

White Paper

Best Practices in Water Data Management

Towards a European Water Data Ecosystem

Version 1.5

October 23rd, 2024



Summary

The context for European Water Data is changing rapidly. Digitalization of the Water Sector is perceived as a key lever for addressing the challenges facing the European Water Sector, maybe even the lever. At the same time, Water Data are scattered, difficult to find and access, and in many cases not interoperable and reusable without significant transactional costs.

A solid framework for a European Water Data Ecosystem is therefore highly needed to sustain a foundation for a sector-wide digitalization to achieve the aims of a Water Resilience Strategy as outlined in the Mission Letter to Commissioner-Designate for Environment, Water Resilience and a Competitive Circular Economy.¹

Importantly, the Mission Letter adds utility data to the picture. This allows for data from the entire water cycle to be included in important Decision Support Systems e.g. Integrated Urban Water Pollution Management, Cloudburst Management, Source to Sea Analysis and Planning, Cross-border Ecosystem Management, Environmental Impact Assessment, Flood Risk Management and Flood Warning Systems.

Data as the raw material of digital tools should be seen as a strategic asset and given priority attention at leadership level, which, however, by far is not the case in all organizations. A comprehensive communication effort is needed. Written in a non-technical language, this White Paper aims to contribute to this effort.

The Water Sector is categorized by having a rather decentralized structure and a strong focus on specific sub-domains. Most water data management systems of utilities, cities, industries and river commissions function independently, e.g., compared to the heavily interconnected electricity and road infrastructure systems.

This sub-domain focus pays limited attention to interactions across the water cycle. Hence data related to ground water, surface water, wastewater and the marine are often structured without a systemic focus on data sharing between these sub-domains, thus preventing interoperability. The vision is to create a data ecosystem which follows the natural ecosystems, allowing for Decision Support Systems to cover the scale of these ecosystems.

The potential for value creation is easily measured in hundreds of millions of Euros: reducing costs and time of public sector work processes; improving Technical, operational and Production Processes as well as construction works; substantially increasing solutions and knowledge of researchers and innovators due to access to relevant data; and, at the core of matters, improving Environment, Climate, Quality of Life.

Data is the foundation for making good decisions and in respect to water it is often data over time which allows decision makers to make the right decisions. However, data is often not treated as a business asset nor as a national asset. Hence it is often left to technical people building and collecting data and deciding on which data are important. Often, they only take a sub-domain perspective and don't see it from the perspective of the featured problem – the river catchment, the whole urban area, sometimes not even the whole utility.

This implies that for the really big decisions where you have to go across various sub-domains it is hard to get access to the right data. Willingness to share data constitute a challenge but combining data across organizations often pose a critical and very costly challenge.

This is not just a problem at the organizational level but also at the national and international level, where different government departments and international organizations set different standards for basically the same

¹ https://commission.europa.eu/document/10a1fd18-2f1b-4363-828e-bb72851ffce1_en

data. E.g. most governments and international organizations do not have unique identifiers on chemical compounds sometimes they even just publish the name without reference to a classification.

Interoperability has so far mostly been handled as an add on to the data collected. Hence the focus has been to develop standards for interoperability on top of data collected, but only limited attention given to ensure that data is produced in an interoperable manner. The result is, e.g., that Chemical compounds collected by different actors are subject to interpretation before they can be shared. An interpretation which can be difficult to make as the area is complex. A real-life example is a government agency with highly skilled experts reported the same chemical compound with three different names in an official document – not being aware that it was the same.

At its core, this White Paper offers a vision that *public authorities and private actors should share, use and reuse observable and measurable data in a standardized, harmonized and uniform way.*

While getting agreement to the above statement might be simple, translating it into real action will be difficult. This is due to a number of factors. Though organizations may agree to the vision of data sharing in principle, data sharing comes with a cost. Hence, organizations tend to shy away from taking on these extra costs. By making the data interoperable at the production of data these costs will be reduced as the need to interpretate is reduced.

Water4All call for a European Water Data Ecosystem to focus on lowering the cost of sharing data by ensuring that data is collected as uniformly as possible. This could, e.g. be based on on the ISO standard for “Observations and Measurements and Samples”.

