Summary of Session 1 (21.09.2023): Making Data, Making Worlds - An Introduction to Data Practices

The Ontology of Data

- Emphasised that data do not exist independently of the practices that create them.
- Introduced the idea that data are "cooked" according to particular recipes, challenging the notion of raw data.
- Discussed the influence of institutional factors on scientific practices.

The importance of data

- Emphasised the ubiquity of data in contemporary society and the need to understand how data is created and how it shapes our world.
- Discussed the impact of data on knowledge production, social order, power relations, ethics and justice.

Big Data Knowledge and Data Science

- Discussed the big data "gold rush" and the V4 model (volume, variety, velocity, veracity).
- Introduced Data Science as the professional study and analysis of data, highlighting its multidisciplinary nature.
- Addressed the assumptions of data science: objectivity, neutrality and the belief that data is a given.

Critical data studies

- Explores the emergence of critical data studies as a field that challenges the positivist assumptions of data science.
- Highlights key perspectives from scholars such as Boyd and Crawford, Dalton and Thatcher, Kitchin and Lauriault, and Iliadis and Russo.
- Emphasised the critical mindset that seeks to understand the power and social ordering effects of data.

Conclusion

The session concluded with questions and the announcement of no class for the following week.

Session 2 Summary (05.10.2023): Making Data, Making Worlds

Housekeeping

- Lunch is welcome as long as it doesn't disturb others.
- 10 min bio-break in each session.

Review of Session 1

- Discuss key takeaways in pairs.
- Emphasis on putting "science in perspective" and developing a "critical mindset".

Understanding Data

- Data are not taken for granted; they are actively produced.
- Production involves choices that shape the form and value of data.
- Data as a powerful resource for knowledge and social order.
- Critical data studies advocate the study of how data is made and remade.

Medicine and Data

- Medical practice, especially in hospitals, as an important site for critical data studies.
- Hospitals produce massive amounts of data about patients, diseases, therapies, etc.
- Emphasis on how doctors create data from ambiguous phenomena, emphasizing the role of knowledge and tools.

Case Study: Atherosclerosis (Body Multiple)

- Study of the development and treatment of atherosclerosis in a Dutch hospital.
- Emphasis on different ways of knowing, healing, and the impact on coordination, distribution, and inclusion.

Atherosclerosis Explained

- Common condition in which plaque builds up in the arteries.
- Consequences: reduced blood flow, potential complications such as a heart attack or stroke.

Diagnosing Disease - Outpatient vs. Pathology

- Different diagnostic approaches in outpatient clinics and pathology.
- Clinical atherosclerosis more important but uncertain.
- Emphasis on the importance of reconciling different diagnostic approaches.

Medical Practice and Diversity

- Mol's emphasis on the study of medical practices and their effects.
- Reality is multiple and varies between different medical practices.

Enactment and Performativity

- Introduction to enactment and performativity in the production of knowledge.
- Methods are performative, they shape realities.
- Ontology is political, influencing which social realities are enacted.

Example of Predictive Policing

- Study of predictive policing in Switzerland and Germany.
- Focus on prediction of domestic burglaries using different databases and algorithms.
- Challenges: reliability, mutation, and decisions affecting real-life consequences.

Conclusion

- Summarize the main points of the session.
- Encourage questions and further exploration.

Session 3 (12.10.2023) - Classification and Standards

Rear-view Mirror:

- Data are tools for understanding the world and making decisions.
- Different tools for producing data can represent the same phenomenon in different ways.
- Data "enact" different realities and influence interventions in the world.
- Responsibilities are linked to how we produce data as they shape the world.

Let's Talk About Death:

- The importance of gathering knowledge about death is discussed.
- Participants work in pairs to discuss how they would gather such knowledge and potential problems.

Knowing Death:

- In the 19th century, disease and death shifted from fate to microbial causes.
- Public health policy relies on understanding the causes of death to take preventive measures.

Tracking Death:

- Challenges in harmonizing the global diagnosis and recording of deaths.
- Cultural and medical differences affect what is considered a legitimate cause of death.

