UNIVERSITY OF OSLO

IN5210: Data Work

Troels Sune Mønsted
Associate Professor
Department of Informatics

October 15th, 2025



Agenda

- Intensified datafication of organizations and the importance of understanding the impact on data work.

- 2 perspectives on data work:
  - What work does it take to make data work? The concept of data stewardship (Parmiggiani et al, 2022).
  - What work does intensified datafication create and what are the sociotechnical implications? (Hoeyer, 2023).



Credit: Dall-E

Data work

Core definition:

- *”(…) any human activity related to **creating**, **collecting**, **managing**, **curating**, **analyzing**, **interpreting**, and **communicating** data (Bossen, Pine, Cabitza, & Piras (2019)".*

- Or in other words: **"The work that makes data work"**

- Some characteristics:
  - Difficult to define, **emergent** type of work.
  - A **craft**, requires creativity, discretion, and improvisation.
  - **Interpretive** labor.
  - **Collaborative**, involves **various actors**.

Mixed theoritical underpinnings:

- STS (Science and Technology Studies) heritage, e.g. mutual shaping of work and technologies (Pinch & Bijker, etc).

- Empirically driven studies, e.g. in CSCW producing rich empirical insights but with limited development of theory.

- Critical studies: Critique of how data work challenges or reproduces social, political, and economic power structures.

Why the data work perspective?

Questions the data work perspective may help answer:

- What types of work does datafication create?
- Who does this work?
- What skills does it require?
- What tools are needed?
- How is the work organized?
- Etc.

Intensified datafication of organizations

- Increasingly large volumes of digital data is produced and stored (sometimes referred to as big data).
- Increased expectations to the benifits that can be drawn from data, for instance in optimization, decision-making, product development, service provision, monitoring, research etc.
- To gain these benefits a core challenge is to harness the 4 Vs: the volume, variety, velocity, and veracity of data (Raghupati & Raghupati, 2014). Someone needs to do work to achieve this.



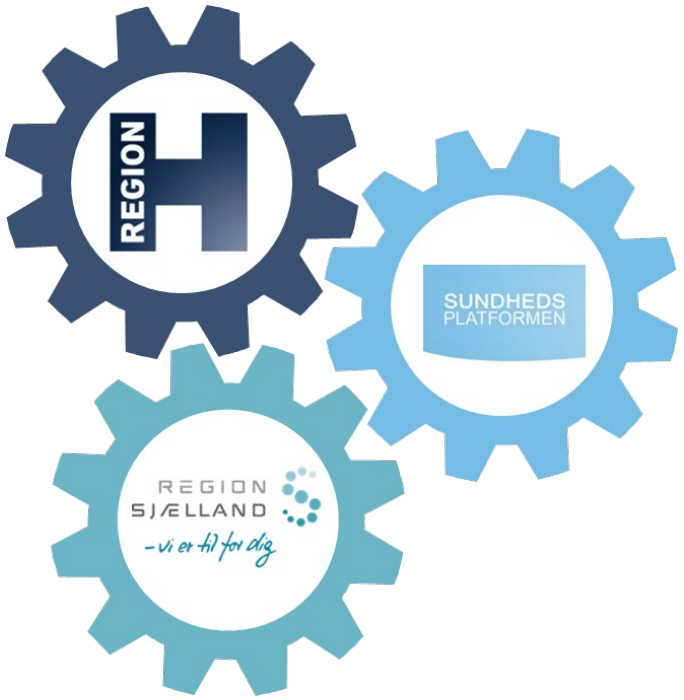
Intensified datafication of Danish Healthcare

The Health Platform (Sundhedsplatformen):

- Large-scale implementation of a configuration of the American EPIC electronic medical record.
- Implemented in 2 (out of 5) regions, covering app. 44.000 end users and 2.5M citizens.

Purpose

- To form a unified infrastructure for work in secondary healthcare (hospitals).
- To improve coordination and collaboration between health providers.
- To improve information and data flow and usage.



Data interests in Danish healthcare

Research (knowledge)

- Research on clinical matters, health services, epidemiology, laboratory methods, big data etc.

Clinicians (health)

- Optimization of treatments, coordination, monitoring of quality, health surveillance etc.

Politicians and administrators (governance)

- Resource allocation, political negotiations, preparation of reforms, goal setting, budget control, accountability etc.

Industry (wealth)

- Product development, R&D, monitoring of drugs, market expansion etc.