Further, Water4All call for making the registration of data across various sub-domain more aligned and less complex. A domain driven approach where different sub-domains such as groundwater, surface water and marine water could use very similar or perhaps even similar data models for physical parameters, chemical parameters and biological parameters should be explored. If data models across sub-domains could be made more uniform this would increase interoperability and make the IT infrastructure less complex and hence cheaper.

At the dawn of digitization, collection of data was mainly driven by the research community. This role was gradually taken over by utilities and in particular public monitoring, not least with the endorsement of legislation like the Water Framework Directive, the Drinking Water Directive and the Wastewater Directive.

In recent years, however, we have seen a new change in the primary users of water data. Private sector companies have emerged providing sophisticated digital services improving water data management. However, the lack of interoperable data and uniform data standards act as an obstacle to the scalability of these solutions.

The public sector is increasingly using water data as an integrated part of major economic decisions. An example of this is the use of water data in environmental impact assessments. Here the lack of uniform water data slows down the whole permitting process delaying new plans and projects related to the Green Transition.

Recalling the decentralized character of the water sector with data management systems being developed in many different ways, vested interests are deeply rooted and will not make it an easy task to form a European Water Data Ecosystem. The White Paper therefore puts forward a number of detailed recommendations, divided into Organizational and Political, Technical, Cultural and Resource sections, all of huge importance.

Strategically, it will be crucial to bring regulatory steps to the forefront and to have public authorities leading the way. Technically, giving importance to implementation of standards will be necessary. In a cost-efficiency context, ensuring interoperability should especially happen by making data harmonized at the data production levels.

An example is achieving interoperability regarding chemical parameters by digitalizing relevant EU directives and providing access to a shared code list with unique identifiers and semantics for the chemical parameters.

About this white paper

This white paper aims to introduce the principles behind a data sharing ecosystem in Water4All and the benefits we can achieve. The paper also describes the barriers we are facing and provides recommendations to overcome them. Important to notice the data management principles suggested can be widely used in many other areas. This is important as issues around water interact with many other areas, and where other areas will have impact on water. In that way we strive for a holistic approach to all topics that will influence our water theme.

The paper is deliberately intended to be non-technical, as the objective is to create leadership attention, eventually at a scale of the attention leadership pays to economic management.

In brief the content of this paper is:

- Preface
- A vision for a European Water Sector Data Ecosystem
- The potential in sharing data
- Barriers and problems in data sharing
- Introduction to the concept for data management and data sharing
- Recommendations for the next steps

In order to keep the white paper short and non-technical, we have an annex with

- Short introduction to Water4All
- examples of use
- technical references

The White Paper has been produced a Working Group under the Water4All Data Sharing Task, led by the Danish Environmental Portal and the French Geological Survey. Key elements of the White Paper have been presented to Water4All governance bodies – Pillar C Plenum, Executive Board, Governance Board and Advisory Board.

With this version 1.5, the White Paper will not be distributed to a wider circle of European and International stakeholders, in order facilitate a discussion on the strategic importance of structured data and a mutual data ecosystem as preconditions for building a digitalized water sector.

Preface- Towards a European Water Data Ecosystem

With the endorsement of the EU Green Deal, a framework for significant new steps to address Europe's vast water sector challenges have been taken, encompassing increased problem awareness, increased ambitions, including targeting zero emissions, implying a focus on integrated management – and: placing digitalization at the core of the tool-box to deliver a solid response. Which in turn requires leadership attention to data management as a strategic asset – thus this White Paper!

Digitalization is perceived as a key lever for addressing the challenges facing the European Water Sector, maybe even the lever. Accordingly, data as the raw material of digital tools and also the perspectives in terms both of substantial cost reductions and the potentials for value creation related to data sharing should be given strategic attention at leadership level, however this is by far not the case in all organizations. Instead, focus is based on investments in new facilities, infrastructure, IT software not realizing that investments in data would have a higher return on investment.

Furthermore, the Water Sector is categorized by having a rather decentralized structure and a strong focus on specific sub-domains. Decentralized as with most water managers of utilities, cities, industries and river commissions functioning independently e.g. compared to the heavily interconnected electricity and road infrastructure systems. It is sub-domain focused with limited attention paid to the systemic interaction within the water domain. Hence data related to ground water, surface water, wastewater and the marine are structured without a systemic focus on data sharing between the sub-domains. As most the adjoined data management systems can be expected to be similarly decentralized without substantial focus on interoperability of the systems.

