics in order to avoid artificially constraining the emergence of a broader theory of information justice. (Urquhart, 2007, pp. 350–351)

However, I must stress that understanding the creation of data using grounded theory was not intent at the outset of this project; grounded theory is itself emergent in this research. It does not rely on the formal data collection processes of open coding and memo writing,[4] nor

[4] Kelle (2005) and Charmaz (2008) provide exceptional reviews of these specific techniques, defending respectively the two distinct approaches created by the schism between Glaser and

does it remotely approach theoretical saturation. The latter, of course, provides the opportunity to view this quite weak implementation of grounded theory methodology as an initial iteration of the process and thus valuable as a preliminary approach to the question of what data represents. Ultimately, while written in conversation with and as an interpretation of experience, this paper is a work of social theory that aims to understand the structural context of a normative question, and thus makes no claim to any sort of methodological rigor appropriate to more strictly empirical research.

3 INFORGS IN DATA-DRIVEN DECISION PROCESSES

Data-driven decisions are the hegemonic paradigm in public management today, with education policy and administration as one of the central cases. Accrediting bodies demand that mission fulfillment and student learning be demonstrated through “meaningful, assessable, and verifiable data—quantitative and/or qualitative, as appropriate to its indicators of achievement” (Northwest Commission on Colleges and Universities, 2010, sec. 4.A.1) and that institutions practice “regular, systematic, participatory, self-reflective, and evidence-based assessment of its accomplishments.” (Northwest Commission on Colleges and Universities, 2010, sec. 5.A.1) The

Strauss, the founders of grounded theory. While it is impossible to fully argue the point here, I would suggest that the focus on specific methods in that schism has missed the real strength of grounded theory: its reliance on abductively created theoretical concepts that are iteratively tested and refined. This would, especially, be more consistent with the approach’s Peircean roots, in which, I have previously argued (2000), the origin of theory is a creative act and science consists not in the body of knowledge but in subjecting claims abstracted from experience to the examination of further experience. Specific approaches to code development are not necessary for the success of grounded theory in the same way that, for example, successfully passing tests of statistical significance is for quantitative research using a hypothetical-deductive methodology. From this perspective the test of good grounded theory is its tendency to approach theoretical saturation rather than its compliance with any particular research procedure, and specific coding processes are evaluated from a purely instrumental perspective (i.e., is it helpful for moving toward theoretical saturation). The lack of compliance with such procedures in this paper would thus argue for its inefficiency but not its inadequacy as a work of grounded theory.

results of these data-driven analyses are “used for improvement by informing planning, decision making, and allocation of resources and capacity” (Northwest Commission on Colleges and Universities, 2010, sec. 4.B.1) Institutions that fail to use appropriate data-driven processes to evaluate mission fulfillment and student learning risk punitive actions by accreditors. For example, the most recent meeting of the Middle States Commission on Higher Education issued warnings that the accreditation of 10 schools was in jeopardy; nine of these institutions had failed to demonstrate compliance with standards relating to planning, effectiveness, and learning assessment. (Middle States Commission on Higher Education, 2014)

The reliance on data in assessment, evaluation, and planning—arguably the most important decision processes in a university—is a paradigmatic case of the broader model of data-driven decision making in educational administration. Mandated at the K-12 level by the No Child Left Behind Act, data-driven decision making compels institutions to use data “to stimulate and inform continuous improvement, providing a foundation for educators to examine multiple sources of data and align appropriate instructional strategies with the needs of individual students.” (Mandinach, 2012, p. 72) The model is based on business management theories (especially those derived from manufacturing), including Total Quality Management and Continuous Improvement. The model organizes and interprets multiple types of data into information that is meaningful to the users. This then becomes actionable knowledge when users evaluate and synthesize the available information, ultimately using the information to either inform discussion or to choose actions. This process is cyclical and takes place within a range of varying organizational contexts. (Marsh, Pane, & Hamilton, 2006) The result is held to be a more rigorous and informed decision process that allows educators to teach more effectively and administrators to operate more efficiently and reliably. (Mandinach, 2012)