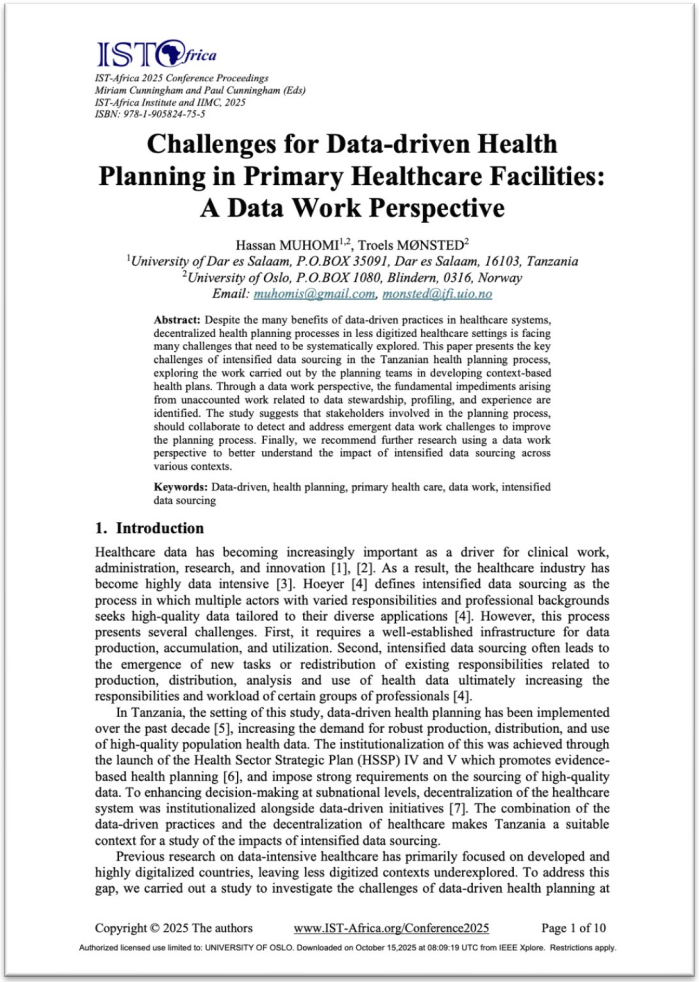
Table 1.1

Purposes for which actors want more data

Actors/Goods	Stated Purposes Claimed to Justify Intensified Data Sourcing
Researchers Knowledge	Clinical research, such as the use of patient data to • Document the effect of changes in clinical procedures and evaluate the effects (and side effects) of treatments • Identify patients for randomized clinical drug trials or experiments with precision medicine Health services research focusing on • Usage of patient data to identify, e.g., patient needs or patterns of inequality • Usage of patient data to assess, monitor, and compare, e.g., different forms of organization Epidemiological research, such as the use of patient data to • Monitor and predict disease burden and distribution • Identify population health problems and potential causes of disease Laboratory research, such as the use of patient data and samples to • Validate new laboratory methods • Carry out -omics research to understand mechanisms of disease Big data research using population data sets, also beyond health-care, to • Identify patient pathways and interactions • Identify patterns of disease or raise new hypotheses based on unexpected correlations
Clinicians Health	Systematic approaches to ensure consistency when dealing with, for example: • Screening, triage and assessment of needs • Follow-up (e.g., discontinuation of drugs, side effects, chronic care, polypharmacy) Optimized treatment options • Use of experiences from similar patients to make targeted treatment plans (e.g., in personalized medicine) • Enhanced monitoring of patients at risk Coordination and easy information exchange • Communication and coordination between clinicians, across units and sectors • Communication between clinicians and patients Monitoring of quality, such as by using patient data to • Identify safety problems (e.g., infection rates) and unusual patterns (e.g., in prescriptions or deaths) • Document progress or achievement of goals Public health surveillance, such as • Pandemic surveillance • Preventive measures
Industry Wealth	such as by • Moving indication thresholds or documenting off-label uses to enhance marketing • Instigating value-based pricing through documentation of effect to differentiate prices To use data as assets in their own right such as by • Delivering data access and analysis to other companies

Hoeyer (2023), p. 37-38

A global phenomenon



IST-Africa 2025 Conference Proceedings
Miriam Cunningham and Paul Cunningham (Eds)
IST-Africa Institute and IIMC, 2025
ISBN: 978-1-905824-75-5

Challenges for Data-driven Health Planning in Primary Healthcare Facilities: A Data Work Perspective

Hassan MUHOMI[1,2], Troels MØNSTED[2]
[1]*University of Dar es Salaam, P.O.BOX 35091, Dar es Salaam, 16103, Tanzania*
[2]*University of Oslo, P.O.BOX 1080, Blindern, 0316, Norway*
Email: muhomis@gmail.com, monsted@ifi.uio.no

Abstract: Despite the many benefits of data-driven practices in healthcare systems, decentralized health planning processes in less digitized healthcare settings is facing many challenges that need to be systematically explored. This paper presents the key challenges of intensified data sourcing in the Tanzanian health planning process, exploring the work carried out by the planning teams in developing context-based health plans. Through a data work perspective, the fundamental impediments arising from unaccounted work related to data stewardship, profiling, and experience are identified. The study suggests that stakeholders involved in the planning process, should collaborate to detect and address emergent data work challenges to improve the planning process. Finally, we recommend further research using a data work perspective to better understand the impact of intensified data sourcing across various contexts.

Keywords: Data-driven, health planning, primary health care, data work, intensified data sourcing

1. Introduction

Healthcare data has becoming increasingly important as a driver for clinical work, administration, research, and innovation [1], [2]. As a result, the healthcare industry has become highly data intensive [3]. Hoeyer [4] defines intensified data sourcing as the process in which multiple actors with varied responsibilities and professional backgrounds seeks high-quality data tailored to their diverse applications [4]. However, this process presents several challenges. First, it requires a well-established infrastructure for data production, accumulation, and utilization. Second, intensified data sourcing often leads to the emergence of new tasks or redistribution of existing responsibilities related to production, distribution, analysis and use of health data ultimately increasing the responsibilities and workload of certain groups of professionals [4].

In Tanzania, the setting of this study, data-driven health planning has been implemented over the past decade [5], increasing the demand for robust production, distribution, and use of high-quality population health data. The institutionalization of this was achieved through the launch of the Health Sector Strategic Plan (HSSP) IV and V which promotes evidence-based health planning [6], and impose strong requirements on the sourcing of high-quality data. To enhancing decision-making at subnational levels, decentralization of the healthcare system was institutionalized alongside data-driven initiatives [7]. The combination of the data-driven practices and the decentralization of healthcare makes Tanzania a suitable context for a study of the impacts of intensified data sourcing.

Previous research on data-intensive healthcare has primarily focused on developed and highly digitalized countries, leaving less digitized contexts underexplored. To address this gap, we carried out a study to investigate the challenges of data-driven health planning at

Copyright © 2025 The authors
Authorized licensed use limited to: UNIVERSITY OF OSLO. Downloaded on October 15,2025 at 08:09:19 UTC from IEEE Xplore. Restrictions apply.

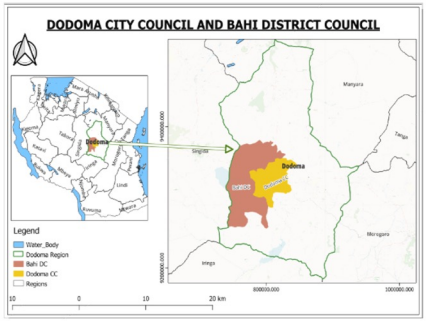
Muhomi & Mønsted (2025)

A global phenomenon

IST frica

Research Objectives

- The research is guided by the key question: What are the key challenges experienced by subnational level teams in the development of data-driven health plans?
  - Particularly additional work resulted from:
    - decentralization and datafication processes
    - Complying with national guidelines and templates
    - Access of information from siloed systems

