International Standard

ISO 24591-2

Smart water management —

Part 2: Data management guidelines

Gestion intelligente de l'eau —

Partie 2: Lignes directrices pour la gestion des données

First edition
2024-04

Reference number
ISO 24591-2:2024(en)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 224, *Drinking water, wastewater and stormwater systems and services*.

A list of all parts in the ISO 24591 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

To better manage the entire life cycle of water systems, water system owners and operators continually improve operational efficiency, reduce costs and communicate to stakeholders or other systems. In addition, they address safety, regulatory and public authority requirements. One effective approach to achieving these goals is to take advantage of data generated by water systems.

Information-sharing facilities and models established based on these data can provide optimal solutions for the owner or operator of the water systems to meet stakeholder demand for, e.g. drinking water production, transmission and distribution, asset management, risk management, wastewater collection and sanitation, stormwater management and water resource protection. Over the past few years, advances in digital technologies have enhanced the capabilities of data generation; meanwhile, data-processing capacities have also significantly improved.

With the rapid development of new digital technologies, the data generated from water systems are increasing drastically. This "data explosion" has enabled the delivery of new services that:

— increase the operational efficiency of assets and networks;

— reduce or optimize capital expenditures and operating expenses;

— allow better anticipation and assessment of risks;

— enable a smaller environmental footprint;

— enhance regulatory compliance;

— support oversight and substantive accountability to local or national stakeholders;

— improve the level of service to water system customers.

However, large-scale data also dramatically increase the requirements on data storage and data transfer facilities. In addition, it is important to ensure that large-scale data do not result in negative impacts on the environment. Therefore, data management is a challenge for water system owners and operators.

To ensure that the data and information generated by water systems produce maximal values, proper data management approaches should be applied in organizations that work with water systems, e.g. using consistent nomenclature, specifying ownership rules, performing data validation and applying standardization and normalization.

Smart water management —

Part 2:
Data management guidelines

1 Scope

This document provides a general foundation for data management in services, systems and facilities related to drinking water, wastewater and stormwater. It emphasizes data as an asset and introduces basic rules for efficient data acquisition, storage and processing. It aims to help water system owners and operators manage water facilities more efficiently based on large-scale data.

The following aspects are within the scope of this document:

— management of data as an asset in water systems;

— data management principles and guidelines;

— people organization in relation to data management.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 24513, *Service activities relating to drinking water supply, wastewater and stormwater systems — Vocabulary*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 24513 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
data
set of values of qualitative or quantitative variables

[SOURCE: ISO 21378:2019, 3.1]

3.2
artificial intelligence
AI
branch of computer science devoted to developing *data* (3.1) processing systems that perform functions normally associated with human intelligence, such as reasoning, learning and self-improvement

[SOURCE: ISO/IEC/IEEE 24765:2017, 3.234]

3.3
Internet of Things
IoT
infrastructure of interconnected entities, people, systems and information resources, together with services, which processes and reacts to information from the physical and virtual world

[SOURCE: ISO/IEC 20924:2021, 3.2.4]

3.4
data flow
movement of *data* (3.1) through the active parts of a data processing system in the course of the performance of specific work

[SOURCE: ISO/IEC 2382:2015, 2121825, modified — Notes to entry removed.]

3.5
privacy
right of individuals to control or influence what information related to them may be collected and stored and by whom that information may be disclosed

[SOURCE: ISO/IEC TR 26927:2011, 3.34]

3.6
digital twin
digital asset on which services can be performed that provide value to an organization

Note 1 to entry: The descriptions comprising the digital twin can include properties of the described asset, industrial *Internet of Things* (3.3), collected data, simulated or real behaviour patterns, processes that use it, software that operates on it, and other types of information.

Note 2 to entry: The services can include simulation, analytics such as diagnostics or prognostics, recording of provenance and service history.

[SOURCE: ISO/TS 18101-1:2019, 3.9, modified — Example removed.]

3.7
data governance
property or ability that needs to be coordinated and implemented by a set of activities aimed to design, implement and monitor a strategic plan for *data asset* (3.10) management

Note 1 to entry: More information on data governance can be found in ISO/IEC 38505-1.

