



AARHUS UNIVERSITY



Cover sheet

This is the accepted manuscript (post-print version) of the article.

The content in the accepted manuscript version is identical to the final published version, although typography and layout may differ.

How to cite this publication

Please cite the final published version:

Bonde, M., Bossen, C., & Danholt, P. (2019). Data-work and friction: Investigating the practices of repurposing healthcare data. *Health Informatics Journal*, 25(3), 558-566.

<https://doi.org/10.1177/1460458219856462>

Publication metadata

Title:	Data-work and friction: Investigating the practices of repurposing healthcare data
Author(s):	Morten Bonde, Claus Bossen, Peter Danholt
Journal:	Health Informatics Journal
DOI/Link:	10.1177/1460458219856462
Document version:	Accepted manuscript (post-print)

© The Author(s) 2019

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Data-work and friction: Investigating the practices of repurposing healthcare data

Morten Bonde*, Claus Bossen**, Peter Danholt**

* VIVE – The Danish Center for Social Science Research, Denmark, mbok@vive.dk

** Digital Design & Information Studies, Aarhus University, Denmark, clausbossen@cc.au.dk; pdanholt@cc.au.dk

Abstract

The focus on digital data for improved management and quality of healthcare is paramount. In particular, the vast volumes of accumulated data in clinical systems has created high hopes for repurposing data to serve secondary purposes beyond the practices of direct clinical care, such as research, improvement and efficiency. This paper contributes with an understanding of the pivotal, but often unnoticed ‘data-work’ involved in such efforts. The paper is based a regional project in Danish healthcare, in which nine hospital departments were given the task of developing new indicators for quality to substitute the previous accountability regime based on Diagnosis-Related Groups. Using the concept of ‘friction’, we analyze the challenges of turning clinical ideas into data-supported indicators and of collecting data from existing repositories. Especially, we turn attention to the interaction between clinicians and it-personnel to focus on the interdisciplinary and collaborative aspects of this work.

Introduction

The focus on data for managing and governing society, organizations and work practices is paramount and the healthcare sector is no exception (1,2). Within healthcare, attention in particular is directed at collecting existing data from clinical systems, and utilizing them for other, secondary purposes (repurposing) such as quality improvement, efficiency, safety, research and analysis. Allegedly, repurposing the vast volumes of existing data is key to maximizing the benefits of digitized, clinical systems, and for transforming and improving healthcare (3–5). While the potentials may be great and many, there is only very limited research into the concrete practices of repurposing healthcare data. Therefore, the overall aim of this paper is to contribute to an understanding of ‘data-work’ in healthcare. Data-work can designate any work associated with acquiring, filtering, storing, calculating, visualizing and making sense of data. However, in this paper we focus on practices of repurposing data.

We draw on insights and concepts from data-studies (6,7). Here, a central observation is that data is never “raw” factual resources, but always “cooked”, that is recorded, ordered, formatted, and processed for specific purposes (6). Consequently, data is not in itself a resource for knowledge, but must be ‘(re)-cooked’ carefully to become useful knowledge for a specific practice. Furthermore, as Edwards has shown in his study of global knowledge infrastructures, data is material and physical, and the use of data across contexts, practices and technologies inevitably entails ‘friction’ (7). In his analysis, Edwards defines various kinds of friction are identified and the concept thus appears potent for analyzing practices of data production and (re-)use. Most basically ‘data friction’ denotes “the costs in time, energy, and attention required simply to collect, check, store, move, receive, and access data” (7). Closely related, ‘computational friction’ denotes the work of combining and turning data into useful information. ‘Metadata friction’ occurs when information about data is missing or is irreconcilable (7), and ‘science friction’ relates to communicative and collaborative issues when different professional cultures interact around data (8). In addition to this set of concepts, Boyce introduce the term ‘second-order friction’ to denote a resistive force that occurs when systems built for one purpose are repurposed to underpin ‘second-order systems’ (9, see also 10). In her analysis, Boyce unfolds social and political consequences of connecting heterogeneous data and infrastructures, thereby demonstrating the analytical value of friction in relation to practices of repurposing data.