Towards a Systematic Approach:

- Two main challenges identified: building consensus on causes of death and ways of producing data.

The Birth of the ICD:

- The International Statistical Institute adopts the first international classification of diseases in 1893.
- The ICD has evolved over time and is currently in its 11th iteration (ICD-11, published in 2018).

The Purpose of the ICD:

- To systematically record, analyze, interpret, and compare mortality and morbidity data.
- Coding translates diagnoses into alphanumeric codes for easy storage, retrieval, and analysis.
- Multiple use cases beyond health statistics.

ICD Use Cases Include:

- Cause of death certification and reporting.
- Morbidity coding, diagnosis-related grouping, safety surveillance, cancer registries, etc.

The Structure of Today's ICD:

- ICD-11 contains 17,000 diagnostic categories and over 100,000 medical diagnostic index terms.
- It reflects the density of collisions of classification schemes that have evolved over time.

Building Boxes:

- Classification systems involve choices and are not self-evident.
- Two ideal types: Aristotelian (empirical, top-down) and prototypical (discursive).

ICD as a Compromise:

- A pragmatic compromise that embraces complexity and multiplicity.
- Reflects the socio-technical embeddedness of health data.

Death Reporting:

- The challenge of selecting a single cause of death.
- Recognizes the ambiguity of causes of death, especially in cases of multimorbidity.

Standardization of Reporting:

- A globally standardized medical certificate of cause of death takes account of ambiguity.
- Provides space to tell a story and explain a causal chain of different conditions.

Data and Imagination:

- Data must be imagined to exist and function, to be articulated against the seamlessness of phenomena.
- Understanding basic assumptions through the study of classification systems.

Summary:

- Classification systems are constructed in accordance with practical goals and constraints.
- This doesn't diminish the scientific nature of the data generated.
- It highlights the embeddedness of knowledge systems in the real, imperfect world.

Session 4 (17.10.2023)

Focused on data infrastructures in the context of the course "Making data, making worlds: An introduction to data practices." Here are the main points discussed:

Rear view mirror:

- Building boxes is critical to assembling comparable data at scale, but it is not easy.
- Classification systems need to take account of both theoretical and practical considerations.
- Global statistics and models involve trade-offs and choices that should be acknowledged.

What is an infrastructure?

- Infrastructures are sets of collective facilities necessary for human activities.
- They can be physical (bricks, wires) or intangible (protocols, standards).
- Infrastructures are often invisible when functioning properly, but become noticeable when they break down.

Characteristics of Infrastructure (Bowker and Star 1999):

- Embeddedness: Infrastructure is within and part of other structures.
- Transparency: It supports tasks invisibly.
- Reach: Extends spatially or temporally beyond a single event.
- Learned as part of membership: Outsiders learn about the infrastructure.
- Embodies standards: Adapts and connects to other infrastructures in a standardized way.
- Built on an installed base: Grows from an existing base, inheriting its strengths and limitations.
- Becomes visible on failure: Invisibility becomes visible on failure.
- Fixes in modular increments: Changes are not global but incremental.

Data Infrastructures:

- Digital means of storing, sharing, connecting and consuming data over the Internet.
- Includes servers, cables, catalogues, directories and more.
- Critical to many aspects of life, including information sharing, communication, government and business.

Application areas:

- Data infrastructures are essential for information exchange, governance, economics and are emerging as business models.

Databases for EU internal security

- The Schengen visa system relies on data infrastructures to exchange information.
- Supranational databases pool data from individual Member States to manage security and border control.

Development of a database for information exchange:

- One example is the Visa Information System (VIS).
- A study of the construction of the VIS reveals frictions and coordination processes in a multi-level governance system.

Implications:

- Building and maintaining data infrastructures involve social and political negotiations.
- Ongoing coordination is required for data generation, formats, standards, security and more.
- Database design is not predetermined and has power implications.

Broader perspective on infrastructure (Crawford 2021):

- Artificial Intelligence is embodied, material and connected to various global systems.
- The environmental and social impacts of data infrastructures are significant.

