



Journal of Data and Information Quality

Special Issue on Software Engineering and AI for Data Quality

Guest Editors:

- **Sagar Sen**, SINTEF (Norway), sagar.sen@sintef.no
- **Andreas Metzger**, The University of Duisburg-Essen (Germany), andreas.metzger@paluno.de
- **Foutse Khomh**, Polytechnique Montréal (Canada), foutse.khomh@polymtl.ca
- **Phu Nguyen**, SINTEF (Norway), phu.nguyen@sintef.no

Modern systems are centered on data, using data in novel and intelligent ways. One key driver for increased data availability are data spaces and data sharing platforms, which are booming¹. Another key driver is cyber-physical systems (CPS)/Internet of Things (IoT), which have been realized in many industrial sectors and application domains facilitating data acquisition from physical sensors and devices on unprecedented scale.

All this makes data quality become crucial. As an example, the quality of the data acquired and used for decision support is a common factor across industrial sectors that underpin our reliability and trust in CPS/IoT. However, data quality can deteriorate due to several factors. For example, CPS/IoT systems can experience sensor faults and failures due to operating in harsh and uncertain environments, bias in measurement in digital health, intermittent loss in connectivity due to physical barriers or unreliable communication protocols, and inconsistencies in data being transformed/pre-processed and duplicated. Poor data quality reduces our trust and reliance on such systems.

In this ACM JDIQ special issue, we invite submissions on novel software engineering and Artificial Intelligence (AI) techniques that address data quality issues of modern systems. As an example, emerging modern distributed software systems must consider data quality when acquiring and processing along the complete edge-fog-cloud continuum. New software engineering approaches should interact hand in hand with AI models to pre-process data, detect anomalies within short-term streaming data and long-term historical data, repair erroneous data, replace missing data, and detect ethical issues such as bias in data. In addition, we look at data quality in the social, cybersecurity, and distributed dimension. And not least, we are interested in public sensor datasets with data quality metadata, e.g., from CPS/IoT (manufacturing, digital health, energy, etc.) to enforce data quality in AI and Machine Learning (ML) applications.

Topics

The topics of interest are inspired from the themes above and include, but are not limited to:

- Software/hardware architectures and frameworks for data quality management, (e.g., in CPS/IoT)
- Software engineering and AI to pre-process and clean data
- Software engineering and AI to detect and repair erroneous values in data
- Software engineering and AI for robust abstraction from data
- Software engineering and AI for data fusion
- Software tools for data quality management, testing, and profiling
- Data quality metrics for Software Engineering/AI
- Public sensor datasets with data quality metadata, e.g., from CPS/IoT (manufacturing, digital health, energy, etc.) to enforce data quality in AI/ML applications
- Distributed ledger and blockchain technologies for quality tracking
- Augmenting data quality using AI/ML techniques for data cleansing, data enrichment, and data

¹ <https://www.bdva.eu/facilitating-interoperable-data-sharing-new-data-spaces-support-centre-now-launched%C2%A0>

integration/synchronization

- Case studies that have evaluated an existing technique or tool on real systems, not only toy problems, to manage data quality in cyber-physical systems in different sectors
- Certification and standardization of data quality (e.g., in CPS/IoT)
- Approaches for secure and trusted data sharing, especially for data quality (including confidentiality, integrity, availability, accountability, and privacy), data management, and governance (e.g., in CPS/IoT)
- Trade-offs between data quality and data security (e.g., in CPS/IoT)

Expected Contributions

We welcome the following types of research contributions:

- **Survey papers:** Should present a coherent review of scientific work related to data quality issues in data preparation, together with interesting future research directions in the field (up to 25 pages).
- **Technical papers:** Should present novel research contributions on the topics above, clearly describing the progress from the state of the art and providing evidence for the benefits of the contributions (up to 25 pages).
- **Experience papers:** Should detail recent applications of data quality techniques in practice and industry, providing pertinent application scenario(s), lessons learned, and open problems (up to 15 pages).
- **Resource papers:** Should present a new resource, such as a dataset or tool, or an interesting compilation of multiple datasets (up to 15 pages).

Important Dates

- Submission deadline: 30 April 2023
- First-round review decisions: 15 August 2023
- Deadline for revision submissions: 15 October 2023
- Notification of final decisions: 15 December 2023
- Camera-ready Manuscript: 15 January 2024
- Tentative publication (early access): February 2024

Submission Information

JDIQ welcomes manuscripts that extend prior published work, provided they contain at least 30% new material, and that the significant new contributions are clearly identified in the introduction. Submission guidelines with Latex (preferred) or Word templates are available at <https://dl.acm.org/journal/jdiq/author-guidelines#subm>

Please submit the paper by selecting as the type of submission: “**SI: SEA4DQ**”.

For questions and further information, please contact **Sagar Sen**, sagar.sen@sintef.no