At the same time, data and data management is often left to the data experts in various sub-sector domains like wastewater management, groundwater or surface water. The result of this is that decisions on rules, regulations and investments in water is often based on inadequate silo-based information instead of taking a holistic approach to water management.

Overall, systematic data is key to making the right investments in data systems and to further unleash the potentials for increased value creation, which include a.o.:

Public sector work processes

- **Reducing costs and time** spend by Environmental Authorities for a range of tasks including , permits and enforcement processes, management and analysis related to water pollution problems as well as EIAs (Environmental Impact Assessments) and cross-sector analysis and priorities
- **Smoother transfer of data** within national levels, as well as at international scale especially when water flows across countries, including reporting to EEA and other EU structures
- Improving **flood warning systems** and other emergency systems

Technical, operational and Production Processes

- Allowing for **better integration** of water flows and levels of rivers, cloudbursts, floods, droughts and groundwater close to the surface, altogether improving climate change adaptation
- **Reducing operational costs** for operating data management systems and ensuring harmonizing of data from different sources, leading to increased interoperability and reusability
- Overall, implementation of **better communication** lines between organizational leadership and IT developers, allowing for new IT solutions to have a stronger emphasis on value creation.

- Construction works more precise, **reducing both costs and resources** used

Research and Innovation

- Improving researchers and innovative companies **finding and accessing data**, not least results of monitoring programs, leading to strengthened relevance of new solutions and knowledge
- Development of **better digital decision support** tools based on harmonized data, including increased frequency of planning cycle, e.g. nutrient management and reduction of eutrophication

Environment, Climate, Quality of Life

- Improved environmental quality – ecosystems, water recipients, biodiversity protection
- Improved health conditions due to improved drinking water quality and pollution reduction
- Decreasing carbon and other greenhouse gas emissions
- Speeding up implementation processes for renewable energy

Data Interoperability Foundation for Decision Support Systems

Decision Support Systems using advanced digital tools are no better than the data input they are based on. In the following, examples of these dependencies are listed:

Decision Support System	Data Sharing and Interoperability dependency
Integrated Urban Water Pollution Management	Several Monitoring Entities from different government levels; Physical parameters (hydrometri) as foundation for digital twins and models, combined with chemical parameters
Source to Sea Analysis and Planning	Cross Border, both from source to sea, but also marine data across borders; Connecting Surface Water Quality and Marine Water Quality
Cross-border Ecosystem Management	Ecosystems are often functioning across administrative and national boundaries; Chemical parameters, Biological parameters
Environmental Impact Assessment	A multitude of entities, geographical information, data harvesting time consuming
Cloudburst Management, Flood Risk Management and Flood Warning Systems	Flood risk and Flood Warning involves upstream and downstream entities, geographical information; River flows and levels, groundwater levels close to surface, technical infrastructure sewers and overflows
Monitoring-as-a-Service	A harmonized and pan-European approach to monitoring methods and procedures for analysis, unambiguous reporting formats, and data management is a prerequisite for M-a-a-S

We are not there yet

Data is the foundation for making good decisions and in respect to water it is often data over time which allows decision makers to make the right decisions. However, data is often not treated as a business asset nor as a national asset. Hence it is often left to technical people building and collecting data and deciding on which data

are important. Often, they only take a sub-domain perspective and don't see it from the perspective of the featured problem – the river catchment, the whole urban area, sometimes not even the whole utility.

This implies that for the really big decisions where you have to go across various sub-domains it is hard to get access to the right data. Willingness to share data constitute a challenge, but combining data across organizations often pose a critical and very costly challenge.

Major decisions may be based on data which are not suited, leading to increased risk of taking the wrong decisions. The White Paper is therefore intentionally in a non-technical language, aiming at contributing to a wide range of stakeholders to become more familiar with the topic, especially at leadership level of organizations, thus striving for data management in an age of digitalization to be considered of strategic importance at the level of economics, human resources, and innovation.

A precondition for materializing the potential of digitalization is an adequate focus given to its raw material: data. Accordingly, Best Practices in Water Data Management must be given center stage in the Water Sector, however the importance of adequate Data Management Practices is often neglected when talking about the perspectives of digitalization.