Note 2 to entry: A strategic plan for data asset management is a document specifying how *data management* (3.8) is to be aligned to the organizational strategy. This term has the same meaning as strategic asset management plan defined in ISO 55000 with a data point of view.

[SOURCE: ISO/IEC 20547-3:2020, 3.7, modified — Note 1 to entry revised, Note 2 to entry removed and Note 3 to entry given as Note 2 to entry.]

3.8
data management
process of keeping track of all *data* (3.1) and/or information related to the creation, production, distribution, storage and use of e-media and associated processes

[SOURCE: ISO 20294:2018, 3.5.4]

3.9
data quality
degree to which a set of inherent characteristics of *data* (3.1) fulfils requirements

[SOURCE: ISO 8000-2:2022, 3.8.1, modified — Note 1 to entry removed.]

3.10
data asset
set of *data* (3.1) items, or data entities, that have a real or potential benefit for an organization

4 Data as an asset

4.1 General

Effective use of data can create significant value. Therefore, data should be considered as an intangible asset which should be managed in a manner consistent with the principles of ISO 55001 to maximize the value to the organization.

Although data acquisition, storage and display devices are physical assets, data is immaterial and can be easily copied, modified or corrupted. As a consequence, data should be effectively stored, maintained and protected to prevent the following:

— misinformed decisions with potential social, environmental and economic impacts;

— misinterpretation of asset behaviour and misalignment of asset maintenance timing with actual needs;

— an inability to effectively harness benefits from automation or other digital technologies.

Before implementing a process of data collection, the requirements for this data should be clearly stated (e.g. precision, timestep, freshness, storage duration, environmental impacts). It is recommended that only the necessary data is collected and stored.

Data flows with no use case should be avoided. A good practice is to pre-process the data close to the data source before generating new data flows and storing the data.

The data life cycle should also be considered. The end-of-life data should be destroyed properly. Retention time of collected data should be considered in order to organize data destruction when required.

For water systems, the same data asset can be required by different stakeholders. One effective practice to deal with this situation is to catalogue and share data, which will prevent other stakeholders from organizing and capturing the same data again. Cataloguing data is a way to inform stakeholders that the data is available without needing to share it automatically.

4.2 Data quality

High-quality data can improve production efficiency and drive the company’s business to a higher level. The information about the quality of the data is valuable for making decisions on the practical application of the data. For water systems, high-quality data is beneficial for enhancing system performance, reducing operational costs and making strategic decisions. Therefore, data quality in smart water systems should be analysed and measured.

Data quality control is part of data governance. According to the application practice, the requirements of data quality control should be assessed and the procedures to implement data quality assessment should be defined and applied. The data quality control and validation process can be conducted automatically or manually. Figure 1 shows an illustration of a data quality check process.

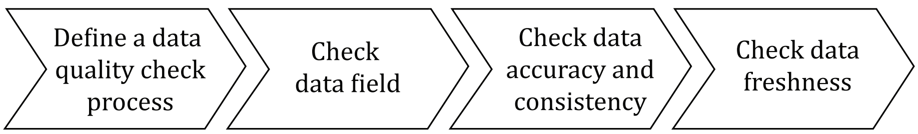

Figure 1 — Example of a data quality check process

The consistency of the data compared with other values of the time series or consistency with other related parameters should also be checked.

During the data quality check process, experts can also be consulted to confirm the validity of the data.

Before data is used in engineering or management practices, quality checks should be performed based on the quality requirements and the defined data quality check procedures. According to the quality check results, proper data quality filtration or data processing methods should be applied to the downstream analysis and applications.

5 Data management

5.1 General

Data management rules and data governance should be defined and implemented across the organization from top management to field operations, with active involvement from the senior management team. This also means setting a clear data strategy with regular reviews and action plans. Clear rules should be implemented across the organization to ensure data accuracy and consistency, data validation and curation, and data integrity and security. Some rules are given in 5.4.

The data flow should be monitored and maintained for the different users of the data (humans or machines).