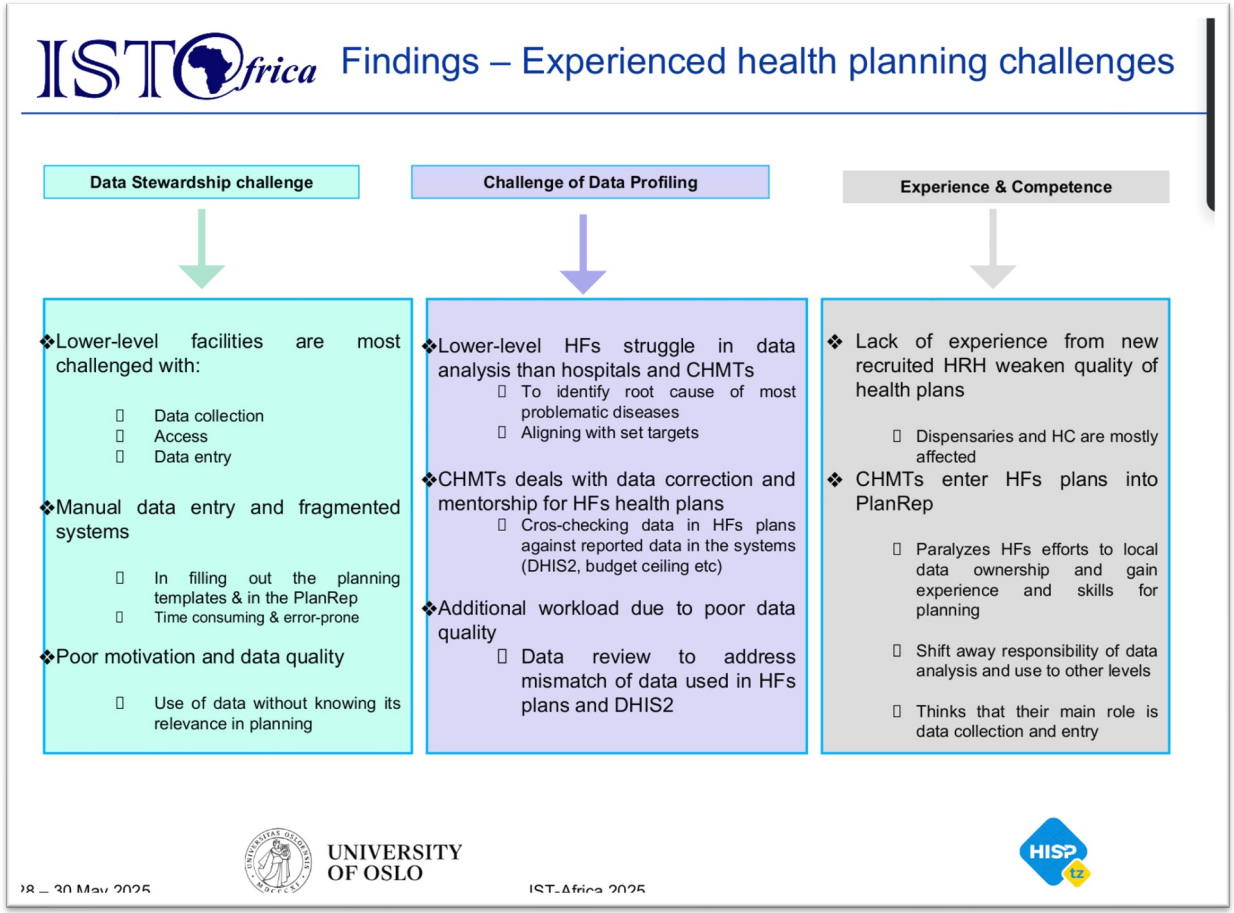


- Case study of 10 primary HFs from 2 councils of Dodoma region
- Involved fieldwork within DoE project
  - Signed MoU to strengthen data management practices, innovation, and research
  - MoH, Ministry responsible for LGA, UiO, HISP, and the two districts

UNIVERSITY OF OSLO

HISP tz

A global phenomenon



Muhomi & Mønsted (2025)

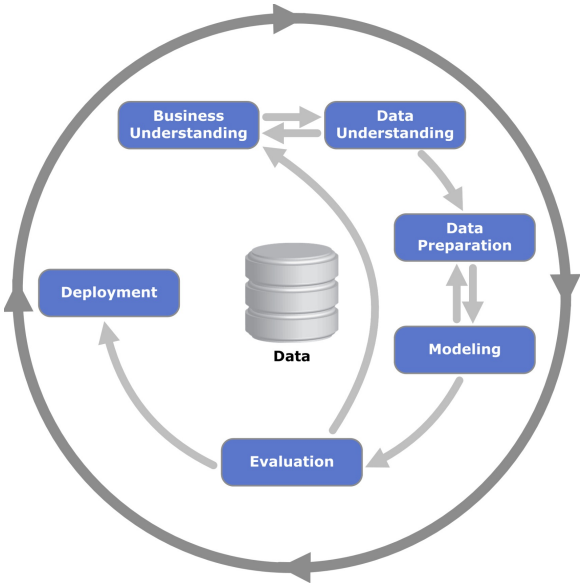
The emergence of the data scientist

Role

- To ensure robust data mining (collection and preprocessing), analysis of data, development of models (e.g. AI and ML), develop visualizations.

Limitations of this perspective

- Does not account for the *backstage* of data science - the *invisible* work and *broader impact* of datafication in organizations.
- In response, data work represents a *practice perspective* on datafication.



CRISP-DM (cross-industry standard process for data mining)

What does data scientists *really* do?



Example: The no-show project

Purpose

- To utilize existing data recorded in the EMR to develop an ML model that was able to predict patient no-shows.
- To develop, evaluate and implement an ML system in a hospital outpatient clinic.

Project organization

- Organized as a PhD project conducted by C. Gyldenkærne based on an ML model developed during his masters' project.
- Co-funded by Roskilde University and Region H (the Capital Region of Denmark).