Increasingly data is collected across their originally intended purpose, however often found in many different IT formats, structures and in different ways of documentation. This can make it challenging to work with and combine data across disciplines, and even across providers within the same discipline, and create a coherent overview of the specific object, person or place we are seeking knowledge about.

For instance, it is not unusual that governments in the EU spends months on harmonizing and standardizing data in order to prepare water management plans, reporting to the parliament, to evaluate the impact of initiatives etc. Time and money could have been spent on addressing the problems faced, if the required data was already produced using common practices.

Making data interoperable and reusable (report once, reuse many times) - allowing for increased value to society by utilizing data for more purposes than originally anticipated, comes with a cost: merging data streams derived from different data models and building complicated technical interfaces comes with even higher costs. These problems were analyzed and constituted the foundation for launch of the INSPIRE Directive.

As documented by Forrester (2019)². This is not only the situation in the water sector: In general, Data Scientists roughly spend 80% of their time mining and harmonizing data, and 20% analyzing data due to the fact that Data Systems have been implemented in a wide variety of ways.

It is not just a problem at the organizational level but also at the national and international level. Where different government departments and international organizations set different standards for basically the same data. E.g. most governments and international organizations do not have unique identifiers on chemical compounds sometimes they even just publish the name of chemical compounds without any reference to a chemical compound classification.

The result is a mess. Chemical compounds collected by different actors cannot be compared without interpretation. An interpretation which can be difficult to make as the area is complex. A real life example is a

² Insert reference link to Forrester analysis

government agency with highly skilled experts reported the same chemical compound with three different names in an official document – without being aware that it was the same.

Both within and across organizations it is a problem as water flows across sub-sectors and organizational and administrative boundaries. If data cannot be exchanged easily it makes it difficult and time consuming to make the right decisions about water management. Within the EU large-scale initiatives on data spaces are being funded. These initiatives are focusing on facilitating data exchange across organization.

In the following Water4All presents its vision for a Water Data Ecosystem. The vision is accompanied by an outline of the cornerstones of the technical solution as well as recommendations to initiate a joint, multistakeholder process towards such an ecosystem.

It is important to note that Water4All is not the only initiative to focus on strengthening sharing of environment and water data.

In Annex a non-exhaustive list of initiatives is presented. Eventually, these should be invited to take part in a mutual dialogue.

A vision for a European Water Sector Data Ecosystem

Public authorities and private actors should
SHARE, USE and REUSE
observable and measurable data in a
standardized, harmonized and uniform way

The bullets below constitute a general goal for data sharing, no matter if it is water-based data or not. All sectors working with data sharing should work with these principles.

- Data collected for a local perspective should be re-usable for many other purposes
- Data collected in proprietary formats should be shared in an open and well documented format
- Data understood in a local context should be transferred to widely understandable data
- Use of internationally agreed FAIR Water Data Standards'
- Use commonly agreed classifications lists and code lists, ideally made mandatory via EU regulation
- Overall, move away from sub-sector approaches to domain-solutions, e.g. don't make solutions for river flows, but make them for continuous measurements.

Water4All call for a European Water Data Ecosystem to build on the "Observations and Measurements and Samples" principles, already laid down in an Iso-standard, which places the object ("Something") and location ("Somewhere") at the core of matters. This is the feature, which we aim to observe.

Further, Water4All call for making the registration of data across various sub-domain more aligned and less complex. A domain driven approach where it is explored if data from different sub-domains such as groundwater, surface water and marine water could use very similar or perhaps even similar data models for physical parameters, chemical parameters and biological parameters should be explored. If datamodels across sub-domains could be made more uniform this would increase the interoperability of data across sub-domains and make the IT infrastructure to support these data far less complex and hence cheaper.

An example is achieving interoperability regarding chemical parameters by digitalizing relevant EU directives and providing access to a shared codelists with unique identifiers and semantics for the chemical parameters. See Annex for more details.

