

# BEYOND DATA QUALITY: DATA EXCELLENCE CHALLENGES FROM AN ENTERPRISE, RESEARCH AND CITY PERSPECTIVE

28th of July 2018, 7th International DATA Conference 2018, Porto, Portugal

Johannes Sautter<sup>1</sup>, Rebecca Litauer<sup>1</sup>, Rudolf Fischer<sup>1</sup>, Tina Klages<sup>2</sup>, Andrea Wuchner<sup>2</sup>, Elena Müller<sup>1</sup>, Gretel Schaj<sup>1</sup>, Ekaterina Dobrokhotova<sup>1</sup>, Patrick Drews<sup>1</sup> and Stefan Riess<sup>3</sup>

<sup>1</sup> Fraunhofer IAO, Stuttgart, Germany

<sup>2</sup> Fraunhofer IRB, Stuttgart, Germany

<sup>3</sup> KPMG Wirtschaftsprüfungsgesellschaft, Köln, Germany



# Organisational data challenges



# AGENDA

Beyond Data Quality:

Data Excellence Challenges from an Enterprise, Research and City Perspective

- Introduction
- Method
- Related Work
- Domain Cases
  - Enterprises
  - Research Institutes
  - Cities
- Data Excellence – Cross-domain analysis and preliminary definition
- Conclusion and Outlook

# Challenges beyond data quality

As consequence of the digital transformation

## Data quality (DQ): fitness for use by data consumers

(Wang and Strong, 1996)

However, insufficient DQ is not the only realized shortcoming:

- Operational excellence issues
- Non-compliance to legal issues can lead to significant financial penalties or even cause criminal liability (European Comission, 2016).
- Insufficient process quality is not directly measured by DQ, but has e.g. indirect impact on DQ dimension believability (Piro, 2014)

### Data Quality (DQ)

#### Contextual DQ

- Relevancy
- Value-Added
- Timeliness
- Completeness
- Vollständigkeit
- Amount of data

#### Accessability DQ

- Accessability
- Access security

#### Representational DQ

- Interpretability
- Ease of understanding
- Representational consistency
- Concise representation

#### Intrinsic DQ

- Accuracy
- Glaubwürdigkeit
- Hohes Ansehen
- Objektivität

(Wang and Strong, 1996)

**Operational excellence:**  
"execution of the business strategy more consistently and reliably than the competition"

(Soto, 2013)

Goal of organisations: Shape future of organisations using data.

# Organisational data challenges



# Measuring and addressing data challenges

## Solution vs. problem perspective

Consensus in scientific and practitioner's community:

**Only structured management leads to better performance in dealing with data**

- However, in many organizations resort silos dominate
- Independent non-integrated data sources

### **Solution perspective:**

Data management concepts and methods propose:

- Data cleansing (Maletic and Marcus, 2005),
- Master data management (Otto et al., 2011; Scheuch et al., 2012)
- Data quality management (Otto and Österle, 2016; Morbey, 2011)
- Data governance (Otto et al., 2011; Otto and Österle, 2016).

### **Problem perspective:**

- DQ framework allows structured measuring, assessment and monitoring
- No framework is able to explain or measure existing challenges beyond DQ

This paper: Structured challenge assessment in very different organization types: enterprise, research institute and city administration

# Method

- Comprehensive literature review  
→ research gap of few existing frameworks for data challenges, especially for non-DQ issues

4 challenge dimensions from literature & project experience:

- **Operational excellence** (including internal compliance and standards)
- **Legal challenges** (obligations anchored within laws, external compliance)
- **Data management process quality** (data management maturity)
- **Data quality** (fitness for use by data consumers)

- Cross-project study with selected available case studies:
  - **Enterprise:** case studies of data governance projects and triggering challenges within 4 DAX 30-enterprises (4 semi-structured interviews + project experience)
  - **Research institutes:** Single case study at an institute of the German Fraunhofer Gesellschaft (19 half-standardized interviews)
  - **Cities:** Workshops with 15 scientific employees and 10 international strategic officials and smart city leaders
- Table approach for empirical and derived issues
- Cross-domain analysis with literature enhancements