In the present paper, we take these notions of friction to imply that repurposing health data is not simply a matter of applying data-management tools and techniques in order to maximize the benefits of accumulated health data. Rather, achieving such benefits come with the prize of managing frictions between data, clinical systems and the various practices and purposes they underpin. Morrison et al. (11) provides a poignant example of this, describing how the introduction of a system for repurposing data from electronic patient records (EPR) imposed an extra workload on clinicians; additional data entries had to be made manually, and existing data had to be manipulated to be ‘repurpose-able’. In addition to the extra workload, issues about the responsibility for managing all this data friction arose among the clinicians. This kind of data-work is at crucial for successful repurposing of health data, but often it remains ‘invisible’ and goes unnoticed and unaccounted for (12).

Therefore, in this paper we take up the notion of ‘friction’ to become attentive to the strenuous and less evident work efforts required for repurposing data in healthcare. However, inspired by Tsing (13), we take friction not only as analytical marker of a troublesome and resistive force. Friction is also generative, like the heat generated when rubbing two sticks together (13). Taking this into account, the concept helps us to identify and surface the invisible work of repurposing data, as in Morrison et al.’s account, as well as also to become attentive to, and appreciative of, how the mundane work of managing friction might be generative in itself. Thus, our contribution aligns with Morrison et al. account, with the addition of the generative aspect of friction. Furthermore, whereas the focus in Morrison et al.’s study is primarily on clinicians and their responses to the extra workload caused by a repurposing agenda, the focus in this paper is mainly on the side of IT-technical personnel and their role in repurposing data for improved quality and management of healthcare. We focus on three types of processes and practices: 1) the role of technical personnel in operationalizing quality indicators defined by clinicians 2)

how IT-infrastructures and interfaces generate different expectations about the retrievability of data 3) how the use of data produces further data-work.

Case and Methods

Our analysis is based on our research into a governance experiment from 2015 to 2017 initiated and promoted by one of the five regions in Denmark in charge of secondary healthcare. The aim of this experiment was to transform the governance of healthcare from a focus on activity and production to a focus on value and cost-effectiveness. The experiment involved nine hospital departments that were asked to propose and develop new indicators to measure, improve and account for quality and patient-value, hoping that this would be conducive to a value-based healthcare system. To make the process of indicator development feasible, the departments were encouraged to mainly repurpose data already available, rather than developing indicators requiring new data. Over the next months, the nine departments thus engaged in developing indicators on the expectation that data could be retrieved from existing IT systems, notable the EPR. During the three-year project, half-annually reports were worked out by the region on the performance of the nine departments based on the chosen indicators. This paper focuses on the processes of defining and implementing new indicators by repurposing existing data and infrastructures.

The empirical material consists of 27 semi-structured interviews with managers, head doctors and head nurses from the nine departments and two interviews with technical personnel with degrees in computer science from the region's Business Intelligence (BI) unit responsible for operationalizing indicators. Interview guides were designed by the authors, and focused on how the heads of departments perceived and approached the governance experiment, and the process of developing and using the indicators. Interviews with technical personnel focused on the challenges of operationalizing clinician's ideas of indicators in data. The interviews lasted between 1-1½ hours and were conducted in two rounds in 2015 and 2016. Furthermore, the authors have observed meetings between the departments and regional officials. Interviews were transcribed and coded using qualitative software by all three authors using grounded theory techniques (14). Codes were derived from the thematic questions of the interview guide, and interviews were blind-coded by two of the authors to reveal and resolve inconsistencies in code application and ensure inter-coder reliability in a final application of codes. For the purpose of this paper, we have been through the material again focusing on questions of indicators and data, as well as the interviews with the BI staff.

Analysis

The following analysis falls in three parts focusing on 1) the interactions between clinicians and technical personnel in the development of indicators 2) the challenges of retrieving data for indicators, and 3) how data generates further data and classification work.