Int. J. Human-Computer Studies 181 (2024) 103162

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Innovation tactics for implementing an ML application in healthcare: A long and winding road

Christopher Gyldenkærne [a], Jens Ulrik Hansen [a], Morten Hertzum [a,*], Troels Mønsted [b]

[a] Department of People and Technology, Roskilde University, Denmark
[b] Department of Informatics, University of Oslo, Norway

ARTICLE INFO

Keywords:
Innovation tactics
Organizational implementation
Machine learning
Data work
Last mile

ABSTRACT

Artificial intelligence techniques, including machine learning (ML), have shown remarkable test results over the past decade but struggled with the transfer to practical application. The present study applies action research to investigate this last stage of a project to implement an ML algorithm for predicting no-shows at a Danish hospital. We approach the implementation of the no-show algorithm as an innovation process and identify 14 tactics that were employed to provide the innovation necessary at the implementation stage. The tactics span three analytic levels – organization, project, and practice – and alternate between efforts to train the algorithm and to establish trust in its predictions. These efforts are interdependent, highly sociotechnical, and hence blur the boundary between technical development and organizational implementation. They also show the intricacies involved in innovating during ML implementation. Despite sustained support at the organization level, the implementation of the no-show algorithm at the practice level remained unsettled.

1. Introduction

Within healthcare, the use of machine learning (ML) and other artificial intelligence (AI) techniques has attracted widespread attention over the past decade as a way of deriving new benefits from the big pools of data that are continuously produced and accumulated (Panesar, 2021; Qayyum et al., 2021). ML algorithms have shown remarkable results and accuracy during the development and testing phases, for instance in automated image analysis (Poostchi et al., 2018) and the early detection of disease outbreaks (Chen et al., 2017). However, the results have often been confined to the experimental stage and have not been possible to realize in clinical practice. This study contributes to recent research on the innovativeness required of projects that aim to apply ML in practice.

ML research focuses mainly on the technical aspects of developing reliable and accurate algorithms. This research has substantially furthered our understanding of how data can be curated, and algorithmic models built and validated. The challenges of applying ML algorithms in practice have attracted less research attention until the emergence of the discourse on *the last mile* (Coiera, 2019). This discourse pictures the development of AI and ML technologies as consisting of three successive and equal-size efforts: data capture and cleaning (the first mile), model building and testing (the middle mile), and real-world implementation (the last mile). Two chasms must be bridged to achieve successful real-world implementation (Cabitza et al., 2020). The first, referred to as the hiatus of human trust, concerns the challenges related to gaining acceptance of ML in practice, including the resolution of any cognitive dissonance between the system and its users. The second chasm, referred to as the hiatus of machine experience, concerns the challenges involved in ensuring reliable data in amounts sufficient to allow the proper training and continued adjustment of the ML algorithm. That is, the last mile comes with its own challenges; its success does not simply ensue from successfully completing the first and middle miles.

While the discourse on the last mile provides a conceptual framing, it says little about how people, in practice, work to implement and scale ML solutions. Yet, this work is pivotal in realizing innovations, as recognized in studies of digital entrepreneurs (Arvidsson and Mønsted, 2018) and trust in data science (Passi and Jackson, 2018). With this paper, we focus on the tactics applied by digital entrepreneurs working with the implementation of ML in healthcare. The studied tactics were applied to respond innovatively to local circumstances. Empirically, we conducted an action-research study of a project for using ML to predict and reduce patient no-shows at a Danish hospital. The project remained unsuccessful at delivering the intended reduction in no-shows but

* Corresponding author.
E-mail address: mhz@ruc.dk (M. Hertzum).

https://doi.org/10.1016/j.ijhcs.2023.103162
Received 18 March 2023; Received in revised form 19 September 2023; Accepted 28 September 2023
Available online 29 September 2023
1071-5819/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Gyldenkærne, Hansen, Hertzum, & Mønsted (2024): Innovation tactics for implementing an ML application in healthcare: A long and winding road.

Example: The no-show project

Results

- The model showed satisfactory precision on test data: App. 60% prediction rate.
- In real-life settings the prediction rate was significantly lower.

Challenges that showed during deployment

- Incomplete data reporting.
- Inconsistent data reporting.
- Extensive need for data curation and preparation.
- ***Conclusively: Lack of data stewardship***.



Int. J. Human-Computer Studies 181 (2024) 103162

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Innovation tactics for implementing an ML application in healthcare: A long and winding road

Christopher Gyldenkærne [a], Jens Ulrik Hansen [a], Morten Hertzum [a,*], Troels Mønsted [b]

[a] Department of People and Technology, Roskilde University, Denmark
[b] Department of Informatics, University of Oslo, Norway

ARTICLE INFO

Keywords:
Innovation tactics
Organizational implementation
Machine learning
Data work
Last mile

ABSTRACT

Artificial intelligence techniques, including machine learning (ML), have shown remarkable test results over the past decade but struggled with the transfer to practical application. The present study applies action research to investigate this last stage of a project to implement an ML algorithm for predicting no-shows at a Danish hospital. We approach the implementation of the no-show algorithm as an innovation process and identify 14 tactics that were employed to provide the innovation necessary at the implementation stage. The tactics span three analytic levels – organization, project, and practice – and alternate between efforts to train the algorithm and to establish trust in its predictions. These efforts are interdependent, highly sociotechnical, and hence blur the boundary between technical development and organizational implementation. They also show the intricacies involved in innovating during ML implementation. Despite sustained support at the organization level, the implementation of the no-show algorithm at the practice level remained unsettled.

1. Introduction

Within healthcare, the use of machine learning (ML) and other artificial intelligence (AI) techniques has attracted widespread attention over the past decade as a way of deriving new benefits from the big pools of data that are continuously produced and accumulated (Panesar, 2021; Qayyum et al., 2021). ML algorithms have shown remarkable results and accuracy during the development and testing phases, for instance in automated image analysis (Poostchi et al., 2018) and the early detection of disease outbreaks (Chen et al., 2017). However, the results have often been confined to the experimental stage and have not been possible to realize in clinical practice. This study contributes to recent research on the innovativeness required of projects that aim to apply ML in practice.

