V viewpoints

Jeffrey Johnson

Computing Ethics
The Question of Information Justice

Information justice is both a business concern and a moral question.

"ONCE THE ROCKETS go up, who cares where they come down; That's not my department," says Wernher von Braun.
—*Tom Lehrer*

In the 1990s, the government of India began a program to digitize and open land records, one rooted in what open data proponents tout as its chief virtue: "The Internet is the public space of the modern world, and through it governments now have the opportunity to better understand the needs of their citizens and citizens may participate more fully in their government. Information becomes more valuable as it is shared, less valuable as it is hoarded. Open data promotes increased civil discourse, improved public welfare, and a more efficient use of public resources."[a]

Digitizing the Record of Rights, Tenants, and Crops (RTC) along with demographic and spatial data was intended to empower citizens against state bureaucracies and corrupt officials through transparency and accountability. Sunshine would be the best disinfectant, securing citizens' land claims against conflicting re-

cords. In fact, what happened was anything but democratic. The claims of the lowest classes of Indian society were completely excluded from the records, leading to the loss of their historic land tenancies to groups better able to support their land claims within the process defined by the data systems. Far from empowering the least well off, the digitization program reinforced the power of bureaucracies, public officials, and developers.

This case illustrates an underappreciated challenge in data science:

creating systems that promote justice. Data systems, similar to von Braun's rockets, are too often assumed to be value-neutral representations of fact that produce justice and social welfare as an inevitable by-product of efficiency and openness. Rarely are questions raised about how they affect the position of individuals and groups in society. But data systems both arbitrate among competing claims to material and moral goods and shape how much control one has over one's life. These are the two classic questions

---

a See 8 Principles of Open Government, 2007; http://bit.ly/1KbMC0I. For further discussion see the author's article "From Open Data to Information Justice," in *Ethics and Information Technology 14*, 4 (Dec. 2014), 263–274; http://dx.doi.org/10.1007/s10676-014-9351-8 and http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=1459381.

of philosophical justice, raising the question of information justice.

Data is a social process, not just a technical one. Data, the structures that store it, and the algorithms that analyze it are assumed to be objective reflections of an underlying reality that are neutral among all outcomes. But data scientists constantly make choices about those systems that reflect both technical and social perspectives. One common validation table for a gender field contains only two values, "Male" and "Female." But many include an "Unspecified" value as well; Facebook began allowing dozens of different values in 2014. Or the validation table might not exist at all, storing whatever text subjects want to use to describe their gender identities.

A data architect charged with storing such data must choose the specific architecture to be implemented, and there are few truly technical constraints on it; yet practice often depends on adopting one answer around which a system can be designed. The design of the gender field and its associated validation table are thus, in part, social choices. They might be decided based on conscious decisions about gender norms, or based on organizational data standards that are the result of social or political processes. The Utah System of Higher Education, for example, deprecated the "Unspecified" value in 2012 to make reporting comply with data standards for the U.S. Integrated Postsecondary Education Data System. This makes them, as much as any act of the political systems, choices about how society will be organized.

Information systems cannot be neutral with respect to justice. Information justice is a reflection of Kranzberg's first law: "Technology is neither good nor bad, nor is it neutral."[3] Many questions of information justice arise as a consequence of a problem familiar to data scientists: the quality of data. "Injustice in, injustice out" is a fundamental principle of information systems. Gender is a good example of what I call the "translation regime" of a data system. Because many different data frameworks can represent the same reality, there must be a process of translating reality into a single data state. That incorporates technical constraints but also the social assump-

This is not simply a problem in data architecture that can be overcome by better architecture.

tions and processes of that reality. Together, these structures translate reality into a definitive data state. When the translation regime is built around injustice—when it is built around one group's prejudices about the world, for instance—the associated data system perpetuates that injustice by embedding it in the processes the data informs while at the same time hiding it behind a veil of technicity.

Translation is not the only way that injustice enters data systems. Data collection processes may bias data in favor of the privileged. The undercount of minority households in the decennial U.S. Census is a consequence of a wide range of barriers to participation, barriers that are not equally distributed. The result is that the process is more likely to capture data about those who are relatively privileged rather than those who work two jobs and are rarely home, who do not live as a single family unit, who move frequently, who do not speak the most common languages in their community, who have learned to be wary of government.[5] These kinds of collection issues can bias conclusions in favor of the privileged, as when cities conclude that building code violations are as common among the wealthy as among the impoverished because the number of violations reported by residents is equal, not considering the wealthy have far lower thresholds for reporting violations when they see them.[1]

This is not simply a problem in data architecture that can be overcome by better architecture. The problem exists because the data architecture is embedded in a wider constellation of problems, models, and actions. As Lehrer's caricature of the famous rocket scientist suggests, data

scientists cannot be content to say the use of their systems is someone else's problem: where the rockets are meant to come down determines the specifications of the system. Learning analytics systems becoming increasingly common in higher education are built to address the problem of, as Austin Peay State University's provost described it, students who "find it difficult to make wise choices."[4] But wise choices do not necessarily mean the same thing to the provost as they do to any particular student. Austin Peay expects students to graduate on time; decisions that lead away from that are unwise from the institution's perspective. Hence a data system that includes the information and models needed to predict students' course grades but not how much they are challenged. The design specifications and intended uses of a system are key sources of its social elements: data systems only capture what is made legible to them, and that depends on what the data system exists to do.