5.2 Data governance

Digital technologies can be applied in all entities of water companies or organizations and the data generated by these digital technologies are important assets. Therefore, a strategy should be defined at the top level of the organization (e.g. executive committee). Someone can be appointed to be responsible for this, such as a chief data officer (CDO). This person can interact with all departments of the organization, including the information technology (IT) department, to implement the data governance strategy. More information about people organization can be found in Clause 6.

Data governance of smart water systems should be consistent with the overall data governance of the organization. Effective data governance is necessary to ensure that data ownership is clear and the data management strategies are consistent across all departments of the organization. The key aspects of data governance are as follows:

— Data ownership: the definition of ownership roles should start with the appointment of a manager to take accountability within the organization for the accuracy, quality and management of the data. Based on this, other roles for data accountability can be defined.

— Data strategy: the data strategy should outline the policies and guidelines for all aspects of data management in the organization, from data generation to disposal. The protocols for regular review of the data strategy should also be established.

— Data rules: clear rules, supporting the data strategy, should be defined to ensure data accuracy, consistency, validation, curation, integrity and security. Examples of these rules are provided in 5.4.

5.3 Data value chain

Effective data management is an essential element for extracting maximum value from the data. To make the value extraction process more effective, data management should be implemented at all levels of the organization.

For a smart water management organization, data collected by different sectors should be made accessible to other sectors when necessary, and aggregated data should be made available to administrative sectors. It is important to ensure data security is checked during the data access process.

The data collected locally can be used spatiotemporally at the collection site and at upper levels of consolidation for real-time calculation or data analysis spanning various time frames. It can also be processed (possibly combined with data from other sources) and potentially used under different spatiotemporal scenarios. Data value can be augmented with data transfer and data transformation to fit different use cases (see Figure 2).

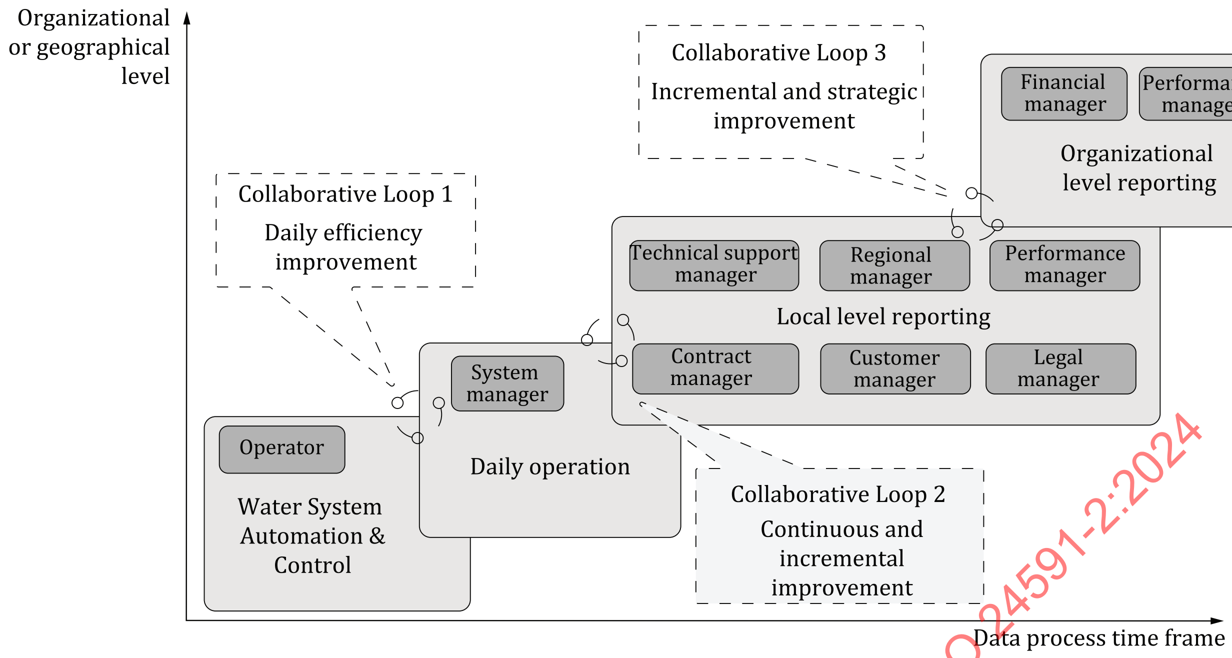

Figure 2 — Illustration of the data value chain

As shown in Figure 2, the raw data collected from water systems can be processed and used at different levels to generate value at different times as follows:

— Raw data serves to control processes in real or near-real time. It helps deliver the services and is used by data analysts and data scientists for modelling, simulation, predictive or prescriptive analysis.

