This forum is dedicated to personal health in all its many facets: decision making, goal setting, celebration, discovery, reflection, and coordination, among others. We look at innovations in interactive technologies and how they help address current critical healthcare challenges. **— Yunan Chen, Editor**

Who Does the Work of Data?

Naja Holten Møller, University of Copenhagen, **Claus Bossen,** Aarhus University, **Kathleen H. Pine,** Arizona State University, **Trine Rask Nielsen**, University of Copenhagen, **Gina Neff,** University of Oxford

Many people are involved in making large-scale data, yet only some of the tasks involved are getting attention from researchers or recognition by the managers who are reorganizing the data-driven workplace. Despite the emergence of occupations like data analyst and data scientist, much of the work that makes data analysis, interpretation, and responsible use possible happens in administrative or clerical jobs. As a result, this work is often not recognized as vital to producing quality data. New kinds of data and new kinds of uses of data mean that people in traditional roles are working with data in new ways, requiring new skills and knowledge. But these tasks and competencies in existing occupations have been undervalued and slow to come to scholars' attention.

This type of work is called *data work*. Research on data work turns the focus to the sociotechnical practices of producing and using data. Data workers help with the interpretation and contextualization of data, ensure that results are fair and inclusive, and communicate with multiple stakeholders about the data, including information about its context and the privacy concerns raised [1]. Data workers produce data, but they are also required to do additional work adding to and combining datasets, interacting with data, and helping data move to different departments and contexts [2,3]. Data work is increasingly demanded of clerical and administrative workers in a whole range of organizations. Yet the practices of data work are often invisible to managers, and data-work tasks get neither the necessary resources nor the proper compensation.

Scholars, designers, managers, and data workers alike should be concerned about the lack of attention to data work [4]. Consider the healthcare industry. In both Denmark and the U.S., the countries where we studied data work in practice, as well as in healthcare organizations around the world, data-driven approaches to healthcare promise new opportunities to monitor and manage healthcare services, improve service quality, and use data-intensive science for research to advance medicine. But the clerical work that forms critical components of the social and organizational data infrastructure is often missing from the discussion. In Denmark, for example, the business case for a new electronic health record (EHR) system was based on the

Insights

→ Data work carried out by nonclinical hospital workers is complex, skillful, and effortful. It ensures, for example, that documentation is meaningful and ready for coding and analysis.

→ Data work is also concerned with the human and ethics. It interfaces with people's questions, rights, and concerns about data and includes the tasks of managing and mitigating those concerns.

→ When designing the future workplace, data workers must make their work visible so it can be protected and supported.

assumption that medical secretaries over time would become obsolete. As a result, newly designed hospitals did not incorporate any dedicated physical space for medical secretaries, who have long been vital to hospital workflow and whose data-work tasks will remain, even with a new electronic medical records system. We argue that it is essential to legitimize the tasks necessary for the ecosystems of data collection, analysis, interpretation, and ongoing meaning-making, especially when that work is done by workers with lower status in the workplace.

WHO DOES THE DATA WORK OF HOSPITALS?

We studied data work in hospitals in Denmark and the U.S., where efforts to design future hospitals created an opportunity to examine how assumptions about data work also change the physical workplace [4,5]. Consider the data work of clerical workers in both contexts.

In Denmark, a medical secretary sees that a patient checked herself in with the hospital's automatic registration system and updates the physician and nurse programs of the day accordingly. This way, the physician will know that a patient has arrived, and the nurse seeing the patient afterward can prepare for the tests that will follow. The secretary checks the patient's contact information upon arrival. As the patient moves through the clinic's workflow, her status is updated digitally with color codes, and workflows are coordinated as other patients arrive to have tests or routine checkups. In the quiet moments at work, the medical secretary verifies

ICD-10 codes entered into the chart by the physician and makes sure to add the mandatory information about the visit. This could include logging the patient's trajectory to ensure the hospital meets the conditions set forth in patients' rights frameworks.

In the U.S., a clinical documentation improvement specialist (CDIS) reviews a physician's documentation for a patient still in the hospital and, realizing that the coder will not be able to code the chart for an appropriate ICD-10 code related to a patient's development of sepsis, sends a query to the physician. The physician receives the query and reviews her documentation, adding a detail to the patient record in a specific format that is now codeable. Once the patient leaves the hospital, a coder finalizes the code set, entering an ICD-10 code related to sepsis. Elsewhere in the hospital, a quality analyst in the hospital system is working with a special information systems team to query the hospital's database for patients with sepsis, and crafting a slide deck to present to hospital administration to explain the hospital's rise in sepsis cases, using data from the ICD-10 field of the patient's records.