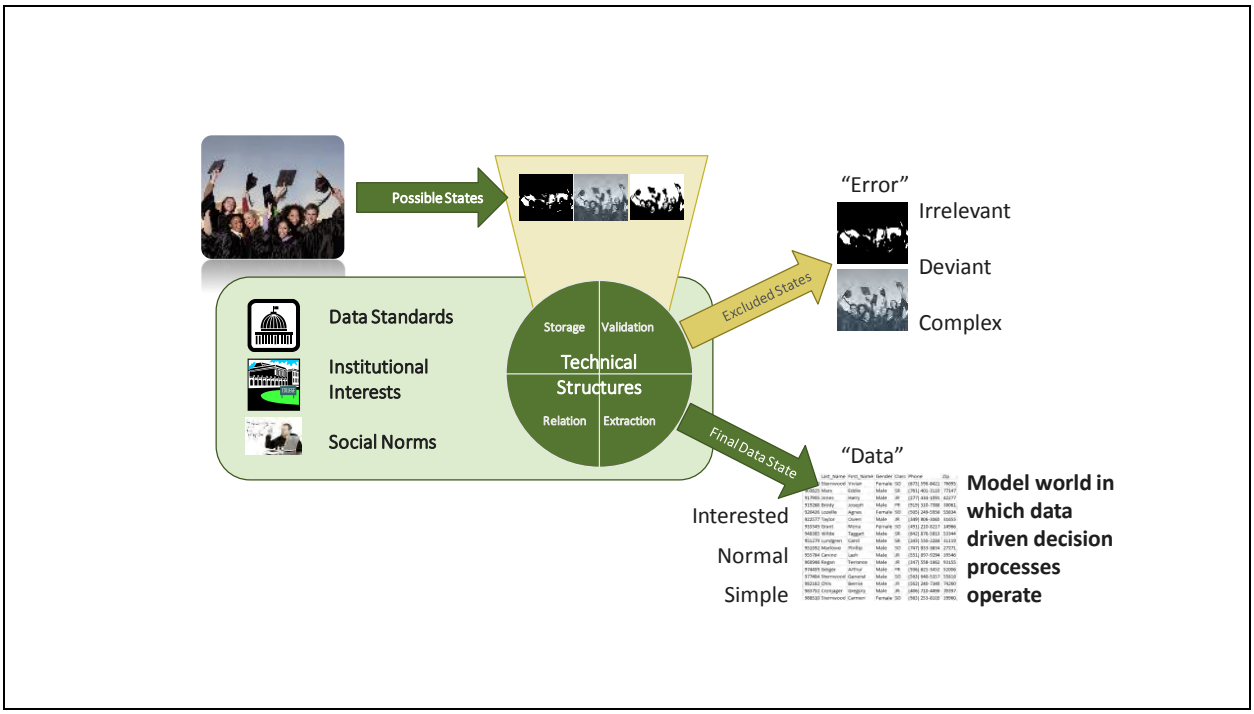
Unexamined in this model is the nature of the data that is driving decision making. The ubiquity of data obscures the fact that it is a form of knowledge unique to the modern, bureaucratic organization, and rooted in such organizations’ needs to make knowledge of its

subjects legible. Legible knowledge transforms reality into standardized, aggregated, static facts that are capable of consistent documentation and limited to the matters in which there is official interest. Such facts emerge from a process in which common representations are created into which cases are classified and which can then be aggregated to create new facts on which the state will rely in making decisions. (Scott, 1998, pp. 80–81) But while the process of producing legibility probably requires something like the illusion of objective representation, it cannot actually produce such a data state. The relationship between a "reality" which is to be represented and a data state representing it is one-to-many; hence a data system that attempts to provide a definitive representation must include a process that selects only one data state from among the many possible representations to the exclusion of all others. It also must do so systematically rather than exegetically, as the final data state chosen by such a process for any one set of circumstances must be commensurable with those states for all other possible circumstances chosen by that same process. These final data states are the *de facto* reality within which data-based decision processes operate.

I have elsewhere (2015) called this process the *translation regime*. Following Krasner's (1982, p. 186) definition of regimes in international relations, one might define the translation regime as the set of implicit or explicit principles, norms, rules, and decisionmaking procedures through which single, commensurable data states are selected to represent states of the world. The translation regime acts as an external source of stability for the data system and allows it to bring legibility to the represented conditions. (Mitev, 2005) Translation regimes are technosocial hybrids (Latour, 1993) composed of two domains: a technical domain that includes the formal rules and procedures of the data system that collect, store, validate, relate, extract, and apply the data; and a social domain of implicit and explicit norms and standards that provide the content of the technical structures. The technical structures embed, enforce, and constrain the substantive worldview of the social domain while at the same time hiding the inherently social nature of data systems behind the appearance of pure technicity. Through these translations,

the reality that the data system purports to represent is irrevocably changed, creating a new reality. That translated reality, not the state of the world from which it originated, informs data-driven processes of decision. One could look to gender as a paradigmatic case of translation, with myriad possible gender expressions reduced to a small number of values, most commonly "male" or "female."

Figure 3.1: The Data Translation Regime



From this perspective, data-driven decision making takes place within an abstracted model world that resembles any reality external to it in one of many possible ways selected by the translation regime. In a data table, data exists in columns where the data has a common framework, but it also exists in rows that relate data points in different columns to each other through association with some sort of entity: data is information *about some things*, students and courses in the case of UVU's core data systems. These things in the database can have no more objective existence than the characteristics that the database attributes to them. The PERSON view contains information about all of the individuals within the UVU data system, including a *PERSON_IDENTIFIER* field that contains a unique identification number for an

person in the data system. The ADDRESS view, however, contains a comparable field called *ENTITIY_IDENTIFIER*, which is identical to *PERSON_IDENTIFIER* for people but also includes identifiers for non-human entities in the system such as vendors. Thus the system creates at least two types of data entities that can be tracked. In the COURSE view, courses are identified uniquely at the level of individual sections within a course number for a given term rather than treating sections and terms as characteristics of a particular course. Courses are thus constructed as groups of students meeting regularly with one or more faculty members to study a set topic rather than as topics themselves. One can see that the translation regime does not simply translate the characteristics of objectively existing entities into the columns of a database; those entities that make up the rows are also translations.