— The same data aggregated on an hourly or daily basis will help the plant or network manager to operate the assets they are responsible for through daily performance indicator (PI) follow-up.

— The same PIs are used on a monthly, quarterly or yearly basis and consolidated at different geographical or organizational levels for different usages; they can be used for internal or external reporting, contract management, transparency, performance analysis and environmental commitments,.

A strong data acquisition chain, including quality validation, should be implemented and maintained across the data chain to ensure that each level takes full advantage of data usage and delivers the right value at the right level.

The facilities and devices related to water systems deliver values only if the data they generated are right and meet the minimum level of data quality requirements. Decisions made through automated processes or help-decision tools can be strongly affected by incorrect data. To illustrate this case, the flow measurement on raw water inflow into a water treatment plant can be used as an example.

Local operation teams need real-time value and information because the operators need to know exactly what the situation is in order to make appropriate operational decisions.

Local equipment also needs real-time information to potentially pilot the treatment process, manage pumps and adjust chemical dosing systems.

At a higher level, an aggregated value of inlet flow can be required for reporting issues (e.g. daily volume entering the plant) or performance evaluation issues (e.g. ratio calculations, such as daily water loss in the treatment process considering raw water, treated water and stored water). In this case, erratic values or outliers should be removed from the time series and sometimes replaced by estimated values to enhance a consistent indicator.

It is recommended that this process is documented and auditable, in particular when the calculated indicator is mandatory.

This process can be evaluated through the scoring process, e.g. 5 % of the aggregated time series has been extrapolated (default of the raw measure) and then the daily inlet flow has a reliability of 95 %.

In addition, for the calculation of indicators, a rule can be established to consider the scoring of each data set used in the calculation of the indicator.

5.4 Basic rules for data management

5.4.1 General

The following subclauses present some rules that are easy to implement and follow in order to build a consistent data management discipline for smart water system management.

5.4.2 Data property, security, privacy and confidentiality

Water system owners (e.g. a municipality that owns a drinking water network) and water system operators (e.g. a municipality or private company that operates a municipal water utility) capture or generate data. These stakeholders are proprietors of the data and they need to ensure data security, privacy and confidentiality.

The following practices can be considered to ensure data security, privacy and confidentiality:

— forming a team to deal with the data management issues;

— assigning different data access permissions to different people;

— applying encryption storage.

5.4.3 Data acquisition

5.4.3.1 General

For water systems, data can be acquired from different sources, e.g. measured automatically by sensors, manually recorded, analysed in a laboratory, extrapolated from indirect measurements, captured from internet of things devices and pushed directly to a remote platform.

No matter which sources are used, the data quality and accuracy should be guaranteed.

5.4.3.2 Data quality management

5.4.3.2.1 General

The quality of data can be impacted by the data acquisition methods. For example, some data is manually measured by operators with specific devices in situ. In such cases, errors can be introduced when the data is acquired and entered into computer systems. Moreover, the data quality can also be affected by the sensors and equipment when the data is acquired automatically.

5.4.3.2.2 Choice of sensor technology

The sensor needs to detect the parameter of interest at a level of accuracy reflecting relevant changes.

For example, in the raw water clarification of a drinking water treatment plant, the turbidity can reach 5 000 NTU, while the turbidity of treated drinking water is often lower than 5 NTU. In both cases, the choice of sensor technology with a suitable measurement range is key for detecting relevant changes and providing accurate data.

5.4.3.2.3 Location of the sensor in process or networks

Some recommendations for sensor installations should be adopted (e.g. electromagnetic flowmeter with upstream and downstream pipe configuration).