Data-work between clinicians and technical personnel

Developing indicators was not simply a matter of what should be measured considered from the viewpoint of good medical quality, but also a matter of what was ‘doable’ in terms of data and IT-architecture. Indicator development therefore entailed a close collaboration between technical and medical personnel, in which the indicators were refined and defined more and more precisely.

The BI employees were regularly in contact with the clinicians in the course of identifying indicators, what data should be included, and how results were to be presented. A central BI employee explained how his starting point in these discussions and negotiations typically was to clarify logical inconsistencies in the ways indicators were defined by clinicians. For example, in an effort to reduce the number of non-attending patients, the heads of a center defined their indicator and goal as; ‘the number of non-attendances should decrease by 20 percent’. At a first glance, this goal might seem straightforward and logical, but for the data-worker, it implied computational friction with the data, because the indicator was mix of absolute and relative criteria:

“Logically, this goal does not make sense. How should I compare a decrease in percentage points with a decrease in numbers of non-attending patients? Is it the non-attendance percentage they want to decrease by 20 percent? That does not make sense. ... Therefore, in these cases I have to say [to the clinicians]: ‘You need to explain what it is that you want. Do you want the number to decrease by 20%? And would that be 20% annually, or in the next 2 years, or do you really mean, that the number of non-attending patients shall decrease one fifth in eternity? What is your baseline? Is it 2013?’” (Interview 1, BI-employee)

The example demonstrates how seemingly simple, straightforward and well-defined indicators entailed friction with the data they should be developed from. It required clarification-work to attain the strict, logical quality that is necessary for an indicator to be implemented. As the data-worker put it “...I actually have to tell the data how to behave...” meaning that logical consistency was a prerequisite for indicators to be implementable and to overcome the computational friction between data and indicator.

Besides ensuring logical consistency, data-workers also played an active part in making indicators clinically meaningful and useful. As such, data-work did not only rest on technical skills, but on knowledge and insight into how clinicians work with data. For example, one department had proposed an indicator measuring whether post-operative plans were kept to schedule or not. Based on former experience and insight into how physicians make use of such data, the data-worker singlehandedly split this indicator up into 11 sub-indicators based on different patient groups. As he explained:

”...we know very well that it is useless if we simply show for example that ‘72% adhered to the postoperative plan’. Because then they [clinicians] want to know; ‘Well, is there any difference across the patient-groups? And which patient groups is it that do not adhere to the plans?’” (Interview 2, BI-employee)

This demonstrates how making indicators clinically useful was also the result of data-workers’ initiatives and knowledge of clinical work. In general, data-workers such as the BI-employee played an active part in the processes of getting from an idea of an indicator

to a final definition of the indicators. For example, the management of five departments wanted to measure the number of patients in ambulatory care that were registered as 'completed' in the EPR-system. In the initial definition, this indicator was conceived of as a measure of clinical quality: Using the EPR system it should be calculated how many patients were fully treated and by good health. However, as the data-worker explained, the data in the EPR do not provide the possibility to measure whether patient trajectories in the system are completed, because the patients recovered, or, for example, because the patients transitioned to palliative care. This underscores the point that data is not in itself a knowledge resource, as they are cooked for specific purposes, which complicates the repurposing of the data. Through a dialogue with the BI employee on the challenges of measuring the number of recovered patients, the department managers then began to reconceive the purpose of this indicator. Rather than a measure of clinical quality counting the number of recovered patients, the indicator was conceived of as a means to make sure that patients, whose actual course of treatment had been completed, were also registered as 'completed' in the EPR-system. This in order to avoid that patients in fact completed still remained digitally active in the EPR-system, and would be called in mistakenly for unnecessary control examinations. The point here is that the dialogue with the data-worker was not merely about how to support the indicator with data, but also a process in which the indicator was reconceived as means for process management (making sure that patient charts do not remain open in the EPR-system), instead of a measure for clinical quality as it was initially conceived. Thus, in this case friction was not simply mitigated or managed to achieve a predetermined goal. Rather, the friction was also generative as it was part in causing the goal to be reconceived.