# Related Work

## Organisational data excellence and related aspects

- DQ
    - "Information delivered [as] a total product" (Wang et al. 1998)
    - "Suitability to fulfil determined requirements" (Paskaleva et al., 2017)
    - Concept containing several dimensions (Morbey, 2011, p. 26–27) (Wang and Strong, 1996)
  - Process quality
    - High process quality thereby has a direct impact on DQ (Glowalla and Sunyaev, 2013).
    - Maturity models can serve for the measurement of data management process quality (Mosley 2008, Otto and Österle, 2016; Pentek et al., 2017).
  - Operational excellence (OE)
    - Ideal business performance across all domains: OE is "the consequence of an enterprise-wide practice of ideal behaviours based on the correct principles" (Rusev and Salonitis, 2016)
- "the impact of data management on the data itself, [...] with regard to **data quality** [...], but also with regard to additional data related aspects, such as **data compliance, data security and privacy**"  
(Pentek et al., 2017)
- Morbey (2011) identifies data owners, and not the DQ team, in charge of content accuracy investigations (Morbey, 2011, p. 28–30)
  - Not all domain needs regarding data are transferable to computer measurable criteria (Piro, 2014)
- Legal challenges
- Compliance requires the observance of "rules and regulations imposed by any regulatory bodies to which a firm is subject" (Edwards and Wolfe, 2005).
    - the influence of the law in operational activity
    - counteract against possible quality losses and to achieve operational excellence
  - The concept of compliance management consists of the functions risk minimization, harm reduction, liability obligation, and corporate efficiency increase (Wecker and Ohl, 2013)

# Domain Cases

## Enterprises

Industry 4.0: fully wired shop floors producing intelligent products

- Operational excellence means economic success
- Efficiency and ability to compete
- Networking and collaboration

## Research Institutes

Open Science: maximum access to scientific knowledge for research, society and economy

- Digital interchange of data, ideas and results
- FAIR research data  
(findable, accessible, interoperable and reusable)  
(Wilkinson et al., 2016)
- Data peer-review and citation
- Duty to publish data (EC2016)  
(Opt-out for privacy, security and exploitation)
- Institutions and research associations responsible (OECD 2007, DFG 2015)

## Cities

Smart City: Creating new data-based services for citizens

- New available data allows for better performance
- Smart city data are owned by different stakeholders  
(Saujot and Erard, 2015)
- Exchange standards exist  
(KoSIT, 2018) (DIN, 2017)
- Predominant and outdated silo mentality" (Pfau-Weller and Radecki, 2018)

# Domain case: Enterprises

Empirics, theoretical derivations

4 Case studies were main causes for data governance in DAX 30 companies  
(automotive, bank, chemistry and energy sector).

Main Challenges	Operational Excellence Challenges	Legal Challenges	Data Management Process Challenges	Data Quality Challenges
Unreliable contract data	Lack in transparency whether contract contents have been entered and approved correctly (automotive OEM procurement)	AO, HGB, IFRS 2015 and 2016ii demand for contract transparency	Unclear responsibilities, no clear contract master data creation processes	Inconsistent contract details (especially believability, completeness, concise representation)
No reliable risk management	No identification of business risks from bank data	BCBS 239iii and MaRiskiv demand data governance & IT infrastructure, aggregation and report of risk data	Unclear and unreliable data management processes (bank)	Low data quality in all dimensions
Insufficient business performance	Insufficient performance in procurement and sales	Dangerous goods remain at customs officev	No data governance as basis for clear data ownership	No reliable supplier and customer master data (chemical company)
Regulatory risks	Inefficient steering, risks for inaccurate cable excavations, customer complaints (grid operator)	Regulatory risks regarding network performance and maintenance	No clear responsibilities	Double data entries, accuracy

# Domain case: Research Institutes

Empirics, theoretical derivations

1 case study, 19 interviews, leading/scientific/service employees  
Institute of the German Fraunhofer Gesellschaft, ~600 employees

Main Challenges	Operational Excellence Challenges	Legal Challenges	Data Management Process Challenges	Data Quality Challenges
Internal Competition	Not existing exchange post-usage hinders research excellence	–	Competitive thinking between scientists	Lack of accessibility for excellent research
No internal Exchange	No efficient collaboration possible	–	Lack of communication between scientists within research field	Lack of representational data quality and accessibility data quality
Effort-benefit Balance	Value for effort in metadata structure is not seen	–	Easy and low-effort processes for metadata standards are missing	Lack of representational and intrinsic data quality
Inattention on duties to funder	No provision of data to research community	Lack of knowledge on funder compliance	Not sufficient knowledge management	–
No process Standards	Scientists collect their data in different ways	–	Lack of tools and processes guaranteeing more standardized collection of data	Insufficient standardization of data and meta data (representational data quality)
No data standards	There is no minimum standard for metadata and data storage	Ensure that institute/community standard complies to legal requirements	Lack of process for agreement of scientists and metadata standard coordination	Lack of representational consistency
Shortage of scientist development	Little methodological expertise	–	Provision of little methodological support in form of workshops/consultancy services	Lack of data quality in all dimensions