These entities' existence is defined strictly by the information with which they can be associated. This includes not only the columns of the data table in which they exist, but also data in other tables with which they can be associated. UVU's Retention and Graduation Rates Dashboard is organized around students as entities, but in fact relates 17 different data tables built on a wide array of entities: students, degrees awarded, data freezes, visas, admissions application questions, standardized test scores. The information about these entities and the other entities to which they can be related is truly staggering, and both information overload and personal discomfort with the information available are common problem among new analysts in institutional research. But there are things that cannot define the entities. As a public institution with an overwhelming majority of students from a single religious group, UVU has been reticent to collect data on students' religious preferences. Students, as data entities, thus exist without religion. Information about students that cannot be joined to Banner data is similarly nonexistent. In spite of questions on the university's Graduating Student Survey asking whether students intended to complete their degrees at UVU or transfer elsewhere, the Retention and Graduation Dashboard cannot account for that intent because the survey data is in a different data system. Intent is defined strictly by declared program: a student who has declared a major

is defined as a degree-seeking student, one of the core ontological categories of existence in American higher education.

These data entities are best described as what philosopher of information Luciano Floridi terms “inforgs”:

> In many respects we are not stand-alone entities but rather interconnected informational organisms or *inforgs*, sharing with biological agents and engineered artefacts a global environment ultimately made of information, the infosphere. This is the informational environment constituted by all informational processes, services, and entities thus including informational agents as well as their properties, interactions, and mutual relations. (Floridi, 2010, p. 9, emphasis in original)

An inforg is characterized as an entity that is de-physicalized, typified (represented as an instance of a class of identical objects), perfectly clonable, and existing only through its interactions with other inforgs. While the extent to which this ontology, which Floridi calls “informational structural realism,” is an adequate description of being more broadly remains controversial, the sense of inforgs inhabiting an infosphere captures well the ontology of the model world in which a data-driven decision process takes place. In such a model world, data consists of signifiers of states that attach to inforgs. In a star schema, for instance, data is divided into fact tables that describe entities and dimension tables that describe conditions that those entities can take on. Each row in the fact table represents one entity, named by the data table’s primary key, and that entity has no characteristics other than the facts stored in the row, that can be joined to the row, or that are stored in the related dimension tables.

If data systems exist as informational entities inhabiting an infosphere, then Inforgs are the only kind of entity that can exist within a data-driven decision process; all that which is not data is reduced to “context.” The place of context in a system of data-driven decisionmaking is that it can only inform the interpretation of data: only data can drive decisions, and only in the presence of data can a decision be made regardless of the presence or absence of context. Contextual knowledge thus exists only in relation to data, and thus only in relation to entities capable of being described with data. These entities are the only entities that can be said to

exist within data-driven decision processes, and the data about them are the only characteristics that can be said to exist within the model world of a data-driven decision process.

This distinction between data entities and non-data context is central to the evolution of policymaking at UVU's completion efforts. In 2013, UVU implemented a program called "15 to Finish" to improve its low graduation rates. The standard measure of graduation is the IPEDS Graduation Rate Survey, which is a cohort-based measure including only first-time, full-time, bachelor's degree-seeking students who entered in the fall. Among the most recent GRS cohort for which the standard six-year graduation rates were available (the 2005 cohort), only 25% had graduated. The program sought to improve the graduation rates by encouraging students to take 15 credit hours per semester rather than the 12 credit hours that was the minimum for full-time financial aid eligibility.

UVU serves a relatively small number of full-time traditional students. This is not just a matter of numbers; UVU's identity is built in part around inclusivity, which serves as one of the four core themes with which it articulates and evaluates its mission. It maintains an open admissions policy, and most policy discussions involve considering effects on part-time students, non-traditional students, working students, and students with families. UVU's culture is in important ways fundamentally challenging to policies centered on students for whom school is their only significant responsibility. And yet 15 to Finish was implemented with relatively little opposition. Discussions raised the issue of its appropriateness for all students and even the potential to reduce graduation rates among students with significant outside responsibilities. And to its credit, the office implementing the program did so in a way that made clear that this should not be a requirement and that students for whom it was ill suited should not be encouraged to take 15 credits hours.