Summary:

- Infrastructures are ubiquitous and form the backbone of communication.
- Construction and maintenance require coordination and alignment.
- Data infrastructures are embedded in global trade and capital structures.

Session 5 (02.11.2023) "Data Journeys and Data Friction"

The main points discussed are:

Reflections on Data Infrastructures:

- Data infrastructures are powerful tools for communication, business, and governance.
- Building and maintaining these infrastructures requires coordination and alignment.
- Data infrastructures are embedded in global structures of trade and capital.

Challenges in Classification Systems and Standards:

- Classification systems and standards involve trade-offs, especially on a large scale.
- Challenges arise when these systems need to be universally applicable and compatible with different local infrastructures, cultures, and practices.

Measuring climate and data frictions:

- Discussion of how climate scientists produce knowledge about climate and patterns of climate change.
- Introduction to the concept of "data frictions" in the context of global climate models.

Climate modelling and data frictions (Edwards 2010):

- Examine global climate models and the history of efforts to collect weather and climate records.
- Highlights the existence of "data frictions" that complicate the production of global data sets.

Definitions in meteorology:

- Weather vs. climate and the longitudinal function of climate.
- Climate change as long-term shifts in temperature and weather patterns.

Modelling weather/climate:

- Weather and climate modelling involves complex processes and a large number of variables.
- The need for global data has led to the development of computational infrastructures for weather prediction.

Generating global data:

- Efforts since the 1940s to make global data possible.
- Challenges and the role of standards in facilitating global data flows.

Concept of friction:

- Friction in physical systems as resistance at interfaces.
- Introduction of computational friction in information systems and the challenges it poses.

Data Friction:

- Define "data friction" as the cost in time, energy, and attention required for various data-related activities.

Data Journeys (Leonelli 2020):

- Introduces the data journey approach, focusing on the life of data as it moves through space and time.
- Emphasis on data friction at different stages of data movement.

Examples of making data global:

- Strategies used to make data global, including automation, manual inspection, and interpolation.
- The role of computer models in creating consistent global datasets.

Implications of global data for meteorology:

- The changing meaning of "data" in meteorology today.
- The merging of data and theory to create a "data image" rather than a traditional data set.

Data frictions in security contexts:

- Study of Passenger Information Units in the EU and the data frictions in reusing commercial data for security.
- Several efforts are needed to make data suitable for security intelligence.

Reducing friction in security data:

- Challenges in reducing friction, conflicts between quality requirements, and establishing quality control processes.
- Practical adaptations to make data-driven intelligence actionable.

Summary:

- Data, as it travels, creates friction that must be resolved.
- Trust in data is rare due to deviations and inconsistencies, requiring active labor to fix and extract meaningful knowledge.
- Fixing data changes its shape and informational value.

Session 6 (9.11.2023)

Focused on the role of metadata, particularly in the context of surveillance practices. The main points discussed were:

Metadata Overview:

- Metadata, in the context of this session, refers to structured information that describes, explains, locates, or manages an information resource.
- Metadata is often generated in an automated way and plays a crucial role in ensuring access to information resources both in time and in space.

Types of metadata: Descriptive metadata.

- Descriptive metadata is used to discover and locate data, including keywords for search queries.
- Administrative metadata includes technical, preservation, and legal information necessary for the operational capability of information systems.
- Structural metadata establishes links between smaller pieces of data and describes the structure, types, and relationships of data.

Examples of descriptive metadata:

- The session provided examples of descriptive metadata, such as library catalogue records and online trading platforms, to illustrate how metadata facilitates the identification and retrieval of information.

Challenges in systematising description:

- Uniformity of both categories and category content is critical to metadata systems and requires adherence to standards.
- Standardisation involves addressing issues of semantics, syntax, and content rules.

Example standard: Dublin Core:

- Dublin Core is presented as an example of a descriptive metadata standard, consisting of fifteen key metadata elements for describing digital or physical resources.