The data-worker reflected in more general terms on his role in developing indicators and was critical of the perception that he as technical personnel was simply an implementer of the quality indicators defined by physicians:

“We are not simply programming-machines who receive an input from clinicians, and then we go like ‘tack-tack-tack’ [imitating a machine], and then we return with a number to them. It is most definitely a dialogue, back and forth, ... After having worked some years in the healthcare system, I have gained a rather detailed insight to the kinds of things that interest them [clinicians] ... So that is also how we were able to sort of direct them towards certain kinds of indicators, that we were able to support with data. However, the idea about [clinicians] ordering indicators and us implementing them, that does not hold. Not at all.” (Interview 2, BI-employee)

Hence, he underscored the necessary cooperative aspects of data-work across professions, and the relational dynamism between technical and clinical competencies as key when repurposing healthcare data. Our point is not that technical personnel or others should be granted more responsibility for defining clinical measures, but merely to stress that these interdisciplinary relations and processes are necessary and productive.

Underscoring the dynamism between data-workers and clinicians is different from the view by Donabedian (15) that data-work competencies should be developed and internalized as a part of certain clinical professions. Donabedian argues that successful development and use of quality measures requires what he labels a ‘clinical performance epidemiologist’ “possessing both clinical competencies and skills to gather, analyze, and

interpret the data quality assurance depends on.” (p. 118). Ideally, such a hybrid profession of both clinical, performance management and data-work competencies might be feasible and productive in certain cases. However, we would argue that it is perhaps equally beneficial and more immediately feasible to facilitate close and iterative collaboration between clinicians and data-workers, and the competence to be developed is for both clinicians and technicians to engage in these sorts of relations. Furthermore, one should be careful not to centralize data-work in designated organizational units, as this might hamper the iterative dialogue between technicians and clinicians. This became evident in our case, when it was decided during the project to centralize the data-work and technical personnel in a business intelligence unit, which then discontinued the close back and forth dialogue that had played an important role for the wards. In summary, rather than concentrating data-work in certain clinical persons and organizational units, we argue for a distributed understanding of data-work for repurposing data.

Practices of data production vs. data extraction

The easiness with which digital data can be produced and distributed may lead to the perception that data is an immaterial resource that can be easily collected and redistributed to support organizational development through continuous flows of data. In the present case, this idea was evident on the side of clinicians, whereas from the data-workers perspective identifying and retrieving data is far from trivial. According to the data-worker, most clinicians assumed that the entries they made in EPR's could be easily identified and retrieved as a basis for developing indicators. However, it turned out that there was a substantial difference between the situated practice of entering data in the EPR-system, to the actual extraction, formatting and storage of those data in the main data-warehouse of the Region. The EPR-interface was not a reflection of how data are stored and did not provide much help in terms of extracting data. Making an entry in the EPR was entirely different from extracting that exact data-object from the data-warehouse. As the data-worker explained, this difference between interface and data-storage is incomprehensible and thus invisible for most users of data systems including the clinicians. Instead, the assumption on the clinicians' side was that knowing their data-entries in the EPR-system, was assumed sufficient in order to extract the data from the data-warehouse:

“They say: ‘Those entries we make in the EPR, can't you just give us them?’ But it's a data warehouse! They assume that the things they register are easy and straightforward for me to extract” (Interview 1, BI-employee)

For the data-workers, data must be extracted via database-queries, and contemplating what queries to make in order to identify specific data-entries is an entirely different and complicated task. Another, similar issue occurred when two departments wanted to make their data-entries in clinical quality databases the basis for new performance indicators. This required a more frequent release of these data than the half-year reports they at the time received from the database administrators. Again, this involved data friction: Since these clinical databases were not owned by the Region, accessing the data required manual login, thus making an automatic retrieval of data cumbersome. An interface that could automatize data extraction was required, but this was out of budget limits of the

Business Intelligence unit, which in turn was a great disappointment to the departments. Consequently, one of the departments had a secretary manually extract the relevant data from the database when needed.