But the program was intended from the start as a data-driven program, so discussion focused on the IPEDS graduation rates as the program's target. This did not in any strong sense lead to a deliberate decision to "game the numbers" by instituting a policy specifically

targeting the IPEDS GRS Cohort; indeed, many students who could attend full-time were not in the cohort because they transferred to UVU, entered in spring, or would be excluded from the graduation rate calculation due to missionary service. But once the data by which the program is evaluated is defined, model world composed of a specific set of inforgs becomes the entirety of the world. The data-driven process for evaluating whether the program would be successful was defined in relation to the IPEDS graduating rate and thus to the IPEDS GRS Cohort. The first-time, full-time students were the only ones that could affect the graduation rate and thus the only ones in the universe of the data-driven decision, and their very existence as inforgs was defined by the IPEDS GRS Cohort definition. To that extent, they ceased to exist in this constructive process, relegated to merely context in the form of the recommendation to advisors to use 15 to Finish appropriately, even though they should have represented the majority of students. The only students who existed to the program are those that were represented as inforgs.

4 INFORMATIONAL REPRESENTATION

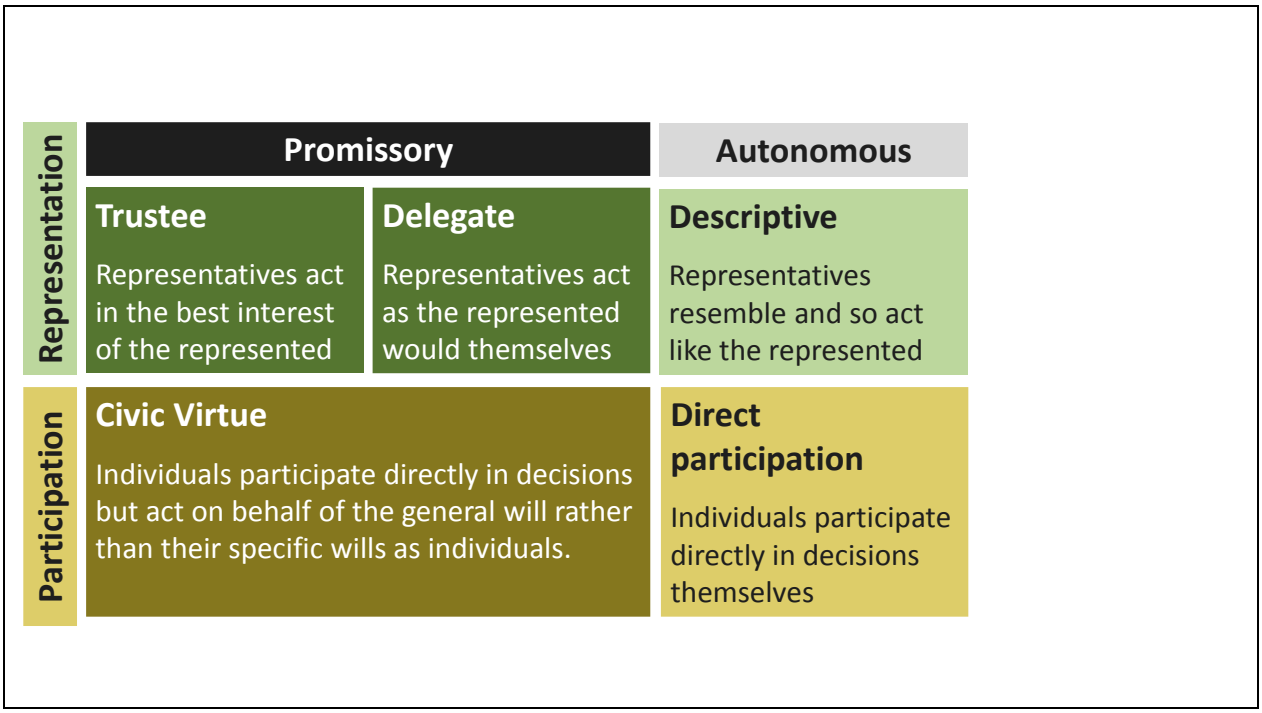
Decisions in public administration are political decisions in the most basic sense: they are decisions made to govern a public entity, in the case above a state university. As such, those that are affected by this decision, as in all political decisions, have a legitimate claim that they ought to have meaningful input into it in some fashion that is not challenged by the fact that the decision takes place within a bureaucratic rather than legislative institution. Presumably, then, decisionmakers in public administration intend for their decisions to represent, in some form and among other considerations, the citizens about whom they are making decisions.

One might analyze different modes of representation along two dimensions. The first concerns the level of participation. Participatory models involve all those who have a claim to input in the process of making the decision; representative models vest that power in a relatively

small group of individuals who act for the group as a whole. A second dimension considers the relationship between the decisionmakers and the group. Promissory models view the decisionmaker as an agent who acts on behalf of those they represent as principals, while autonomous models allow the decisionmakers the freedom to act on their own. The most common models fall into either the autonomous/participatory or the promissory/representative quadrants. Direct democracy, in which all members of the polity participate directly in policymaking, is the standard case of the former; the trustee-delegate dichotomy, in which representatives act respectively in the best interests of the represented or as the represented themselves would, is the basis of the latter.

