Who does the work of data?

Naja L. Holten Møller, Claus Bossen, Kathleen H. Pine, Trine Nielsen & Gina Neff.

Many people are involved in making large-scale data, and only some of these tasks are getting attention from researchers or recognition by managers re-organizing the data-driven workplace. New occupations like 'data analyst' and 'data scientist' have emerged in recent years, but 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 good 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 results are fair and inclusive, and communicate with multiple stakeholders about the data, including information about its context and the privacy concerns raised (Bossen et al. 2019). 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 (Pine et al. 2016; Pine, Bossen et al. 2018). Data work is increasingly demanded of clerical and administrative workers in a whole range of organizations, and yet the practices of data work are often invisible to managers and data work tasks are provided with neither the necessary resources nor compensation.

Scholars, designers, managers, and workers alike should be concerned about the lack of attention to data work (Møller et al. 2017). 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 discussion. In Denmark, for example, the business case for a new electronic health record system was based on the 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 been vital to the workflow in hospitals and whose tasks of data work will still remain with a new electronic medical records system. We argue that it is essential to attend to how 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 (Møller et al. 2017; Neff et al. 2017). 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 afterwards 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 do their routinely check-up. In the quiet moments at work, the medical secretary verifies ICD10-codes entered into the chart by the physician and makes sure to add the necessary information about the visit, including logging the patients' 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 code-able. 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 powerpoint deck to present to hospital administration to explain the hospital's rise in sepsis cases, using data from the ICD-10 field of the patients' 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.

1. *Data work as meaningful registration*. Data work ensures, for example that documentation is meaningful and ready for coding and analysis.
2. *Data work as digital organizing/ infrastructuring*. Data work entails the tasks that 'tie digital infrastructures together across databases and systems.
3. *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 part of Danish clerical workers to develop a tool for bringing out the assumptions that manifest as 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 Work Wheel 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".

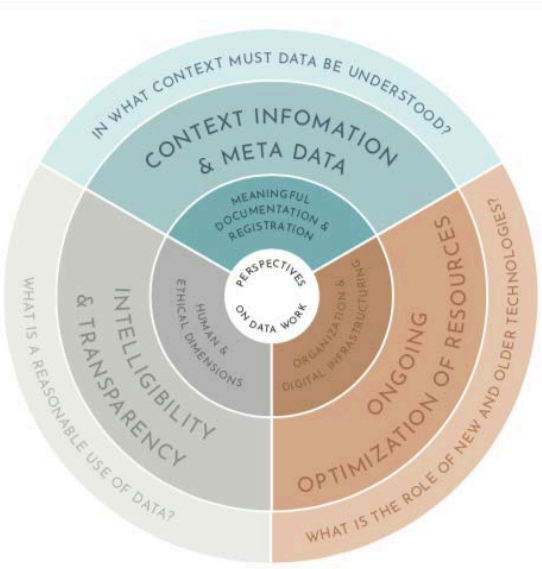

Figure1: The "Data Work Wheel" helps make visible the data work that may otherwise go unnoticed when new digital systems and services are implemented in organizations or workflows are changed.

- **Meaningful documentation and registration**
  - What is your role in documenting and making services and activities measurable?
  - What is your role in ensuring that data are correct, complete and meaningfully registered?

- **Digital organizing and infrastructuring**
  - What is your role in the roll-out and retirement of new and older technologies?
  - What is your role in the continuous optimization of the daily organization of work?

- **Human and ethical dimensions**
  - What is your role in ensuring that data are being used reasonably across purposes?
  - What is your role in translating to others digital knowledge and ensuring that concerns for the human is reflected digitally?

In our cases, this toolkit helped us understand the complexity of organizational work around data. For example, data are 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, can sometimes rely on physical location. In Danish hopsitals 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 ecosystem. In identifying answers to these, stakeholders can map the gaps and opportunities to better support data work.

Question 1: Who Ensures that Data are Meaningful?

Most often data is not enough in itself. The necessity of meta-data 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 is ‘good’ data depends on whether the context they are produced in, including the known differences and similarities with the context in which they are 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 than reimbursement data. Our research shows how data (still) only becomes meaningful information through human interaction (Neff et al. 2017).

However, the demand for data that can be used across several purposes is increasing. This requires more meta-data, 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 are 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 socio-technical infrastructure for data. Data are fundamentally dependent on the existence of the social organization around them and the work practice of *infrastructuring*. The social and technical infrastructures for data must be built and are most often based on pre-existing 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 are to do less or more. This is one of the reasons why the introduction of electronic health records 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 roll-out of new technologies, but also to remember that old technologies must be carefully rolled back—or rolled in—to avoid collapse of the socio-technical infrastructure (Cohn 2016). A deep understanding of infrastructural ‘decay’ is thus crucial to ensure that the associated work is also resolved after the 'retirement' of technologies.

Question 3: Who Will Ensure Concerns for the Human and Ethics?

Concerns for the human and 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 data is used reasonably and for the appropriate purpose. They interface between analytics teams and stakeholders and thus can help translate how data were produced and can help people understand how data about them is produced and used (Møller et al. 2017). Data workers also play a key role in 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 varity 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 and practitioners alike to understand their assumptions about who does the work of data.

Acknowledgements

This piece draws on the work of many prior studies that were critical for pushing recent agendas on data work, and which we could not all cite. A more complete list of studies is cited in the publications below. This work was supported by the Danish union HK Kommunal organizing clerical workers amongst others.

[1] Bossen, C., Pine, K. H., Cabitza, F., Ellingsen, G., & Piras, E. M. (2019). Data work in healthcare: An Introduction.

[2] Cohn, M. (2016). Convivial decay: Entangled lifetimes in a geriatric infrastructure. In Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing.

[3] Møller, N.H; Bjørn, P; Villumsen, J.C; Hancock, T.H; Aritake S. & Tani T. (2017). Data tracking in search of workflows. In proceedings of the 2017 ACM conference on Computer-Supported Cooperative Work & Social Computing, 2153-2165.

[4] Neff, G; Tanweer, A., Fiore-Gartland, B., & Osburn, L. (2017). Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big data*, *5*(2), 85-97.

[5] Pine, K.H; Wolf, C; Mazmanian, M. (2016). The work of reuse: Birth certificate data and healthcare accountability measurements. In *Proceedings of the IConference*.

[6] Pine, K.H; Bossen, C; Chen, Y; Ellingsen, G; Grisot, M; Mazmanian, M; Møller, N.H. (2018). Data Work in Healthcare: Challenges for Patients, Clinicians and Administrators. Companion of the ACM Conference on Computer Supported Cooperative Work and Social Computing, 433-439.