Thus, being attentive of data-work could help mitigate unrealistic expectations by management, physicians and other healthcare staff as to the ease with which data can be identified and extracted. However, the main point here is that the vast production of data in the clinic is not readily available to clinicians in the way one might assume, since entering and extracting data are two different issues and require different kinds of data-work, the latter involving a lot of data-friction in terms of data access and collection.

Data producing work

A common understanding of data is that it has the potential to produce better and more efficient care (16,17). However, in the following example, we see how data produces work and holds potentials for learning.

One of the wards was interested in decreasing cancellations of surgery. They had an initial feeling that cancellations were a substantial problem for the ward as they decrease efficiency and are costly in more than one respect: fewer patients are treated and resources are not optimally used. To better understand the extent of the problem, the ward asked for data. However, what was assumed to be readily available data was not so. The head physician phrased it in the following way:

“What has been challenging is to figure out what cancellations actually are. And how they are registered. Much of this has to do with IT technical stuff, because the way we register a cancellation in our IT-system ...with an ‘Y’ on the list. However, when you go in and look at a cancellation in detail [...] then it could be the patient calling up to say “I do not want surgery after all”. So there is a big difference whether it's the patient's wish or simply because a surgeon has fallen ill [...] *But it's the same ‘Y’, right!* ... There are also instances where we, due to other cancellations, suddenly have time to offer a patient a new and earlier time for surgery, and then that patient also gets a “Y” in the system.”
(Head physician)

The EPR system can provide data, but they lack detail: Cancellations as a general category is not of much use to the ward, and more fine-grained information is required in order to understand and act upon cancellation data. Only if cancellations and their causes and consequences can be further specified, will meaningful and data-informed action be possible. Some cancellations are actually even positive as when moving surgeries forward to earlier timeslots. Consequently, the ward began scrutinizing cancellations and classifying them to identify and reduce the types of cancellations deemed negative.

The example goes counter to the assumption often associated with data that existing data can be utilized in a given practice and may lead to improvement and efficiency. Here instead, asking for data promoted scrutiny of data, analysis of what data represented, and the construction of a new classification of data. Existing data was not translated into useful information and action, because it entailed computational friction and produces work. It is important to note that the ward regarded this as an inevitable and relevant

process, because they would learn more about their practice and about cancellations. Thus, the example also demonstrates the generative side of friction: The creation of new insights about cancellations and their multiple causes. In sum, due to the overall project agenda concerned with promoting quality in healthcare and the initial assumption about addressing a seemingly simple problem, cancellations became unfolded as a rather complex and multifaceted object. By engaging with cancellation data, it multiplied into a variety of different types of cancellations. Consequently, the data-work produced extra work as more had to be learned about the different types of cancellations in order to achieve the quality agenda, but this was also a productive outcome in itself.

Conclusion

This article adds to the empirical and conceptual understanding of the less evident practices of data-work. To this end, we have drawn on the concept of friction to identify and analyze the work in nine Danish hospital departments of repurposing data from clinical systems into indicators for better quality management and improvement. Friction as suggested by Tsing points at the “awkward, unequal, unstable, and creative qualities of interconnection across difference” (13, p4), and within data-studies Edwards and others use the concept to point at the work necessary to establish interconnections between differences of hardware, databases, data format, and disciplines (6, 7, 8, 9).

First, we have analysed how the data-work implied in developing indicators consisted in the mutual collaboration between clinical personnel and BI employees, and we argued that such interdisciplinary collaboration is crucial for repurposing data and developing digital infrastructures in healthcare. Second, with the concept of data friction, we pointed out how existing infrastructures and data repositories often considered readily available resources, in practice require effort from data-workers in order to identify and retrieve data. Third, we pointed out that data in itself produces work, but that this work may be considered valuable in itself as it generates learning and knowledge production.

This article contributes with insights to the crucial, but often invisible work of repurposing healthcare data. As in the ethnographic account provided by Morrison et al. (11), we also found that physicians have to invest extra work to repurpose data. However, our empirical focus is different, highlighting especially the data-work of BI employees in addition to that of clinicians. More importantly, whereas Morrison et al. depict the data-work imposed on clinicians as burdensome, our analysis suggests that this is only one aspect of data-work, and that managing friction may be generative, not only by serving predetermined, managerial goals, but also by enabling clinicians in the wards in unforeseen ways.