ML research focuses mainly on the technical aspects of developing reliable and accurate algorithms. This research has substantially furthered our understanding of how data can be curated, and algorithmic models built and validated. The challenges of applying ML algorithms in practice have attracted less research attention until the emergence of the discourse on *the last mile* (Coiera, 2019). This discourse pictures the development of AI and ML technologies as consisting of three successive and equal-size efforts: data capture and cleaning (the first mile), model building and testing (the middle mile), and real-world implementation (the last mile). Two chasms must be bridged to achieve successful real-world implementation (Cabitza et al., 2020). The first, referred to as the hiatus of human trust, concerns the challenges related to gaining acceptance of ML in practice, including the resolution of any cognitive dissonance between the system and its users. The second chasm, referred to as the hiatus of machine experience, concerns the challenges involved in ensuring reliable data in amounts sufficient to allow the proper training and continued adjustment of the ML algorithm. That is, the last mile comes with its own challenges; its success does not simply ensue from successfully completing the first and middle miles.

While the discourse on the last mile provides a conceptual framing, it says little about how people, in practice, work to implement and scale ML solutions. Yet, this work is pivotal in realizing innovations, as recognized in studies of digital entrepreneurs (Arvidsson and Mønsted, 2018) and trust in data science (Passi and Jackson, 2018). With this paper, we focus on the tactics applied by digital entrepreneurs working with the implementation of ML in healthcare. The studied tactics were applied to respond innovatively to local circumstances. Empirically, we conducted an action-research study of a project for using ML to predict and reduce patient no-shows at a Danish hospital. The project remained unsuccessful at delivering the intended reduction in no-shows but

* Corresponding author.
E-mail address: mhz@ruc.dk (M. Hertzum).

https://doi.org/10.1016/j.ijhcs.2023.103162
Received 18 March 2023; Received in revised form 19 September 2023; Accepted 28 September 2023
Available online 29 September 2023
1071-5819/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Gyldenkærne, Hansen, Hertzum, & Mønsted (2024): Innovation tactics for implementing an ML application in healthcare: A long and winding road.

What is data stewardship?

Definition: The collection of practices that ensure that an organization's data is accessible, usable, safe, and trusted.

Key challenges for data stewardship

Data science practices unfold against a sociotechnical context that is…

- Constantly evolving: The social routines and technical systems used to classify, organize, and structure data evolves over time.
- Epistemologically oriented: The purpose is to gain access to a reality that is otherwise inaccessible to humans, based on data that often have different meaning and/or relevance for different professionals.
- Grounded in informal networks: Problem-solving heuristics are often developed by professionals who rely on informal social networks.



Journal of the Association for Information Systems (2022) **23**(1), 139-164
doi: 10.17705/1jais.00718
RESEARCH ARTICLE
ISSN 1536-9323

In the Backrooms of Data Science

Elena Parmiggiani[1], Thomas Østerlie[2], Petter Grytten Almklov[3]
[1]Department of Computer Science, Norwegian University of Science and Technology, Norway, parmiggi@ntnu.no
[2]Department of Computer Science, Norwegian University of Science and Technology, Norway, thomas.osterlie@ntnu.no
[3]Department of Sociology and Political Science, Norwegian University of Science and Technology, Norway, petter.almklov@ntnu.no

Abstract

Much information systems research on data science treats data as preexisting objects and focuses on how these objects are analyzed. Such a view, however, overlooks the work involved in finding and preparing the data in the first place, such that they are available to be analyzed. In this paper, we draw on a longitudinal study of data management in the oil and gas industry to shed light on this backroom data work. We find that this type of work is qualitatively different from the front-stage data analytics in the realm of data science but is also deeply interwoven with it. We show that this work is unstable and bidirectional. That is, the work practices are constantly changing and must simultaneously take into account what data might be possible to access as well as the potential future uses of the data. It is also a collaborative endeavor involving cross-disciplinary expertise that seeks to establish control over data and is shaped by the epistemological orientation of the oil and gas domain.

Keywords: Data Work, Data Science, Practice, Performative Perspective

Matthew Jones was the accepting senior editor. This research article was submitted on January 10, 2020 and underwent two revisions.

1 Introduction

With the surge in big data and datafication, data science is now recognized as a central discipline in the design and implementation of digital transformation strategies throughout organizations and society (Abbasi et al. 2016; Agarwal & Dhar, 2014; Galliers et al., 2017). Therefore, different professional roles have coalesced into organizational groups responsible for handling the discovery, transformation, modeling, and reporting of large and complex data sets in enterprises under titles such as "data scientist" or "data analyst" (Davenport & Patil, 2012; Kandel et al., 2012).

Researchers in information systems (IS) and related fields have taken a number of perspectives on data science, including technical (Liebman, Saar-Tsechansky, & Stone, 2019; Shi et al., 2016; van der Aalst, 2016), management-oriented (Baesens et al., 2016; Dhar, 2013; Kane, 2014; Mikalef et al., 2020; Saar-Tsechansky, 2015), and performative (Aversa et al., 2018; Jones, 2019; Østerlie & Monteiro 2020; Monteiro & Parmiggiani 2019). Although all three perspectives acknowledge that the work of data science requires a combination of analytical skills and domain-specific knowledge, the first two tend to assume that data remain ontologically separate from the practices that produce and manage them (Provost & Fawcett, 2013, p. 52). Researchers taking these perspectives are inclined to espouse—often implicitly—what Kitchin (2014) refers to as a pre-factual take on data—that is, a view on data as "meaningful independent of format, medium, language, producer, and context [and] a key element in the constitution of . . . facts, evidence, information, and knowledge" (p. 3; see also Abbasi et al., 2016). The consequence of this take on data is that the actual work tasks carried out by individuals to curate and set up the data are typically downplayed (Carter & Sholler, 2016). However, ongoing datafication makes a case for

Parmiggiani, Østerlie, & Almklov (2022):
In the backrooms of data science.

What is data stewardship?

Key tasks for data scientists in practice

- Data discovery: Getting the right data:
  - Knowing the data ecosystem.
  - Knowing historical data-ordering modes.
  - Data usage understanding.
- Data preparation: Getting the data right:
  - Ensuring file formatting consistency.
  - Ensuring cross-platform formatting consistency.
  - Assessing data provenance.
  - Data gatekeeping.