# Domain case: Cities

Empirics, theoretical derivations

1 workshops with scientific employees working in international smart city projects  
1 workshop with strategic smart city leaders

Main Challenges	Operational Excellence Challenges	Legal Challenges	Data Management Process Challenges	Data Quality Challenges
No Data Distribution	Different responsible resorts (E.g. City Planning: Building of nature-based corridor)	Strict legal limits for data exchange between departments	No central data access, no single data source (e.g. water pipe hinders tree planting)	Different data sources, no overview over existing data
Inattention of Regulations	–	No implementation of regulations	–	–
No consistent handling of data	No cross-resort data analysis (use and mix existing data for better performance)	Heterogeneous usage regulations dependent on ownership and licensing	Effortful extraction of information, Distinct processes dependent on ownership	No concise representation and no accessibility
Shortage of methodical knowledge	Unclear definition of methods (indicators, workshops)	–	Unclear standards and processes, uncoordinated collection of data	Data is wrong (Accuracy)
Old data	Data not operationally usable e.g. for safety and security issues	–	Irregular or non-structured data collection	Data is out-dated (Timeliness)
No data standards	Extra data processing effort (e.g. increase urban climate resilience)	–	No standard format defined or practically used	Different data format dependent on source and supplier (interpretability)
No process Standards	Comparison of mobility data of different cities	–	Different standards, units, or methods in different cities	No concise representation, no interpretability

# Data Excellence

## Cross-domain analysis and preliminary definition

### ■ Operational Excellence Challenges

- Operational efficiency (Otto, 2011)
- Exchange and collaboration
- Data availability (Panian, 2010)
- Operational integration and interoperability (Otto, 2011)

### ■ Legal Challenges

- Operational legal requirements (Otto, 2011)
- Awareness of regulations

### ■ Data management process quality challenges

- Clear responsibilities, processes and guidelines
- Data transparency and auditability (Panian, 2010)
- Central data access

### ■ Data quality challenges (Wang, 1998)

- Intrinsic data quality
- Accessibility
- Contextual data quality
- Representational data quality

**Data excellence: the capability of an organisation to execute its strategy consistently and reliably with a suitable, transparent and compliant availability and integration of high quality data.**

The concepts compliance and operational excellence hereby are not part of data excellence, but reach its borders.

# Conclusion

- Deciders in organisations need to take new data into account
- Data excellence challenges such as data availability from an enterprise, research and city perspective were assigned to four challenge dimensions including DQ.
- Problem and solution perspective necessary
  - Solution side: well settled concepts exist
  - Problem side beyond DQ remains disregarded
- Researchers and practitioners need to be able to assess an organisation's data suitability
- This paper provides a first overview of data excellence as well as a preliminary definition.

# Outlook

*Further research is needed for:*

- Structured literature review on non-DQ topics, examine data challenges on a more representative and broader empirical basis
- Profound, sound and generally valid criteria are needed for non-DQ challenges
- Departments on compliance, data management and operational departments need to cooperate more efficiently

# Thank you very much for your attention!

Questions? Feedback?

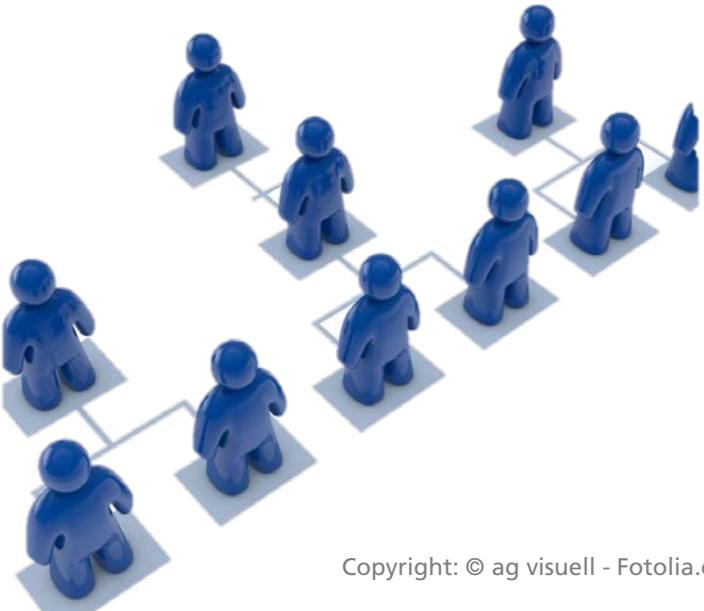
## Contact

Dipl.-Inf. Johannes Sautter  
Johannes.Sautter@iao.fraunhofer.de  
Tel: +49 (0)711 / 970-2387