We find that the concept of data friction implies an important *double perspective* on data, which we think is crucial for developing data infrastructures in reflective and sustainable manners. Such a perspective enables us to consider *simultaneously* the *cumbersome* and *generative* aspects of data and avoid the equally single-minded perspectives on data as either mainly optimistic or pessimistic. Data friction and the empirical accounts analyzed in this article contributes in our understanding to much needed modest, balanced and multifaceted understanding of data infrastructures and their potentials. There are no easy answers or fixes in complicated practices and digitally saturated environments, only close

and detailed reflections and negotiations between different practitioners, concerns and human and technological agents.

References

1. Hogle LF. Data-intensive resourcing in healthcare. *BioSocieties*. 2016 Sep 1;11(3):372–93.
2. Hoeyer K. Denmark at a Crossroad? Intensified Data Sourcing in a Research Radical Country. In: *The Ethics of Biomedical Big Data* [Internet]. Springer, Cham; 2016 [cited 2018 Mar 23]. p. 73–93. (Law, Governance and Technology Series). Available from: https://link.springer.com/chapter/10.1007/978-3-319-33525-4_4
3. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *J Am Med Inform Assoc JAMIA*. 2007;14(1):1–9.
4. Garrett D. Tapping into the value of health data through secondary use: as electronic health records (EHRs) proliferate across the nation, an important new opportunity awaits healthcare organizations that can find meaningful commercial uses for the data contained in their EHR systems. *Healthc Financ Manage* [Internet]. 2010 Feb 1 [cited 2018 Mar 23]; Available from: <http://link.galegroup.com/apps/doc/A225073847/AONE?sid=googlescholar>
5. Pricewaterhouse Coopers. *Transforming Healthcare Through Secondary Use of Health Data*. PricewaterhouseCoopers' Health Research Institute; 2010.
6. Gitelman L, editor. *'Raw Data' Is an Oxymoron*. Cambridge, Massachusetts ; London, England: The MIT Press; 2013. 192 p.
7. Edwards PN. *A vast machine: computer models, climate data, and the politics of global warming*. First paperback edition. Cambridge, Massachusetts London, England: The MIT Press; 2013. 518 p. (Infrastructures series).
8. Edwards P, Mayernik MS, Batcheller A, Bowker G, Borgman C. *Science Friction: Data, Metadata, and Collaboration*. *Soc Stud Sci*. 2011 Aug 15;0306312711413314.
9. Boyce AM. Outbreaks and the management of 'second-order friction': Repurposing materials and data from the health care and food systems for public health surveillance. *Sci Technol Stud* [Internet]. 2016 Feb 12 [cited 2018 Mar 23]; Available from: <https://sciencetechnologystudies.journal.fi/article/view/55409>
10. Greenhalgh T, Potts HWW, Wong G, Bark P, Swinglehurst D. Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method. *Milbank Q*. 2009 Dec;87(4):729–88.

11. Morrison C, Jones M, Jones R, Vuylsteke A. 'You can't just hit a button': an ethnographic study of strategies to repurpose data from advanced clinical information systems for clinical process improvement. *BMC Med.* 2013 Apr 10;11:103.
12. Bowker GC, Star SL. *Sorting Things Out: Classification and Its Consequences.* The MIT Press; 2000. 389 p.
13. Tsing AL. *Friction: An Ethnography of Global Connection.* Princeton University Press; 2005. 348 p.
14. Glaser B, Strauss A. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Aldine Transaction; 1967. 271 p.
15. Donabedian A. *An Introduction to Quality Assurance in Health Care.* Oxford University Press; 2002. 233 p.
16. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst.* 2014;2:3.
17. Wasan S, Bhatnagar V, Kaur H. The impact of data mining techniques on medical diagnostics. *Data Sci J.* 2006;5:119–26.