JAIS
ISSN 1536-9323
Journal of the Association for Information Systems (2022) **23**(1), 139-164
doi: 10.17705/1jais.00718
RESEARCH ARTICLE

In the Backrooms of Data Science

Elena Parmiggiani[1], Thomas Østerlie[2], Petter Grytten Almklov[3]
[1]Department of Computer Science, Norwegian University of Science and Technology, Norway, parmiggi@ntnu.no
[2]Department of Computer Science, Norwegian University of Science and Technology, Norway, thomas.osterlie@ntnu.no
[3]Department of Sociology and Political Science, Norwegian University of Science and Technology, Norway, petter.almklov@ntnu.no

Abstract

Much information systems research on data science treats data as preexisting objects and focuses on how these objects are analyzed. Such a view, however, overlooks the work involved in finding and preparing the data in the first place, such that they are available to be analyzed. In this paper, we draw on a longitudinal study of data management in the oil and gas industry to shed light on this backroom data work. We find that this type of work is qualitatively different from the front-stage data analytics in the realm of data science but is also deeply interwoven with it. We show that this work is unstable and bidirectional. That is, the work practices are constantly changing and must simultaneously take into account what data might be possible to access as well as the potential future uses of the data. It is also a collaborative endeavor involving cross-disciplinary expertise that seeks to establish control over data and is shaped by the epistemological orientation of the oil and gas domain.

Keywords: Data Work, Data Science, Practice, Performative Perspective

Matthew Jones was the accepting senior editor. This research article was submitted on January 10, 2020 and underwent two revisions.

1 Introduction

With the surge in big data and datafication, data science is now recognized as a central discipline in the design and implementation of digital transformation strategies throughout organizations and society (Abbasi et al. 2016; Agarwal & Dhar, 2014; Galliers et al., 2017). Therefore, different professional roles have coalesced into organizational groups responsible for handling the discovery, transformation, modeling, and reporting of large and complex data sets in enterprises under titles such as "data scientist" or "data analyst" (Davenport & Patil, 2012; Kandel et al., 2012).

Researchers in information systems (IS) and related fields have taken a number of perspectives on data science, including technical (Liebman, Saar-Tsechansky, & Stone, 2019; Shi et al., 2016; van der Aalst, 2016), management-oriented (Baesens et al., 2016; Dhar, 2013; Kane, 2014; Mikalef et al., 2020; Saar-Tsechansky, 2015), and performative (Aversa et al., 2018; Jones, 2019; Østerlie & Monteiro 2020; Monteiro & Parmiggiani 2019). Although all three perspectives acknowledge that the work of data science requires a combination of analytical skills and domain-specific knowledge, the first two tend to assume that data remain ontologically separate from the practices that produce and manage them (Provost & Fawcett, 2013, p. 52). Researchers taking these perspectives are inclined to espouse—often implicitly—what Kitchin (2014) refers to as a pre-factual take on data—that is, a view on data as "meaningful independent of format, medium, language, producer, and context [and] a key element in the constitution of . . . facts, evidence, information, and knowledge" (p. 3; see also Abbasi et al., 2016). The consequence of this take on data is that the actual work tasks carried out by individuals to curate and set up the data are typically downplayed (Carter & Sholler, 2016). However, ongoing datafication makes a case for

Parmiggiani, Østerlie, & Almklov (2022):
In the backrooms of data science.

Discussion

How can we *account* or *make room* for data stewardship in clinical settings?



Int. J. Human-Computer Studies 181 (2024) 103162

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Innovation tactics for implementing an ML application in healthcare: A long and winding road

Christopher Gyldenkærne [a], Jens Ulrik Hansen [a], Morten Hertzum [a,*], Troels Mønsted [b]

[a] Department of People and Technology, Roskilde University, Denmark
[b] Department of Informatics, University of Oslo, Norway

ARTICLE INFO

Keywords:
Innovation tactics
Organizational implementation
Machine learning
Data work
Last mile

ABSTRACT

Artificial intelligence techniques, including machine learning (ML), have shown remarkable test results over the past decade but struggled with the transfer to practical application. The present study applies action research to investigate this last stage of a project to implement an ML algorithm for predicting no-shows at a Danish hospital. We approach the implementation of the no-show algorithm as an innovation process and identify 14 tactics that were employed to provide the innovation necessary at the implementation stage. The tactics span three analytic levels – organization, project, and practice – and alternate between efforts to train the algorithm and to establish trust in its predictions. These efforts are interdependent, highly sociotechnical, and hence blur the boundary between technical development and organizational implementation. They also show the intricacies involved in innovating during ML implementation. Despite sustained support at the organization level, the implementation of the no-show algorithm at the practice level remained unsettled.

1. Introduction

Within healthcare, the use of machine learning (ML) and other artificial intelligence (AI) techniques has attracted widespread attention over the past decade as a way of deriving new benefits from the big pools of data that are continuously produced and accumulated (Panesar, 2021; Qayyum et al., 2021). ML algorithms have shown remarkable results and accuracy during the development and testing phases, for instance in automated image analysis (Poostchi et al., 2018) and the early detection of disease outbreaks (Chen et al., 2017). However, the results have often been confined to the experimental stage and have not been possible to realize in clinical practice. This study contributes to recent research on the innovativeness required of projects that aim to apply ML in practice.

ML research focuses mainly on the technical aspects of developing reliable and accurate algorithms. This research has substantially furthered our understanding of how data can be curated, and algorithmic models built and validated. The challenges of applying ML algorithms in practice have attracted less research attention until the emergence of the discourse on *the last mile* (Coiera, 2019). This discourse pictures the development of AI and ML technologies as consisting of three successive and equal-size efforts: data capture and cleaning (the first mile), model building and testing (the middle mile), and real-world implementation (the last mile). Two chasms must be bridged to achieve successful real-world implementation (Cabitza et al., 2020). The first, referred to as the hiatus of human trust, concerns the challenges related to gaining acceptance of ML in practice, including the resolution of any cognitive dissonance between the system and its users. The second chasm, referred to as the hiatus of machine experience, concerns the challenges involved in ensuring reliable data in amounts sufficient to allow the proper training and continued adjustment of the ML algorithm. That is, the last mile comes with its own challenges; its success does not simply ensue from successfully completing the first and middle miles.