This is not to say that the only coherent models of representation fit into one of these two quadrants. Frameworks of representation in the two other quadrants are less commonly observed but nonetheless important. Jean-Jacques Rousseau's (1762) *On the Social Contract* proposes a system in which citizens participate directly in government but represent not their particular individual wills but the "will that one has as a citizen," which he terms "the general will," thus directly participating in government but as an agent of the collective body of citizens that serves as principal. Admittedly, this is a very, and probably impossibly, difficult model to operationalize democratically. Descriptive representation is an important model that is both autonomous and representative used especially to study representation in bureaucracies. (See, for example, Wilkins & Keiser, 2004) In descriptive representation, representatives act without any moral obligation toward the positions of the represented but, "in their own backgrounds mirror some of the more frequent experiences and outward manifestations of belonging to the group" (Mansbridge, 1999, p. 628). This correspondence of backgrounds acts as a mechanism to ensure correspondence between the interests of the representative and the represented so that a representative acting on their own judgment is coincidentally acting on that of the represented as well rather than acting out of an obligation to do so.

Figure 4.1: Models of Democratic Decisionmaking



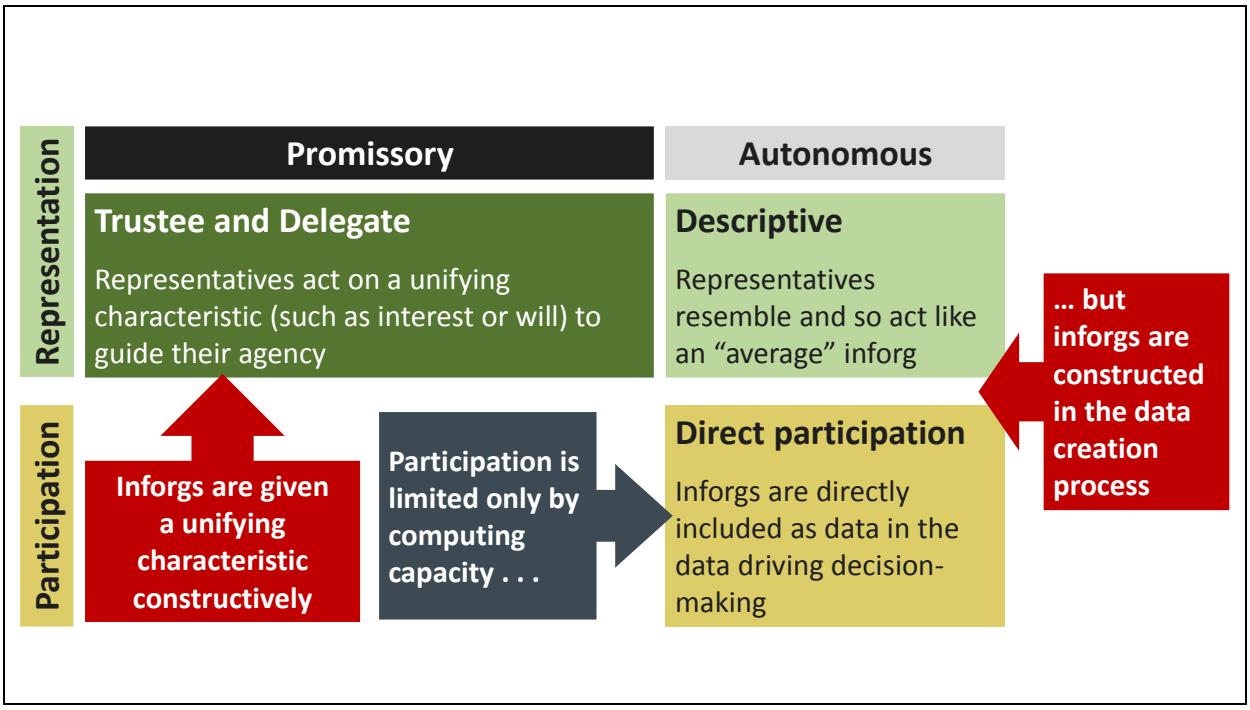
In a personalized decisionmaking context, which we might define in contrast to a data-driven process as one in which either single or multiple decisionmakers use their personal judgment to make what they consider the best decision given the available information under some degree of uncertainty, public administration tends toward a non-autonomous model of representation. Even in the smallest of polities, direct participation in all decisions is impractical because of the number of citizens, number of decisions, and the time-sensitivity of many decisions involved in governing. But while specific institutions may be biased toward one or another of the non-participatory models—students, for instance, "find it difficult to make wise choices" which biases educational institutions toward a trustee model (Parry, 2012)—in general personal decisionmaking can support many models of non-participatory representation.