Administrative and structural metadata:

- Administrative metadata, including technical details and rights information, is essential for data preservation and access.
- Structural metadata establishes links between different parts of the resource, allowing more complex objects to be assembled.

Metadata Surveillance:

- This session discussed the implications of metadata in the context of surveillance, highlighting how metadata surveillance capabilities have expanded the scale and scope of surveillance activities.

The relevance of metadata:

- Despite their technical nature, metadata can reveal profound insights about individuals, their habits, and the connections between entire networks.

Evolution of the meaning of metadata:

- The meaning of metadata has evolved, especially in the wake of the Snowden revelations, from a naive technical understanding to a recognition of its role in socio-technical data assemblages.

Conclusion:

- Metadata serves primarily as an organisational tool for data management but has also become a focal point for global surveillance programmes, raising concerns about individual privacy rights.

The session highlighted the importance of understanding the role of metadata in shaping knowledge and its implications in surveillance contexts.

Session 7 (16.11.2023)

Data capitalism and data justice in the context of the course "Making data, making worlds: An introduction to data practices". Key points discussed included:

Metadata Overview:

- Metadata primarily serves as an organisational tool for data management.
- Although considered technical and neutral, metadata has become a focal point for global surveillance and counter-terrorism programmes.

What is capitalism?

- Capitalism is an economic system based on private property, the profit motive, competitive markets and wage labour.
- Proponents argue that it promotes efficient production, economic growth, innovation and social/financial freedom. Critics point to its exploitative, environmentally unsustainable and inequitable nature.

Data and capitalism:

- Data is described as the "new oil" in the data business model.
- The equation is (more) data + (better) algorithms = value added.
- Surplus value has implications for research, healthcare, governance, international security and the economy.

Data and platforms:

- Platforms are essential to the data business model.
- Characteristics of successful platforms include infrastructure, data monetisation, network effects, monopolisation, cross-subsidisation, rule-setting and extraction.
- Examples of successful platforms include Apple's App Store, Google's search engine, Uber and eBay.

Surveillance capitalism:

- Surveillance capitalism exploits and controls human behaviour through data extraction.
- Key mechanisms include commodification and prediction, leading to behavioural futures markets.

Example: Pokémon Go:

- Pokémon Go illustrates how data can be extensively collected (location, device information) and commodified for targeted advertising.

Instrumental power:

- Surveillance capitalism introduces instrumental power, which shapes human behaviour on a large scale through automated computer architectures.

Eight Theses on Surveillance Capitalism:

- Surveillance capitalism is a new economic order with unprecedented concentrations of wealth, knowledge and power.
- It poses significant threats to humanity and democracy.

Data Justice:

- Data justice examines the social impact of data.
- Data harms include exploitation, discrimination, loss of privacy, surveillance, manipulation, exclusion and injustice.

Data activism:

- Data activism addresses data injustices through digital means, including data collection, visualisation and resistance to surveillance.

Summary:

- Data and capitalism are closely linked, resulting in new forms of capitalism based on platforms and data extraction.
- Data capitalism introduces new forms of harm with implications for justice.

The session highlights the need to integrate data into discussions of justice and introduces the concept of data activism to address data injustices.

Session 8 on Data Policy

Explored the evolving landscape of data regulation and its implications. Key points included:

Flashback:

- Capitalism and justice issues are intertwined.
- Data has led to new business models centered on platforms and data extraction.
- Data capitalism introduces new forms of harm with implications for justice.

Generalizability of Cases:

- Analysis of US tech companies may have global relevance due to de-territorialization of business models.
- Challenges arise due to cultural differences, which require differentiated approaches.

Uber's Business Model:

- Uber's success is based on its platform that connects transport providers and customers.
- Monetization is through commissions and the extraction of interaction data for analysis.
- Uber's disruptive business strategy faces regulatory challenges worldwide.

Regulatory Dilemma:

- The Collingridge Dilemma poses challenges in regulating rapid technological innovation.
- Uber's strategy capitalizes on the dilemma, betting that regulation won't catch up.