Urban Data & Resilience  
Fraunhofer Institute for Industrial Engineering IAO  
Nobelstr. 12 | 70569 Stuttgart | Germany

[www.iao.fraunhofer.de](http://www.iao.fraunhofer.de)



# Literature (1)

- DFG (2015). Leitlinien zum Umgang mit Forschungsdaten. <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeze.html>.
- DIN (2017). DIN SPEC 91357: Referenzarchitekturmodell offene urbane Plattform (oup). <https://www.din.de/de/wdc-beuth:din21:281077528>.
- Edwards, J. and Wolfe, S. (2005). Compliance: A review. *Journal of Financial Regulation and Compliance*, 13(1):48–59.
- European Commission (2016). Open innovation, open science, open to the world: A vision for Europe. Publications Office of the European Union, Luxembourg.
- Glowalla, P. and Sunyaev, A. (2013). Process-driven data quality management through integration of data quality into existing process models. *Business & Information Systems Engineering*, 5(6):433–448.
- KoSIT (2018). Xöv-standards. <https://www.xoev.de/>.
- Maletic, J. I. and Marcus, A. (2005). Data cleansing. In *Data Mining and Knowledge Discovery Handbook*, pages 21–36. Springer.
- Morbey, G. (2011). Datenqualität für Entscheider in Unternehmen: Ein Dialog zwischen einem Unternehmenslenker und einem DQ-Experten. Gabler Verlag / Springer Fachmedien, Wiesbaden, 1. Aufl. Edition.
- Mosley, M. (2008). DAMA-dmbok functional framework version 3. DAMA International.
- OECD (2007). OECD principles and guidelines for access to research data from public funding. <https://www.oecd-ilibrary.org/docserver/9789264034020-en-fr.pdf>.
- Otto, B. (2011). A morphology of the organisation of data governance. In ECIS, volume 20, page 1.
- Otto, B., Kokemüller, J., Weisbecker, A., and Gizanis, D. (2011). Stammdatenmanagement: Datenqualität für Geschäftsprozesse. *HMD - Praxis der Wirtschaftsinformatik*, 48(279):5–15.
- Otto, B. and Österle, H. (2016). Corporate Data Quality: Voraussetzung erfolgreicher Geschäftsmodelle. Gabler and Springer Gabler, Berlin, 1. Aufl. 2016 Edition.

# Literature (2)

- Pentek, T., Legner, C., and Otto, B. (2017). Towards a reference model for data management in the digital economy. In Maedche, A., Vom Brocke, J., and Hevner, A., editors, *Designing the Digital Transformation*. Karlsruhe.
- Pfau-Weller, N. and Radecki, A. v. (2018). Was macht Städte smart? – die Morgenstadt-initiative: In der Morgenstadt initiative entwickelt die Fraunhofer Gesellschaft gemeinsam mit Partnern aus industrie und kommunen Lösungen für die Stadt der Zukunft.
- Piro, A., Editor (2014). *Informationsqualität bewerten: Grundlagen, Methoden, Praxisbeispiele*. Symposion, Düsseldorf, 1. aufl. edition.
- Rusev, S. J. and Salonitis, K. (2016). Operational excellence assessment framework for manufacturing companies.
- Procedia CIRP, 55:272–277.Saujot, M. and Erard, T. (2015). Smart city innovations for sustainable cities? An analysis based on data challenges.  
<https://www.iddri.org/sites/default/files/import/publications>
- Scheuch, R., Gansor, T., and Ziller, C. (2012). *Master Data Management: Strategie, Organisation, Architektur*. dpunkt-Verlag, Heidelberg, 1. Aufl. Edition.
- Soto, F. (2013). A better definition of operational excellence. <http://www.wilsonperumal.com/blog/a-betterdefinition-of-operational-excellence>.
- Wang, R. Y. and Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33.
- Wecker, G. and Ohl, B., editors (2013). *Compliance in der Unternehmerpraxis: Grundlagen, Organisation und Umsetzung*. Springer Gabler,Wiesbaden, 3. aufl. 2013 Wdition. <http://dx.doi.org/10.1007/978-3-658-00893-2>.
- Wilkinson, M. D. et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3:160018.