This model of representation breaks down when citizens are translated into inforgs. Initially, one is tempted to see the translation of those with a claim to voice in a political process as a gain for direct participation. The promissory models both break down when applied to

inforgs. The trustee and delegate approaches both require a unifying concept that acts as the wholeness of the represented (interest or will, respectively) that guides how the agent acts on behalf of the principal, one that is lacking when the principal is no more than a bundle of information: which piece of information defines that unifying concept? But while a personalized process of direct participation requires some complex structure that allows universal participation in the process of developing policy alternatives, manages extensive deliberation among those alternatives, and aggregates preferences into a decision, a data-driven process can bring the participants in as inforgs and then aggregate their informational characteristics. The capacity for participation in data-driven decisionmaking is apparently limited only by the power to collect and process the information that constitutes the inforgs.

This understanding of representation assumes that inforgs have an objective or realist ontological status, existing in their own right rather than being constituted by actors outside of themselves: the data row represents a physically existing citizen as they are in the “real” world rather than existing as an inforg that has been created by someone other than the represented. The analysis of the data structures above shows that this is not the case. Inforgs are themselves social constructs, and both their existence and their characteristics reflect the same social pressures and structures that data fields do. As such, the idea that inforgs are capable of being independently represented in a data-driven decision process is fundamentally unsound; what is represented is the constructive activity of those creating the inforgs. There is the appearance of direct participation, but the participants are not representations of citizens but actants created through the translation regime. The construction of those actants defines the information that constitutes them, supplying the unifying concept of a promissory-representative model independently of the students the inforgs claim to correspond. What is represented is as much the constructors’ understanding of students that is built into the data driving the decision process.

Figure 4.2: Representation of Inforgs in Data-Driven Decision Processes



Data-driven decision processes thus present a fundamental contradiction. While they are instituted as objective processes, it is clear that no process of representing citizens can take place within them without the process of data creation also being a process of imposing external values and assumptions. The inforgs are created by those who create the data system, and decisions about them can only be made if decisionmakers supply their own concepts of interests of will to guide the application of promissory models of representation. This is, to be sure, true of personal decision models as well, but in those models there is a clear connection to individuals against which those assumptions can be checked. In a data-driven model there is nothing to check against beyond the data; the citizens exist solely as data. The objectivity of the process, its supposed virtue, is thus a fiction needed to make the process work.

5 DESTRUCTIVE PRIVACY AMONG INFORGS

Representing inforgs becomes more seriously compromised when considered in relation to information privacy. In the United States, citizens are protected a range of federal and state laws, agency policies, and data handling standards. All of this is meant to ensure that citizens are able to maintain a sphere of personal identity and activity safe from intrusion by others. Most commonly this is protected by the twin principles of consent and anonymity: personal information may only be used or transferred with the consent of the subject; all other information must be stripped of personally identifying characteristics before use or transfer. (van Wel & Royakkers, 2004) Certainly these opt-in or opt-out procedures are the bedrock of most institutions' privacy policies, with the latter likely far more common than the former.

Increasing pressures on personal privacy have given rise to more complex perspectives on privacy. It is increasingly common to interpret privacy as a property right in information about one's self. Subjects hold initial ownership rights in information about them, and can exchange that information contractually in information markets, receiving appropriate compensation—or they can refuse to permit the use of such information in the absence of sufficient compensation to encourage the transaction. (Solove, 2004, pp. 76–81) This approach makes sense, for example, of the willingness of so many to give access to their email to Google: in exchange for an outstanding product, consumers are willing to allow Google to use the information captured to generate profit for itself. Alternatively, Helen Nissenbaum (2010) argues for a reliance on social context to protect privacy. As technosocial systems, the context of information flows is as much a defining feature of data exchange and use as the content of that information flow. The combination of situation, actors, information attributes, and practices of transmission for accepted information exchanges constitute an existing norm of practice that may be violated in the case of new uses of information, such as a data mining practice. Changes in this context that are not supported by its underlying norms are violations of the contextual integrity of the

information flows, and in the absence of separate justification violate one's privacy rights. More recently, the European Court of Justice has embraced a "right to be forgotten" under which individuals are entitled to have information about them essentially destroyed, in the instant case by having Google remove links to information about them from search results. (Case C-131/12 *Google Spain SL v. AEPD* [2014]).

The common thread of each of these approaches to privacy is that they aim to restrict flows of information across parties, transactions, or both. This restriction is frequently considered the essence of data privacy. The centrality of collection (the flow of information from a subject to a data system) and dissemination (the flow of information across data systems or from a data system to subjects) in common definitions of information privacy makes restrictions on flow the *sine qua non* of data privacy. Such a model of privacy is at least plausibly appropriate for the governance of subjects who are persons; preventing the transfer of information will, presumably, prevent those receiving information from using it to do harm to the subjects of that information. This meets the fundamental criteria of a wide range of ethical frameworks, such as Mill's harm principle, which permits the infringement of one's liberty in order to prevent harm to others, or the more recent proposal of a Hippocratic Oath making "do no harm" the first principle in the use of information and communication technology for development. (Mill, 2011, p. 17; Rodrik, 2012)

Restricting the flow of information fundamentally fails, however, when the subjects are constructive inforgs. The flow of information is what translates subjects into inforgs in the first place. To restrict that flow is to change the inforg itself. Such restrictions might, for instance, limit the data known about an inforg in absolute terms as privacy restrictions prevent the transfer of certain types of information (when, for example, the subject opts out of sharing of internet use information). Or it might do so in relative terms as it prevents the transfer of information from one source (when the subject installs a privacy plug-in in Firefox) but allows that same transfer from another source (when the subject doesn't bother reading the 31-page terms and conditions

for the latest iOS update). A difference in the information constituting the inforg violates the principles of typification (the difference resulting in inforgs that are instances of two different types) and clonability (the difference distinguishing two instances as different rather than as clones), and is thus the creation of a different inforg.