Data source unicity is recommended (i.e. data of one parameter can only be generated from one source). All data is attached to the corresponding sources, which gives contextual information to the data. Data for one parameter coming from different sources has different characteristics and accuracy. If one parameter can be measured with different methods (e.g. NH_4^+ can be measured by both sensor and lab analysis), the data should not be aggregated even if one can be used to control the result of the other. If a calculation is required (e.g. aggregate), new data with new characteristics and accuracy will be generated.

5.4.3.3 Data accuracy management

The accuracy component relates to a few things in terms of data acquisition:

— maintenance of the instrument;

— calibration and ongoing verification of the results;

— validation of the results with an alternative measuring device.

5.4.4 Data contextualization

Data without context have limited value. Context provides information about when and where the data were collected, and which other data were collected at the same time. This provides a more complete picture of the system conditions where the data or a particular parameter were collected.

This concept is key to maintaining the value of the data along their lifetime, and when sharing the information with other stakeholders who will possibly not have the same level of information on the origin of the data.

According to the applications of the collected data and the size of the system, the data contextualization can be more or less complete.

For large systems, a feasible way to ensure contextualization is to attach metadata to data. The most critical metadata are the date, time and location of the collection of a specific datum. Contextual information about the sensor is necessary to capture the full value of the measurement (e.g. its ID, its accuracy, its functional location, the rate of sampling and data transmission, and the date of the last calibration). This information then allows linkages to be established among data points in space and time.

To maximize the long-term value of data in system management and operation, the data should be clearly named and tagged with metadata. This facilitates compatibility with newer technologies to harness the data, including artificial intelligence (AI) and digital twins.

5.4.5 Data referencing

One of the main challenges when setting up a smart water system is the heterogeneity of the data sources and the heterogeneity of the data-processing technologies. These issues are common when data is generated from different brands, types and releases of acquisition systems and processed by different people at different times with different programs or software.

In water systems, it is common to use a mix of old and new technologies depending on renewal rhythms and budget allocation. A key consideration is ensuring good identification and referencing of data as it is transferred from the operational technology (OT) to the IT systems.

A data repository or data dictionary should be used to transfer data from OT to IT systems and transfer data among different sections of the organization. This approach will greatly assist in delivering value to the organization at minimum cost.

Another effective way to deal with data heterogeneity issues and make data sustainable is to ease the use of existing data instead of performing data recalculation.

At the level of a water system, a reference list for functional objects and referential definitions should be established and used as much as possible by the different applications involved in the water system. This allows a better understanding of data and eases the share and use of the data by other systems, and potentially other stakeholders or services.

5.4.6 Standardization

To process data from multiple, disparate sources efficiently, users should maintain consistency when presenting and transferring data. To achieve this, clear data standardization rules should be established. These rules should include:

— ensuring that all data has consistent naming across the organization, either through enforcement of a consistent convention or conversion of data into the enterprise convention using a decoder that is used by all parts of the organization;

— using the right units at the right stage of the data management process:

  — at the data acquisition stage, raw data is operated with no unit conversion required;

  — at the data storage or data processing stages, raw data should be standardized and their units should be converted in line with the International System of Units so as to be successfully combined and to make calculations easier (see Figure 3);

  — at the data presentation or publishing stage, data has to be meaningful in order to be efficiently used and understood, therefore data should be converted to the most suitable units.

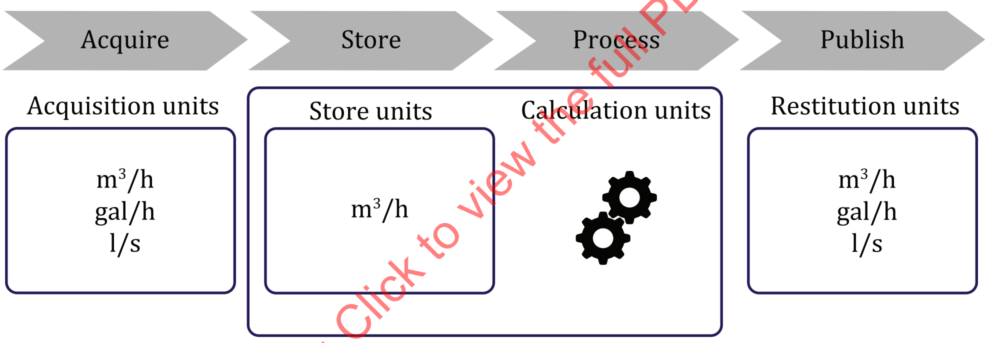

Figure 3 — Example of a unit conversion process during data acquisition, storage, processing and publishing