In both of these cases, data work is much more than filling out empty text fields using available information. Data work unfolds in complex ways across three dimensions:

• *Data work as meaningful registration.* Data work ensures, for example, that documentation is meaningful and ready for coding and analysis.

• *Data work as digital organizing/ infrastructuring.* Data work entails the tasks that tie digital infrastructures together across databases and systems.

• *Data work as concern for the human and ethics.* Data work interfaces with people's questions, rights, and concerns about data and includes the tasks of managing and mitigating those concerns.

Based on our research, we developed a toolkit to help designers, scholars, workers, and other stakeholders identify, surface, and value data work. We think this toolkit may help data workers advocate for their roles and tasks in discussions about the future of work within their organizations. The

toolkit's questions allow data workers, as experts in their own work practices, to engage in debates about data work and data stewardship.

A PROTOTYPE FOR INQUIRY INTO DATA WORK

Taking such insights as our starting point, we worked with hospital medical secretaries and a Danish union representing a large segment of Danish clerical workers to develop a tool for bringing out the assumptions that affect decisions on the design of future data-driven hospitals. In particular, we wanted to translate and bring back prior theoretical insights from research to data workers. Concretely, the Data

Data work is much more than filling out empty text fields using available information.

Work Wheel (Figure 1) poses questions that can enable a balanced and informed dialogue about the assumptions around how organizations do and do not change as their services and products become data driven.

In our cases, this toolkit helped us understand the complexity of organizational work around data. For example, data is often assumed to be quantitative, but hospital data often provides information in mixed forms, such as a diagnostic code and text ("A40 Sepsis caused by streptococci") or images. We also learned that data work is not placeless, and can sometimes rely on physical location. In Danish hospitals, medical secretaries sat in close proximity to clinicians, patients, and their relatives. Physical workflows were continuously "mirrored" digitally in order for the data to be trusted, valuable, and actionable in practice.

We identified three sets of questions that stakeholders can use for surfacing data-work tasks across the data
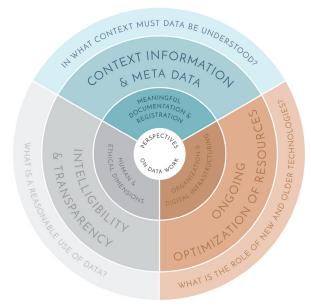
- **Meaningful documentation and registration**
  o What is your role in documenting and making services and activities measurable?
  o What is your role in ensuring that data is correct, complete, and meaningfully registered?
- **Digital organizing and infrastructuring**
  o What is your role in the rollout and retirement of new and older technologies?
  o What is your role in the continuous optimization of the daily organization of work?
- Human and ethical dimensions
  o What is your role in ensuring that data is being used reasonably across purposes?
  o What is your role in translating digital knowledge to others and ensuring that concerns for the human are reflected digitally?

Figure 1. The Data Work Wheel helps make visible the data work that may otherwise go unnoticed when new digital systems and services are implemented in organizations or workflows are changed.

ecosystem. In identifying answers to these, stakeholders can map the gaps and opportunities to better support data work.

Question 1: Who ensures that data is meaningful? Most often data is not enough in itself. The necessity of metadata and context for understanding data goes back to the early research into knowledge-sharing technologies in organizations. Context is crucial for the social processes of making knowledge across organizations. Context is also crucial for us to judge the quality of data. There is a broad consensus in research that documentation is not a trivial task and that understanding what "good" data is depends on whether the context it's produced in, including the known differences and similarities with the context in which it's to be used, can be identified. In other words, documentation work directly influences how people assess the quality of data. In hospitals, for example, people will

judge the quality of data produced for diagnostic purposes differently from reimbursement data. Our research shows how data (still) only becomes meaningful information through human interaction [5].

However, the demand for data that can be used across several purposes is increasing. This requires more metadata and more documentation work. In U.S. healthcare, this means that doctors complain about the increased workload of documentation that has come with digitization. In Denmark, such data is being used to monitor patients' rights to diagnostics and treatment in a timely manner.

Question 2: Who does the work of digital organizing and infrastructuring? Data work most often is part of a larger network of people and technologies that together form the sociotechnical infrastructure for data. Data is fundamentally dependent on the existence of the social organization around it and the work

practice of *infrastructuring*. The social and technical infrastructures for data must be built and are most often based on preexisting systems, which must be adapted. The introduction of new technologies often results in existing tasks disappearing and changing hands, or new ones coming in, and that process involves a large number of negotiations and decisions about which functions and professions should do less or more. This is one of the reasons why the introduction of EHR systems has been difficult.

Data infrastructures must also be maintained on an ongoing basis, and this is true for the social and organizational components of data infrastructure. This work is often undervalued. At the same time, it is important not only to focus on the rollout of new technologies, but also to remember that old technologies must be carefully rolled back—or rolled in—to avoid the collapse of the sociotechnical infrastructure [6]. A deep understanding of infrastructural decay is thus crucial to ensuring that the associated work is also resolved after the retirement of technologies.