This becomes even more problematic when a subject opts out of a data system altogether. For a constructive inforg, a complete data opt-out is not simply a withholding of information; it is a complete destruction of itself as an inforg. Prohibition of data flows prevent the inforg from being constructed in the first place. It perhaps only slightly overdramatic to characterize complete restriction of the flow of data as information suicide for a constructive inforg, as the inforg that protects its privacy ceases to exist in the model world of the data-driven decision process. The physical entity corresponding to the inforg (in this case, the actual citizen) is at best reduced to context—that there are some citizens who are excluded by privacy protections. But context, again, exists only in relation to data, which is to say in relation to inforgs. Citizens who opt to protect their privacy thus exist only as others to the inforgs' selves, defined not individually as entities in themselves but collectively as a typified characteristic of the inforgs (i.e., as a group of identical entities of which the inforgs are not members). Reduced to context that is meaningful only in relation to entities that have corresponding inforgs, those citizens cease to exist analytically and instead are subsumed as information into inforgs corresponding to other citizens.

That further complicates the problem of representation as well. Partial restrictions change how subjects are represented; complete prohibitions exclude subjects from being represented entirely. Citizens are faced with a difficult choice: they can be represented (with varying levels of adequacy given the process of constructing inforgs) in the data-driven decision processes that run the institution that shapes a significant part of their lives, or they can choose to minimize the extent to which that institution is allowed into their sphere of private activity and identity. To exactly the extent that citizens choose one good, they undermine the other. In

personalized decision processes, the unifying concepts of principal-agent representation can moderate this, with decisionmakers taking into account expressions of citizens' best interests and wills regardless of—and perhaps taking into consideration—the privacy status of individual citizens, as these are not data-dependent. In data-driven decision processes, however, with those unifying concepts absent and decisions formally constrained by the available data, representation and privacy are fundamentally irreconcilable.

6 CONCLUSION: WHAT IS TO BE DONE?

Clearly data-driven decisionmaking becomes much more problematic when we recognize that data is made, not collected. As decisionmaking increasingly takes place within model worlds created by the process of collecting, managing, and analyzing data it increasingly transforms people into inforgs, and marginalizes considerations not rooted in data as mere context.[5] Data-driven decisionmaking is part of a larger ethos, one connecting managerialism, technocratic government, and neoliberal politics, that increasingly pervades public administration. The problems of representation and privacy, and especially the tension between the two, stem from the very core of this ethos.

There are thus no clear or easy solutions to these problems. Immediately, one might hope that awareness of the problem might promote decisionmakers to take mitigating steps. Agencies might take efforts to broadly discuss findings before making final decisions based on them in order to identify the ways in which a particular data-driven process has not adequately

[5] To be sure, one might argue that the portrayal of data-driven decisionmaking presented here is something of a straw-man argument that neglects the subtleties of and importance of context in the models advocated in public administration. I would argue to the contrary that those models themselves only pay lip service to context; the more context can be used to override data and the more that conflicting data points are to be considered in the decision process the less data-driven decisionmaking is distinct from personalized decisionmaking. If there is something distinct about data-driven decisionmaking it is that data must take priority over context.

represented citizens. Data analysts might also analyze generally the privacy decisions of citizens to better understand who is exercising their privacy options, what concerns drive those citizens, and what effects those privacy elections are likely to have on data once it becomes part of a data-driven decision process. Steps like this will not solve the problems, but they can at the least create additional context that can moderate the effects of representing citizens as inforgs rather than people.

An emerging agenda exploring “critical institutional research” in higher education might also be of use more generally. Such an agenda aims to explore educational data administration from perspectives of critical thought about current ideas and practice, critical methods that challenge the dominance of positivist or behavioralist research, and critical theories that take seriously diverse perspectives and the places of race, class, gender, sexuality, and especially the intersections of these in educational data. This analysis illustrates the value of critical thought about educational data; critical methods might challenge the dominance of data-driven decisionmaking with methods that emphasize synthesis of research subjects rather than analytically fragmenting them into data bits while critical theories would challenge the claims to objectivity of data and explore connections between data-driven processes and neoliberal political and economic ideology. Together, these idea would enable institutional researchers in higher education (and, by extension, government data analysts more generally) to consider our practices from the perspective philosopher Iris Marion Young takes on critical politics: rather than being the natural or obvious way of doing things, our practice “does not have to be this way, it could be otherwise.” (Young, 1990, p. 6)