5.4.7 Data normalization

Data normalization is the process of converting data collected in relation to a particular outcome into meaningful information. For example, to determine the outcome "measurement of flow in a channel", it can be necessary to aggregate data on water depth or channel dimension or sum data from a number of flow meters. If the data is not in the desired units for analysis, normalization should be applied to convert all data into consistent units.

Normalization allows detailed data analysis to be undertaken, including correlation, aggregation and regression.

In the data normalization process, it is important to ensure that data is time-aligned, i.e. that data collected in times that are suitably close together can be aggregated. The data governance rules should stipulate what are acceptable time periods for aggregation.

Time sequencing (measurement and transmission) is an element of normalization. The time step of a measurement should not match the normalized time step expected (e.g. measure every minute where expected value is an hourly value) or transform a local timestamp to a universal timestamp.

Implementing these calculation steps produces new data that are closely aligned with the data source while adhering to established referencing and standardization guidelines (5.4.6).

When the calculation is performed using data from different sources, the data values should be measured at the same time. The flow calculation example in Figure 4 illustrates this time alignment.

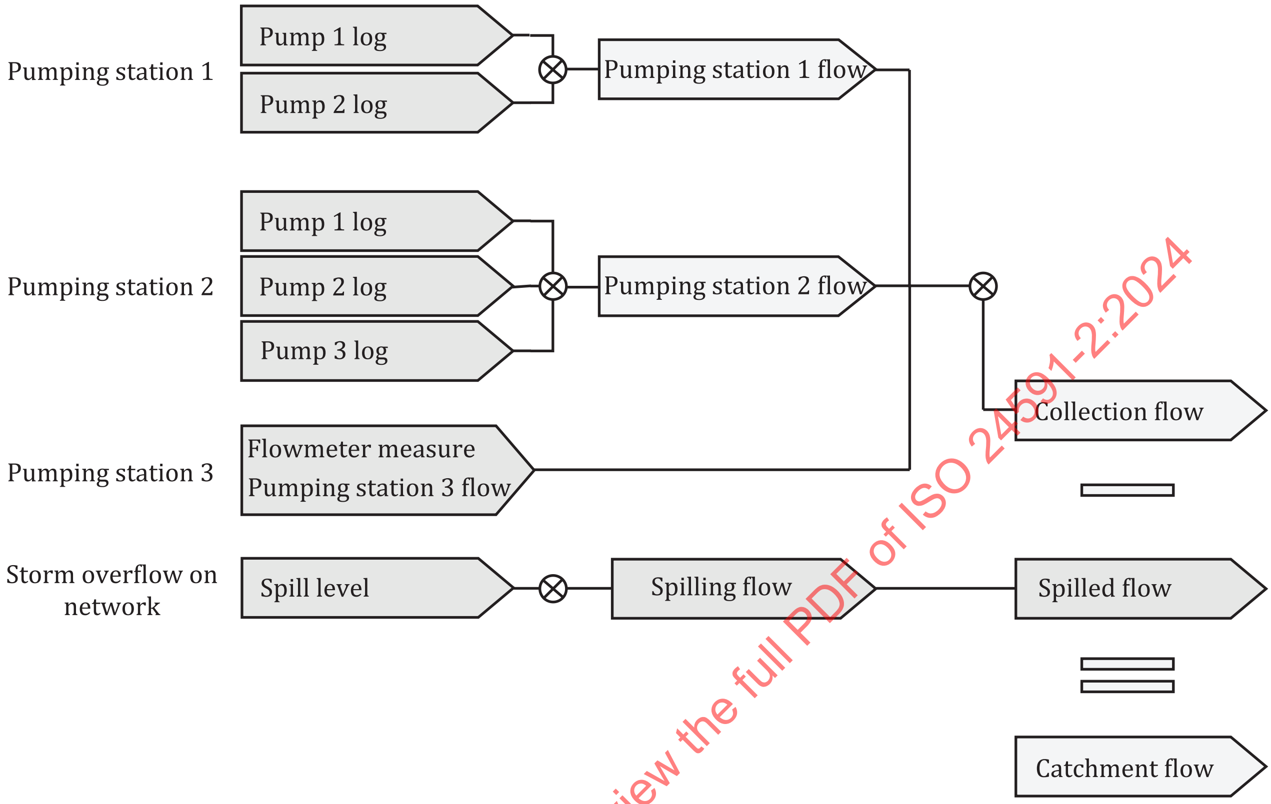

Figure 4 — Calculation of conserved flow at the output of a sub-catchment with heterogeneous raw data sources and timestamps

5.4.8 Pre-processing (including data validation)

Pre-processing data is the automated review of data prior to aggregation and analysis. The process takes raw unprocessed data and uses simple analytics to classify data to different categories (e.g. valid, doubtful, false or misleading). This process then allows for the correction of the data either manually or automatically.

Pre-processed data is then clearly tagged with appropriate metadata so that only valid data is used in future analysis.

Considering Figure 1, implementing a smart water management system allows users to:

— access and monitor their raw data;

— make some correlation between data through visualization;

— tag data with different labels (e.g. valid, doubtful, false or missing);

— correct data (e.g. manually or automatically).

The data correction can include removal of outliers or abnormal values or gap filling with traceability. The methodology used for the pre-processing is driven by expertise. Specific data pre-processing methods should be defined according to the data characteristics and application scenarios. The data passed through pre-processing can be labelled with a quality tag and the tag can be inherited in the downstream data analysis.

6 People organization around data

As mentioned in 5.2, effective data management should be driven by a senior manager or preferably a member of the executive leadership team of the organization. They should champion clear and consistent implementation of data rules across the organization.