All of these factors contributed to the failure of RTC digitalization. The RTC was accepted to the exclusion of the kinds of informal and historical knowledge that had long been the basis of land claims in the region. It was stored in a relational database that could not easily query the kinds of unstructured documents that supported other claims. The RTC itself was based on a model of individual ownership that was not the only ownership practice in the region; some land was historically held in common for the community in ways that could not be reflected in the RTC process.[6] Digital land records maintained in geographic information systems, in spite of being open, became a tool for obscuring the needs of some citizens, for barring participation, and for undermining public welfare.

Information injustice is not a problem of bad data systems, nor are data systems inherently unjust. Robust information systems may just as easily promote as undermine justice. The Map Kibera project used crowdsourced data to identify public services available to residents of slums in Nairobi, Kenya that officials regarded as illegal and thus non-existent.[2] In this case, data acts as a countervailing power to

oppression by government. Alternative data systems that would have improved the outcomes of the Indian case described here might have digitized more than just the RTC, used a data architecture more friendly to unstructured data, built analytical approaches that did not assume all land was privately owned, or aimed to coherently document and resolve land claims in practice rather than identifying a definitive owner for the purpose of public administration.

There are probably no universally right or wrong choices in information justice, but this does not absolve data architects from considering the justice of their choices and choosing the better over the worse, and when that cannot be done through technical means what is left is an of politics. A useful solution to information justice is thus to practice information science in a way that makes politics explicit.

One increasingly common way to do this is for information scientists to work with social scientists and philosophers who study technology. There is precedence for this: anthropologists have become frequent and valued collaborators in user experience design. Expertise in the ethical and political aspects of technology can inform the unavoidable choices among social values as opposed to pretending these choices are merely technical specifications.

The same can result from more participatory development processes. If we understand data systems as part of a broader problem-to-intervention nexus, we see the end user is not the person receiving the data report but the one on whom the intervention acts. Just as consulting the data user is now regularly part of business intelligence processes, consulting the people who are subjects of the system should be routine. Their participation is crucial to promoting information justice.

To be sure, justice should be its own reward. But information scientists must be aware of the consequences of information injustice, consequences that go beyond the compliance concerns with which many are already familiar. Student data management firm inBloom provided data storage and aggregation services to primary and secondary schools enabling them to track student progress and success using not only local data but data aggre-

gated from schools nationwide. Many districts and some entire states adopted inBloom, but the aggregation of such data raised deep concerns about student privacy. After several states backed out of the arrangement because of these concerns, the company ceased operations in 2014.

CEO Iwan Streichenberger attributed inBloom's failure to its "passion" and a need to build public acceptance of its practices—in essence rejecting the legitimacy of the ethical concerns its critics raised.[7] Whether one accepts the legitimacy of those claims or dismisses them as old-fashioned (as is quite common among information technologists), there is no question that inBloom's business failure was not one of inadequate technology but of inadequate ethical vision. InBloom either failed to appreciate the ethical risks of its technologies and business model or failed to convince the public of new ethical principles that would support them. Either way, information justice has become a business concern as much as a moral one.

But whether a business concern, a moral one, or a political one, the challenge information justice presents is one that can be met. It requires that information scientists work with an eye toward the social, asking critical questions about the goals, assumptions, and values behind decisions that are too easily—but mistakenly—seen as merely technical.　ⓒ

References
1. Big data excerpt: How Mike Flowers revolutionized New York's building inspections. *Slate*, 2013; http://slate.me/1huqx0p
2. Donovan, K. Seeing like a slum: Towards open, deliberative development. SSRN Scholarly Paper (Mar. 5, 2013) Social Science Research Network, Rochester, NY; http://papers.ssrn.com/abstract=2045556
3. Kranzberg, M. Technology and history: "Kranzberg's Laws." *Technology and Culture 27*, 3 (July 1986), 544–560.
4. Parry, M. College degrees, designed by the numbers. *Chronicle of Higher Education.* (June 18, 2012); https://chronicle.com/article/College-Degrees-Designed-by/132945/
5. Prewitt, K. The U.S. decennial census: Politics and political science. *Annual Review of Political Science 13*, 1 (May 2010), 237–254.
6. Raman, B. The rhetoric of transparency and its reality: Transparent territories, opaque power and empowerment. *The Journal of Community Informatics 8*, 2 (Apr. 2012).
7. Singer, N. InBloom student data repository to close. *The New York Times* (Apr. 22, 2014), B2; http://nyti.ms/1PUYyIq

Jeffrey Johnson (Jeffrey.Johnson@uvu.edu) is Interim Director of Institutional Effectiveness and Planning at Utah Valley University in Orem, UT.

Calendar of Events