But ultimately the clear failure of existing notions of representation and privacy suggest that they might be better subsumed into a conception of information justice that broadens these specific concerns into a large framework that understands information as a social structure and evaluates it from the perspective of how it contributes to the good of individuals, groups, and society more broadly. Such a perspective would clearly challenge the authoritative claims of

data-driven decision models, showing that these decisions are inherently political acts and thus that the social structure of information is itself of interest to both decisionmakers and the subjects of those decisions. While the details of such a theory are quite far from developed, it would certainly encourage more inclusive decision practices, and consider of both the ends for which information is to be used and the means by which information is analyzed as ethical questions in themselves. A more explicitly just framework for information might even go further to create an environment in which citizens and government cooperate in the pursuit of justice rather than conflict over who has the right to citizens' information. A robust framework for information justice is the long-term solution for protecting citizens as inforgs.

REFERENCES

Charmaz, K. (2008). Grounded Theory as an Emergent Method. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of Emergent Methods* (pp. 155–170). New York: The Guilford Press.

Floridi, L. (2010). *Information: a very short introduction*. Oxford ; New York: Oxford University Press.

Johnson, J. A. (2000). Abductive Inference and the Problem of Explanation in Social Science. Presented at the Midwest Political Science Association, Chicago.

Johnson, J. A. (2015). Information Systems and the Translation of Transgender. *TSQ: Transgender Studies Quarterly*, *2*(1), 160–165. http://doi.org/10.1215/23289252-2848940

Kelle, U. (2005). "Emergence" vs. "Forcing" of Empirical Data? A Crucial Problem of "Grounded Theory" Reconsidered. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *6*(2). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/467/1000

Krasner, S. D. (1982). Structural Causes and Regime Consequences: Regimes as Intervening Variables. *International Organization*, *36*(2), 185–205.

Latour, B. (1993). *We have never been modern*. Cambridge, Mass: Harvard University Press.

Mandinach, E. B. (2012). A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. *Educational Psychologist*, *47*(2), 71–85. http://doi.org/10.1080/00461520.2012.667064

Mansbridge, J. (1999). Should Blacks Represent Blacks and Women Represent Women? A Contingent "Yes." *The Journal of Politics*, *61*(3), 628–657.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making Sense of Data-Driven Decision Making in Education: Evidence from Recent RAND Research*. RAND Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/occasional_papers/2006/RAND_OP170.pdf

Middle States Commission on Higher Education. (2014, June 26). Summary of Commission Actions on Institutions. Retrieved September 22, 2014, from http://www.msche.org/institutions_recentactions_view.asp?dteStart=4/29/2014&dteEnd=6/26/2014&idCommitteeType=0&txtMeeting=Commission

Mill, J. S. (2011). *On Liberty* (Project Gutenberg eBook). London: The Walter Scott Publishing Co., Ltd. Retrieved from http://www.gutenberg.org/files/34901/34901-h/34901-h.htm

Mitev, N. N. (2005). Are Social Constructivist Approaches Critical? The Case of IS Failure. In *Handbook of Critical Information Systems Research: Theory and Application* (pp. 70–103). Northampton, Mass: E. Elgar Pub.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Staford, California: Stanford Law Books.

Northwest Commission on Colleges and Universities. (2010). Accreditation Standards. Retrieved September 22, 2014, from

http://www.nwccu.org/Standards%20and%20Policies/Accreditation%20Standards/Accreditation%20Standards.htm

Parry, M. (2012). *College Degrees , Designed by the Numbers*. Retrieved from https://chronicle.com/article/College-Degrees-Designed-by/132945/

Rodrik, D. (2012). *A Hippocratic Oath for Future Development Policy A Hippocratic Oath for Future Development Policy*. Retrieved from http://www.policyinnovations.org/ideas/commentary/data/000244

Rousseau, J.-J. (1762). *On the social contract*. Indianapolis: Hackett Pub. Co.

Scott, J. C. (1998). *Seeing like a state: how certain schemes to improve the human condition have failed*. New Haven: Yale University Press.

Solove, D. J. (2004). *The digital person: technology and privacy in the information age*. New York: New York University Press.

Urquhart, C. (2007). The Evolving Nature of Grounded Theory Method: The Case of the Information Systems Discipline. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 339–360). Los Angeles; London: SAGE.

Van Wel, L., & Royakkers, L. (2004). Ethical issues in web data mining. *Ethics and Information Technology*, *6*(2), 129–140. http://doi.org/10.1023/B:ETIN.0000047476.05912.3d

Wilkins, V. M., & Keiser, L. R. (2004). Linking Passive and Active Representation by Gender: The Case of Child Support Agencies. *Journal of Public Administration Research and Theory*, *16*(1), 87–102. http://doi.org/10.1093/jopart/mui023

Young, I. M. (1990). *Justice and the politics of difference*. Princeton, N.J: Princeton University Press.