While the discourse on the last mile provides a conceptual framing, it says little about how people, in practice, work to implement and scale ML solutions. Yet, this work is pivotal in realizing innovations, as recognized in studies of digital entrepreneurs (Arvidsson and Mønsted, 2018) and trust in data science (Passi and Jackson, 2018). With this paper, we focus on the tactics applied by digital entrepreneurs working with the implementation of ML in healthcare. The studied tactics were applied to respond innovatively to local circumstances. Empirically, we conducted an action-research study of a project for using ML to predict and reduce patient no-shows at a Danish hospital. The project remained unsuccessful at delivering the intended reduction in no-shows but

* Corresponding author.
E-mail address: mhz@ruc.dk (M. Hertzum).

https://doi.org/10.1016/j.ijhcs.2023.103162
Received 18 March 2023; Received in revised form 19 September 2023; Accepted 28 September 2023
Available online 29 September 2023
1071-5819/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Gyldenkærne, Hansen, Hertzum, & Mønsted (2024): Innovation tactics for implementing an ML application in healthcare: A long and winding road.

What are the implications of datafication on the sociotechnical context and work practices?

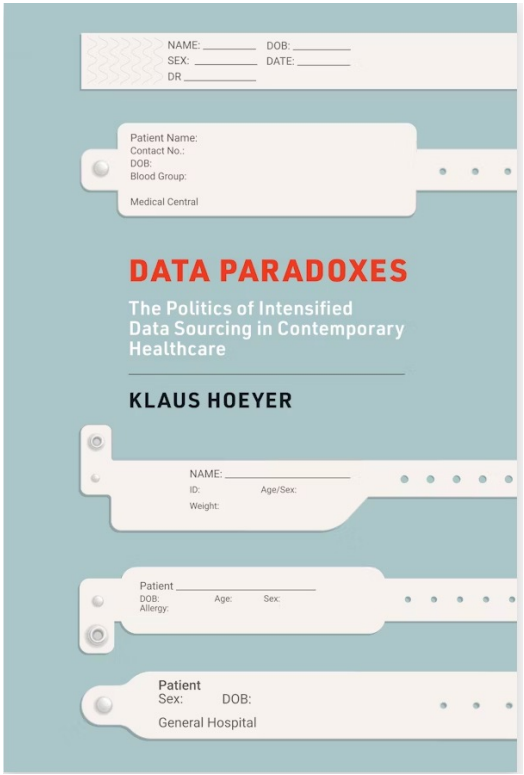


Klaus Hoeyer (2023): Data Paradoxes

- Long-term ethnographic study of the intensified datafication (or *intensified data sourcing*) of Danish healthcare, the associated politics, and the day-to-day impact for health practitioners.

Some observations by Hoeyer

- Intensified data sourcing pushes data work towards health professionals and creates new tasks.
- Datafication creates paradoxes of fundamentally different experiences and opposed interests.



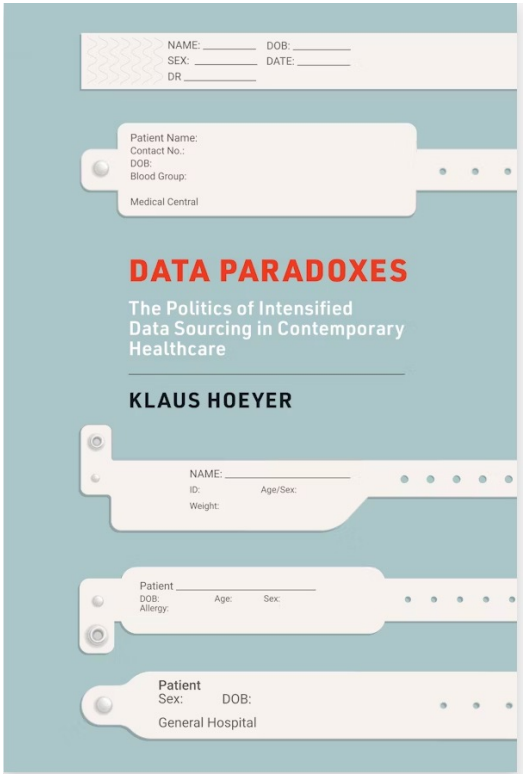
Data paradoxes

Definition

- Data paradoxes is defined as *"(…) situations where ostensibly contradictory stories are both (partly) true. Or, at least, appear equally true to different stakeholders. In short, I am interested in the societal effects of people subscribing to different, but coexisting, truths."*

Examples of paradoxes

- Promises of efficiency vs. increased workload.
- Standardization vs. need for flexibility.
- Data quality vs. the messy reality of data production.
- Tensions between different purposes (research, clinical, administrative.



Four types of data work

Table 3.1
The four main types of work associated with data-intensive healthcare

Data production	The work going into making data through actual documentation (e.g., application of a diagnostic code)
Data analysis	The work that transforms data sets into messages about what they mean (e.g., the making of a table, graph, dashboard or a narrative)
Data instruction	The work involved in governing others with, or being governed by, data
Data use	The work associated with understanding data or data analyses produced by others

Example: NEWS scores

- National Early Warning Score:
- Standardized assessment to inform about early signs of sepsis and level of illness in general
- Collection of data the patient's vital parameters:
  - Respiration rate
  - Oxygen saturation
  - Systolic blood pressure
  - Pulse rate
  - Temperature
  - Level of consciousness or confusion
- Calcutation of risk of sepsis based on total score

