Metadata and their role in the digital transformation of Water Resource Recovery Facility operations

Authors: Daniel Aguado[1], Frank Blumensaat[2], Queralt Plana[3], Victoria Ruano[4], Oscar Samuelsson[5], Kris Villez[6]

[1] Universitat Politècnica de València, Valencia, Spain
[2] Eawag, Dübendorf, Switzerland
[3] SIAAP, Paris, France
[4] University of Valencia, Valencia, Spain
[5] IVL Swedish Environmental Research Institute, Stockholm, Sweden
[6] Oak Ridge National Laboratory, Oak Ridge, TN, USA

The need for meta-data

Digitalization, machine learning and artificial Intelligence are anticipated game-changers that currently are creating a buzz in the water sector. Slowly, the digital transformation is gaining momentum as stakeholders, at all levels, realize the potential of using data. At the same time, the lesson learned from the Big data era was that large datasets do not necessarily contain useful information. Smart decision are based on data that are fit for purpose, meaning they need to be understood, correctly and creatively processed, and interpreted in its context (Ingildsen and Olsson 2016). In this respect, metadata are key to distinguish usable from useless data, and further leverage the value of existing data. Metadata are descriptive information (data about data) that describe

- The data-generating process (e.g., type of sensor, data transfer, filtering)
- The data quality (e.g., validation data, identified anomalies, sensor maintenance routines)
- The context and purpose of collected data (e.g., final goal with data, targeted accuracy, plant and influent conditions, sensor location)

Here, three short examples illustrate how metadata come into play and can, by extension, improve WRRF operations.

1. *Data-driven decision support.* Metadata can provide basic insights about which data that are feasible for model training and validation. Distinguishing garbage from valuables can be a time-consuming and costly activity without metadata (Russo et al. 2020). This activity is currently a bottleneck, but still a key step in both empirical (purely data-driven) and mechanistic (data combined with process knowledge) modelling. Further, metadata in the form of annotations

about the process or sensor condition (faulty or fault free) are central to reduce the training costs, and improve the algorithms detection performance, for any supervised classification algorithms.

2. *Transparent uncertainties and data quality.* Data quality indications obtained through sensor validation and verification procedures are seldom reused as metadata in practice, which is a waste of knowledge. Further, expert knowledge possessed by instrument engineers concerning sensor reliability is neither reused in the form metadata. If these existing data quality indicators instead were automatically used as metadata, the transparency about the sensor data quality would increase. In turn, this would reduce the risk of making poor decisions based on data with a (currently unknown) poor data quality.
3. *Sensor maintenance.* The ubiquitous on-line sensors require substantial maintenance actions, which are at the best recorded in spreadsheets. Predictive maintenance could instead be guided by smart algorithms that exploit these sensor maintenance metadata automatically, in combination with the actual sensor measurements (Samuelsson et al. 2019) and other contextual metadata such as temperature and load conditions.

Take home message

Data governance is a critical, probably the most critical challenge for a successful digital transformation. This transformation will impact both technical systems as well as the organization. New expert roles are anticipated to bridge the understanding about the WRRF specific needs and how algorithms and data relate to this. It has become clear that digitalization does not come for free, and an increased data readiness level is needed (Lawrence 2017). As a first step, a structured approach for governing metadata data may leverage the value of reused data, both for todays and tomorrow's needs. The presentation will introduce metadata and its role in the context of water resource recovery facilities. The IWA Task group Meta-data Collection and Organization (MetaCo) are working to facilitate metadata adoption into practice. On-going MetaCo activities will be presented, emphasizing the knowledge exchange activities in the global Utility groups where many Swedish utilities participate.

Selected references

Ingildsen, P. and Olsson, G. (2016) Smart Water Utilities: Complexity Made Simple, IWA Publishing.

Lawrence, N.D. (2017) Data Readiness Levels.

Russo, S., Lürig, M., Hao, W., Matthews, B. and Villez, K. (2020) Active learning for anomaly detection in environmental data. Environmental Modelling and Software 134.

Samuelsson, O., Zambrano, J., Björk, A. and Carlsson, B. (2019) Automated active fault detection in fouled dissolved oxygen sensors. Water Research 166.