A CDO or someone in a corresponding role should also be appointed to define, manage and maintain the proper implementation of the data management strategy as well as data ownership and data security. The CDO can either be alone or in charge of a data management team, depending on the size of the organization, the contractual requirements and the level of awareness of the organization regarding data management.

The data management strategy defined by the CDO and validated by the executive leadership team should clearly articulate roles and responsibilities for data management at all levels of the organization.

For small and medium organizations, the role of CDO can be assigned to a member of the management team, e.g. the chief information officer or someone in a corresponding position.

For large organizations, senior specialists can complement the CDO to work out some of the key challenges, such as the compliance of data management. They can help the CDO to define and monitor the data governance strategy.

In the IT department, a specialist team responsible for the implementation of the strategy should be formed. Depending on the size of the organization, one specialist in the team can play multiple roles.

For a specific object (e.g. data set, referential, data model), a data owner can be nominated as the temporary manager for this object.

The CDO can be in charge of the following topics:

a) Conformity:

   1) Contractual framework: this person serves as a leading authority on data privacy, sovereignty, stakeholder rules and regulations, and local and national regulations.

   2) Data archiving management: this person takes charge of identifying all demands in terms of data archiving. The CDO should ensure these demands are respected and take action in case of breaches. They should also be in charge of managing the data life cycle and destroying end-of-life data.

   3) Open data management: this person takes charge of maintaining the accessibility of data for all stakeholders through an open data platform. Actions should be taken to monitor this accessibility and deal with failures.

b) Data reliability:

   1) Data storage strategy and operation: data storage is managed according to specifications associated to each data set and demands from the data owner or the contract. This storage takes into account the requirements in terms of accessibility, privileging low-energy consumption architectures for not frequently requested data. This person also manages the destruction process when the data reaches end of life.

   2) Database architecture strategy and operation: this person organizes and maintains the architecture of the database to fulfil the objectives of data storage and usage.

   3) Referential strategy and operation: this person promotes shared data referentials so that all stakeholders share the same data referentials and data produced by each stakeholder can be used by others. They are in charge of updating the referentials according to new needs and rolling them out.

   4) Data quality strategy and operation: this person takes charge of maintaining the expected level of quality for all data (different data sets can have different quality requirements). They take actions if deviations are observed in data produced.