Question 3: Who will ensure attention to human and ethical concerns? Paying attention to the human and to data ethics is another critical role of data work. New ethical dimensions emerge with new types of data. Data workers are central to the task of ensuring that data is used reasonably and for the appropriate purpose. They interface between analytics teams and stakeholders, and thus can help translate how data was produced. Such professionals can also help people understand how data about them is produced and used [4]. Data workers play a key role in the explainabilty of complex analytics systems. At both the technical and social levels, the limits of explainability are increasing as more and new types of data are used. Explainability as a concept in relation to data work is about knowing where data comes from and whether the conclusions we draw based in data are reasonable.

Typically, we perceive explainability in healthcare in relation to a given result. In particular, it is debated whether there should always be a human who can account for a given decision and what forms a meaningful explanation. This question of the ethical dimensions and explainability is perhaps the most open and unresolved question in the latest research in data work.

WORKERS IN DATA-DRIVEN ORGANIZATIONS
While we developed the Data Work Wheel by closely observing hospital work, we think that this toolkit has the potential for wider application across many types of organizations where paraprofessional, administrative, or clerical workers support new kinds of data systems. By creating a toolkit of questions to ask, we hope to help people surface and highlight vital but taken-for-granted tasks that make up data ecosystems. The Data Work Wheel toolkit may help others understand data work in a wider variety along the lines of 1) meaningful documentation and registration, 2) digital organizing and infrastructuring, and 3) human and ethical dimensions. We hope it may also be a participatory tool that will allow scholars, designers, managers, and workers alike to understand their assumptions about who does the work of data.

ACKNOWLEDGMENTS

Endnotes
1. Bossen, C., Pine, K.H., Cabitza, F., Ellingsen, G., and Piras, E.M. Data work in healthcare: An introduction. *Health Informatics Journal 25*, 3, 2019.
2. Pine, K.H., Wolf, C., Mazmanian, M. The work of reuse: Birth certificate data and healthcare accountability measurements. *Proc. of the IConference.* iSchools, 2016.
3. Ismail, A. and Kumar, N. Engaging solidarity in data collection practices for community health. *Proc. of the ACM Hum-Comput. Interact. 2*, CSCW (Nov. 2018).
4. Møller, N.H., Bjørn, P.; Villumsen, J.C., Hancock, T.H., Aritake, S., and Tani, T. Data tracking in search of workflows. *Proc. of the 2017 ACM Conference on Computer-Supported Cooperative Work and Social Computing.* ACM, 2017, 2153−2165.
5. Neff, G., Tanweer, A., Fiore-Gartland, B., and Osburn, L. Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data 5*, 2, 2017, 85−97.
6. Cohn, M. Convivial decay: Entangled lifetimes in a geriatric infrastructure. *Proc. of the ACM Conference on Computer Supported Cooperative Work and Social Computing.* ACM, New York, 2016.

⬤ **Naja Holten Møller** is an assistant professor in the Department of Computer Science, University of Copenhagen, Denmark. She is also a member of ACM's Future of Computing Academy. Her recent work "Citizens' Strategies for Exercising Authority and Personal Autonomy" (2019) addresses ethics in practice when data is reused across healthcare and social welfare.
→ naja@di.ku.dk

⬤ **Claus Bossen** is an associate professor in the Department of Digital Design & Information Studies at Aarhus University, Denmark. His research focuses on data work in a healthcare context across the U.S. and Denmark. Most recently, he has organized the special issue on Data Work in Healthcare in the *Health Informatics Journal* (2019).
→ clausbossen@cc.au.dk

⬤ **Kathleen (Katie) Pine** is an assistant professor in the College of Health Solutions at Arizona State University. She conducts interpretivist qualitative research and mixed-methods, community-engaged research on technology-in-use, data practices, and patient work in health and healthcare. Her work has been published in venues such as ACM CHI, ACM CSCW, and the *Academy of Management Journal*.
→ khpine@asu.edu

⬤ **Trine Rask Nielsen** is a research assistant in the Department of Computer Science, University of Copenhagen, Denmark. She comes from a design background and works with speculative methods for gaining insights into emergent uses of large-scale data and algorithms for decision support in the context of healthcare.
→ trn@di.ku.dk

⬤ **Gina Neff** is associate professor and senior research fellow at the Oxford Internet Institute and the Department of Sociology at the University of Oxford, U.K. She studies the future of work in data-rich environments. She is the author of *Venture Labor* (MIT Press, 2012) and, with Dawn Nafus, *Self-Tracking* (MIT Press, 2016).
→ gina.neff@oii.ox.ac.uk