Data sources in use during NEWS

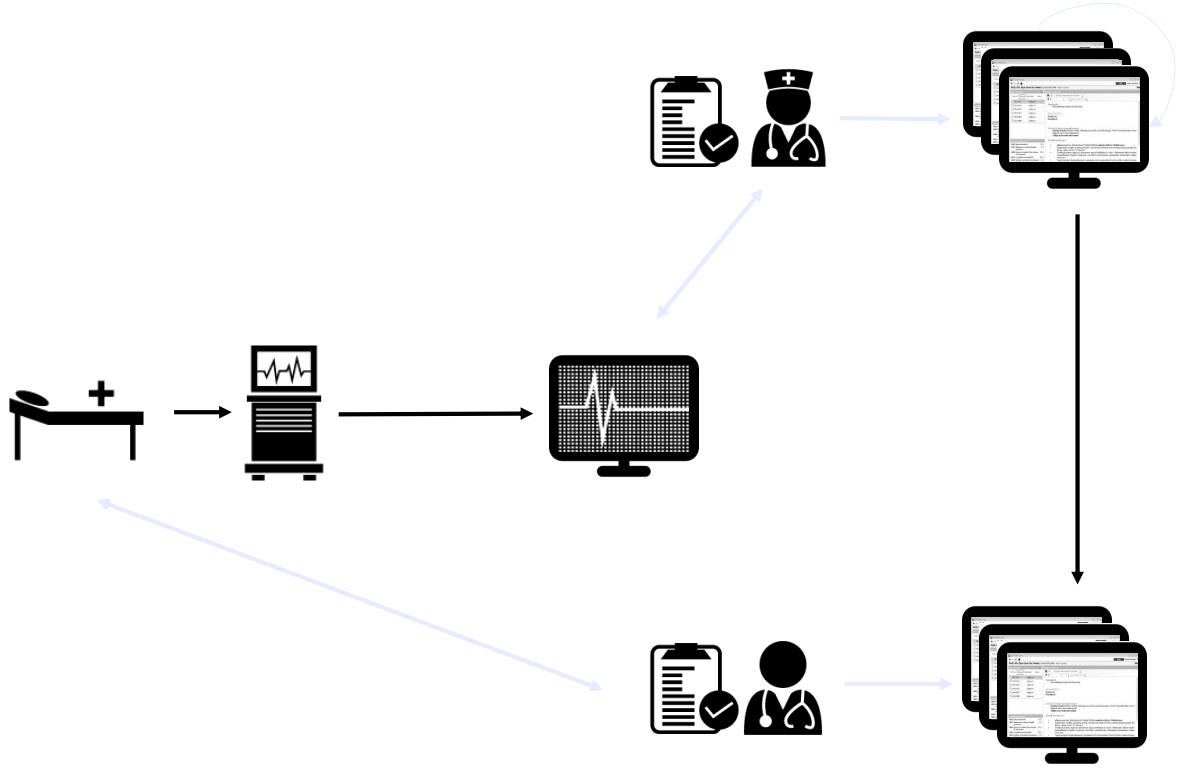






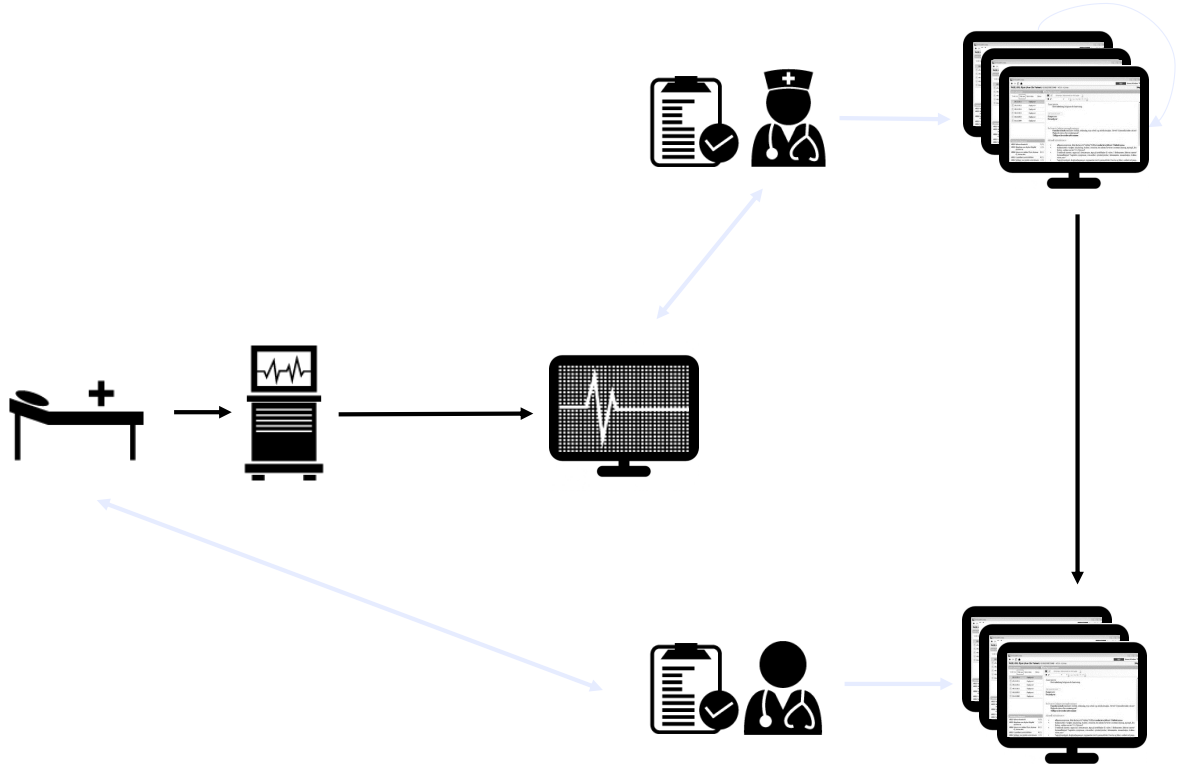
Name	Type	Content	Redundancies
DIPS	EMR	Progress notes, Lab, PACS, PAS, prescriptions.	Pronounced replication of information
Metavision	Medication management and charting	Physical parameters, medication	Replication of data from other HIS
Intellivue	Monitoring	Physical parameters	Duplication (automated and manual) from devices, replication in DIPS and MV

NEWS Score: The formal procedure



Based on: Mønsted, Fossum, & Aanestad (2018).

NEWS Score: The formal procedure



Questions

- What new types of data work were created in this case?
- How does Hoeyers 4 types of data work apply here?
  - Data production
  - Data analysis
  - Data instruction
  - Data use

Based on: Mønsted, Fossum, & Aanestad (2018).

Data work created by the NEWS score:

- Data production: Need for additional data collected manually (temperature, sensing the patient), manual data entry.
- Data analysis: Comparing the NEWS score to the general impression of the patient's condition and previous health history.
- Data instruction: Ensuring that physicians are aware of critical NEWS scores (post-it notes, verbally, telephone).
- Data use: Comparing the NEWS score to the patient's general health condition and past illness trajectory.

Summing up

- Intensified datafication create both opportunities and challenges for organizations.
- In essence, data requires work to work. This work requires new, competences, and expertise.

- The concept of data work can be used as a sensitizing concept to develop our understanding of:
  - The background work of data science.
  - The sociotechnical implications for practice in organizations.

Questions