The Regulatory Response:

- Uber's success has been uneven around the world, with more acceptance in common law countries and bans in civil law countries.
- The lack of regulation for new data-driven business models is driving the data policy debate.

The Right to be Forgotten:

- The Costeja González v Google case led to the establishment of the "right to be forgotten" in the EU.
- The EU's General Data Protection Regulation (GDPR) enshrines this right and regulates the processing of personal data.

GDPR Principles:

- The GDPR limits the use of data, requires purpose limitation, and addresses automated decision making.
- It empowers citizens with rights such as access, rectification, data portability, and challenging algorithmic decisions.

EU as Protector:

- Through the GDPR, the EU is protecting citizens from powerful corporations in the digital age.
- The GDPR serves as a model for data policy worldwide, emphasizing the right to privacy.

Enforcement and Critique:

- GDPR fines indicate active enforcement, but critics cite bureaucratic complexity and stifled innovation.

Clashing Visions:

- The EU leads the way in data regulation, while the US lacks a GDPR equivalent, leading to clashes in global business practices.

Global Trends:

- One global trend emphasizes citizen empowerment in the data ecosystem.
- Data protection is a key regulatory pillar, but a consistent global framework remains elusive.

Summary:

- Data policy is a response to societal datafication, but regulatory development lags behind technological progress.
- The choice of regulation depends on political prioritizations regarding economic order and the rights of citizens/consumers.
- Regulation is coined by jurisdictional questions and different visions of data policy.

This session highlighted the ongoing evolution of data policy and emphasized the need for a comprehensive global regulatory framework.

Session 9 (30.11.2023) - Data Governance

Looking Back:

- Data policy emerges in response to societal datafication and evolving business and government capabilities.
- Technological innovation outpaces the development of an appropriate regulatory environment.
- Regulatory decisions depend on political priorities and vary by jurisdiction and vision of data policy.

Aggregation of data

Data sharing

EU JHA databases

What could go wrong?

- Challenges in EU databases include data quality issues, inconsistencies and misuse.
- Discussion on how to ensure good data quality in centralised EU databases.

The policy strategy I

- European Commission 2021: Implementing Regulation on automated data quality control mechanisms.
- Sorting data into categories: "good quality", "low quality", "rejected".

Policy Strategy II

- Council of the EU 2020: Roadmap for standardisation for data quality.
- Data quality standards for biometric and alphanumeric data, reference catalogue and cybersecurity.

Data quality as an "asset"

- Reliable data is essential for meaningful use; poor data quality increases costs and leads to sub-optimal outcomes.

Dimensions of data quality

I: Accuracy

II: Completeness

III: Trustworthiness of sources

IV: Timeliness

V: Accessibility

Interdependencies and trade-offs

- Different dimensions of data quality are interdependent; trade-offs force organisations to prioritise.

Doing data quality

- Ensuring quality data involves complex processes and data governance.

The CURATE Project

- European Research Council funded project (6/2022 - 5/2027) addressing data quality challenges in law enforcement databases.
- Research questions focus on the origins of "bad" data, its effects and strategies for improvement.

From data policy to data governance

- Data policy governs the production and use of data; data governance governs the use of data within an organisation.

Implications

Practice

- Professionalisation of data quality is essential; this includes defining requirements, roles, responsibilities and allocating resources.

Policy

- Calls for further reform and harmonisation, in particular for reliable data to be entered into EU databases.

Research

- Stresses the need for more academic engagement with data quality, especially in the public sector.

Step 2: Zoom in

- Qualitative research on data journeys in European countries, involving practitioners to gain a deeper understanding of data practices.

summary

- Trustworthiness and reliability of data are key challenges for data-driven decision making.
- Data quality is multidimensional, involves trade-offs and requires appropriate governance.
- An under-researched aspect, especially in the public sector.

Session 10 (07.12.2023) - Data Preservation

The Challenge of Data-Driven Decision Making

- Often defined in practice as data quality.
- Multidimensional concept with trade-offs.
- Organizational governance critical to trustworthiness.

