Data Information Literacy and Application

By now, many people in higher education have at least heard of data management, due in large part to various federal funder requirements of a data management plan for all grant proposals. But in the classroom, if this topic is covered at all, the discussion is often concerned with how students should document the data that they produce in the course of research. This is a valuable skill and one that all students should build into their practice. However, equally important is the need for students to understand how data can be reused, what sort of quality assurances should be made on the data, and which data is implicitly gathered and reused all the time by virtue of our online interactions.

All of these concerns fall under the large umbrella of data information literacy. The term data literacy is sometimes used interchangeably, but it has a historical definition relating to statistical and numerical fields (e.g., understanding standard deviations or how to read a graph). Data information literacy (DIL) has a more expansive definition and concerns the activities of the data creator and consumer. In an effort to create standard approaches to DIL, members of the Data Information Literacy project (http://www.datainfolit.org/), an Institute of Museum and Library Services (IMLS) grant-funded initiative, investigated the DIL needs of researchers and developed a curriculum to address those needs. They also identified twelve competencies associated with DIL: cultures of practice, metadata and data description, data management and organization, data curation and reuse, data ethics and attribution, data conversion and interoperability, data preservation, data processing and analysis, data quality and documentation, data visualization and representation, databases and data formats, and data discovery and acquisition.¹

Competencies like data management, organization, and interoperability factor heavily in the data creation process, as do data preservation and curation. A solid foundation in these areas can help ensure that data is available for reuse by data consumers, who in turn should be well acquainted with other competencies, such as data discovery, attribution, and quality. Students who engage with this ecosystem of competencies would—hopefuly—gain understanding into the role that data plays in the research lifecycle and scholarly communication.

Data generation is prolific, thanks to workflows and capture systems made possible by current and evolving technologies and an increased focus on data-driven decision making. In a lab or research setting, this data is often explicitly collected or created. Data is downloaded from instrumentation, recorded in note-books or software, or pulled from repositories for analyses and manipulation. Good DIL practice would ensure that at the various points in the data lifecycle—for example, creation, documentation, annotation, analysis, assurance, preservation—students and researchers are actively involved. But what about the data that, even though it informs decision making and shapes society, is not explicitly collected or created? What about the data that exists in our lives, rather than in research projects or laboratories?

Our Distant Data

Every day, those of us who interact with the Internet or use networked technology are releasing data to various entities. Data about our shopping and reading habits, our financial information, our location—some of the most personal and precious data about who we are as individuals—is collected, stored, aggregated, analyzed, and sold by and to corporations across the globe. More often than not, people have no idea which parts of their interaction data are being used, nor do they have a mechanism to restrict that use while still engaging with the online service. This is in part due to the onerous and confusing Terms of Service (ToS) that so many of us click to accept without reading.

ToS are written as legal documents, not for a layperson. Major services like Google, Facebook, and Apple do not provide simple synopses for people to understand the agreement they are entering into. Certain websites—like Terms of Service; Didn’t Read (https://tosdr.org/) and tl;dr Legal (https://tldrl.org/-) attempt to provide clarity of ToS statements, while tech reporters and bloggers do the same in their columns.² But for these explanations to be useful, people need to (1) realize that blindly accepting ToS may be unwise and (2) know that there are resources like these out there.

Unfortunately, in the United States there is no authority to mandate that these companies simplify their ToS. In some European Union countries, privacy commissions monitor companies to ensure compliance with consumer protection laws, which include provisions that companies not place an undue burden on the consumer.¹ Turning back to DIL, how could the competencies help mitigate these issues? The “distant relationship” between the individual and that individual’s data—a relationship created by social media and other online interactions—affects both the producer and the consumer roles. The DIL competency that concerns ethics and attribution relates to the intellectual property, confidentiality, and privacy issues around sharing and using data. In my experience, however, most of the training related to those issues concerns health data or other...
sensitive data in the research sphere. How much instruction are students receiving on this topic with respect to routine online interactions?

Student Engagement

Certainly, some colleges and universities are considering the ethical concerns around accessing and using student data as learning analytics and other evaluative metrics. There has also been some effort to educate students on data privacy in the K-12 education system. In higher education, however, concentrations in personal data use and privacy fall into computing, cybersecurity, or law curricula. Students studying marketing or media may learn about the ethical issues related to social media, but the ethical competency related to DIL is not well integrated into general education.

This presents an opportunity for academic libraries and their learning partners. Whereas certain DIL competencies, like data preservation and analysis, may be best learned while handling data, the ethical competency can be bundled with other information literacy strategies, such as the ACRL Framework for Information Literacy for Higher Education (http://www.ala.org/acrl/standards/ilframework). Like the DIL ethics competency, the frame “information has value” has often been applied to highlight the ethical use of information and to create awareness of the economic models that influence information access. Yet it also has a focus on the individual as information creator and active participant. It would not be difficult to parley this into a lesson on the “distant data” that students produce and give away or trade for services.

The myriad ways that we produce and consume data—in research, in learning, and in leisure—can make it difficult to determine how best to scope this instruction. A recent article by Megan Sapp Nelson helps structure scaffolded instruction in DIL. Although Nelson does not explicitly address teaching the ethical complexities that may exist with our distant data, the topic certainly could be addressed in the “personal information management” tier of instruction.

The Result

In some ways, the trading of our data in online interactions occurs as unconsciously as breathing. We do not see packets of data leaving our machines and going into a large bucket of other data, where some process occurs and money falls from the bottom, into a corporate wallet. Yet that is essentially what is happening. Moreover, there may be other entities watching our data and tracking our movements and decisions purely from the data that is produced. One of the core principles of information literacy has been critical evaluation. Essentially, that is what is underlying the DIL ethical competency: not just the critical evaluation of the data collected in the course of research, and its suitability to sharing and the risks associated with collecting it, but the critical evaluation of the systems that we work/live within and the data that those systems collect and use.

There are surely convenience benefits associated with the machine learning that occurs on distant data. Netflix recommendations and Amazon sale alerts are some examples. More education for students on how their data supports these systems does not preclude individuals from using those systems. Rather, it helps demystify the domain and provides students with some level of agency in the data exchange. As individuals become more aware of how critical their data is to the global marketplace and to the intelligence industry, they may feel more enfranchised. Here is an opportunity to inoculate students with that lifelong learning mentality. As they consider the data ownership and privacy limitations that they may be (unwittingly) accepting, maybe some of them will recognize the ethical quandary. Perhaps a DIL intervention across curricula will result in a more engaged populace, ready to interrogate the systems that capture our distant data, whether in the classroom, in a research environment, or in society at large.

Notes


Yasmeen Shorish (shorisyl@jmu.edu; Twitter: @yasmeen_azadi) is Data Services Coordinator and Associate Professor at James Madison University.

© 2017 Yasmeen Shorish. The text of this article is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.