Data Lifecycle

- Data created for a specific purpose.
- Moved, merged, and used in different ways throughout its life.
- Options at the end of lifecycle: preservation or death.

Preservation

- Principles, policies, and rules for extending the life of data.
- Addresses changing technologies and migration needs.
- Active curation to prevent damage and loss.

The Archive

- National archives, libraries, government agencies, museums, etc.
- Preserves the full set of records with supporting documentation.
- Actively structured, curated, documented, and planned.
- Long-term effort for future re-use of data.

Life and Death Decisions

- Not everything is worth preserving, especially in the digital age.
- Archivists and librarians sort and appraise files.
- Some policies, such as the "right to erase," encourage data destruction.

Data Death

- Irreversible deletion without backup.
- Intentional or accidental, implying unavailability.
- Some deleted data may be "resurrected" under certain circumstances.

The Politics of Digital Preservation and Loss

- Thylstrup (2018) examines who preserves digital artifacts.
- Highlights decisions about what to keep and what to discard.
- Raises concerns about power dynamics, legal concepts, and societal implications.

The Internet Never Forgets - Or Does It?

- Data frequently disappears, e.g. the average web page is deleted within 100 days.
- Initiatives such as Internet Archive, Project Gutenberg, Europeana, Google Books aim to digitize cultural artifacts.
- Brewster Kahle's perspective on large-scale data deletion.

Digital Cultural Heritage Projects

- Internet Archive, Project Gutenberg, Europeana, Google Books.
- Aim to digitize, preserve, and make accessible cultural artifacts.
- Controversies surrounding these initiatives.

Internet Archive vs. Publishers

- Lawsuit over copyright infringement.
- Publishers sue Internet Archive for lending digital copies.
- Debate over "controlled digital lending."

Ruling and Implications

- New York Southern District Court sides with publishers.
- Internet Archive to appeal.
- Implications for public access to information vs. intellectual property rights.

Digitalization and Democracy

- Public discourse influenced by social media.
- Challenges associated with content moderation algorithms.
- Fundamental questions for democracy.

Digitization and Culture

- Cultural forms heavily dependent on digital formats.
- Shift from physical to digital availability.
- Power dynamics in the hands of for-profit companies.

Reclaiming Control

- Thylstrup (2018) emphasizes the need to rethink power structures.
- Calls for stronger institutional safeguards, rules, and stable support.

Summary

- Data preservation practices are critical to cultural heritage.
- Driven by choices, not a simple story of progress.
- Growing struggle to define ownership of humanity's digital memory.

The Performativity of Data

- A concept that highlights how data bring phenomena into being.
- Power struggles implicit in the creation and analysis of data.

Classification Systems

- Conventions that define the translation of empirical phenomena into data.
- Rarely fixed, changing over time based on social and political conventions.

Infrastructures

- Key to the creation, sharing, and analysis of data.
- Construction and maintenance require coordination and alignment.
- Raises issues of data inclusivity.

Data Journeys

- Friction in data journeys that require resolution.
- Active labor required to 'fix' data for meaningful knowledge extraction.
- Changes in data form and informational value during fixing.

Metadata

- Primarily organizational, but also used for surveillance.
- Collection of information about activities and networks.
- Extends to uninvolved individuals.

Data Capitalism

- Data productive of new business models.
- Monetization of human nature.
- Implications for behavioral prediction and nudging.

Data Justice

- Harms of data-driven decision making.
- Social sorting, discrimination, lack of transparency.
- Implications for equity and life chances.

Data Politics

- Emerging responses to datafication.
- Lagging behind technological innovation.
- Regulatory choices influenced by political prioritization.

Data Governance

- Challenges in ensuring trustworthiness and reliability.
- Multidimensional concept with trade-offs.
- Under-researched, especially in the public sector.

Data Deaths (Repeated)

- Data preservation practices key to cultural heritage.
- Driven by choices, not a simple story of progress.
- Growing struggle to define ownership of humanity's digital memory.