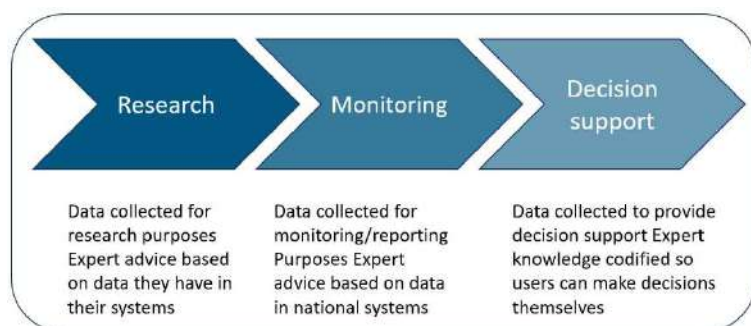
The White Paper puts forward a number of recommendations, divided into Organizational and Political, Technical, Cultural and Resource sections, all of huge importance. Strategically, it will crucial to bring regulatory steps to the forefront and to have public authorities leading the way. Technically, placing even more importance to implementation of standards will be necessary. And in a cost-efficiency context, ensuring interoperability should especially happen by making data harmonization happening at the data production levels.

Probably, achieving interoperability regarding chemical parameters may be easiest via calling for shared codelists and uniform procedures to be used at the laboratories analyzing water quality samples.

Other international structures have defined similar principles, e.g. the US “InternetOfWater”³ (see Annex). Further, several international standards exist, which – if implemented – on the technical side could constitute strong guidance for implementation of best practices in water data management. Regrettably, use of these standards are, at best, scattered, and will have no coherence to regulatory steps as well as addressing organization, cultural and resource barriers. For further details on the barriers, see next chapter.

The value chain for data

In different steps of the life of data, they are used for various purposes. It is our aim that research data (most commonly used today) will also serve as base for the next two steps. For monitoring and especially Decision support one source of data and one theme of data cannot stand alone. For Decision support several data themes should be brought together to form the basis for good decisions.



Barriers of Data Sharing

As mentioned, data sharing is important. However, there is a wide range of barriers we need to face in the project. Historically the water domain has been handled in slim, technical “silos” that primarily has been addressing the needs of the professionals working in a specific “silo”. There has been very little use of a standardized approach both in the data collection and in the applications, collecting the data/results.

- **Organizational and political barriers**
 - **Top management does not pay attention to their organizations data assets.** Data models are like IT architecture – Top management leave it to the experts
 - **The value of data is not monetarized**
 - **To share data is not an “emergency” political agenda item**
 - **It is the organization using data which gets the benefits not the organization sharing data.** There are virtually no incentives or regulatory structures in place to enhance data sharing, except reporting requirements – which in turn rarely are aligned, but rather requesting different formats.
 - **Lack of willingness to share data.** Many organizations and departments in organizations are hesitant to share data as they perceive that other organizations should just look at their website, not realizing that the value of the use of data is far bigger when data is combined. Some are afraid others may do business with ‘their’ data – also called “my precious syndrome”

³ www.internetofwater.org

- **Political and Institutional Factors:** Political differences and institutional barriers, such as bureaucratic processes, competing priorities, and jurisdictional disputes, can hamper efforts to establish common data sharing frameworks and protocols. Lack of political will and leadership may also hinder progress in this area.
- **Cultural barriers**
 - **Cultural and Linguistic Differences:** Cultural and linguistic diversity within the EU can pose challenges to effective communication and collaboration among stakeholders. Language barriers may hinder the sharing of data, information, and knowledge, particularly among regions with different linguistic backgrounds.
 - **What's in it for me?:** Data sharing and providing data in a harmonized format will require resources and economy. Resources that are used for others outside “my project” to understand “my data”. The political pressure, commitment and investment is crucial to make it happen.
 - **Lack of Trust and Transparency:** Concerns about data accuracy, reliability, and misuse may undermine trust among stakeholders and discourage data sharing. Building trust through transparent processes, data validation mechanisms, and accountability measures is crucial for promoting collaboration and cooperation in data sharing initiatives.
- **Resource barriers**
 - **Resource Constraints:** Some Member States may lack the financial, technical, or human resources needed to develop and maintain robust data sharing mechanisms. Limited investment in data infrastructure, capacity building, and training can impede efforts to improve data sharing across borders.
 - **Knowledge constraint:** many domain (water quality, hydrogeology, etc...) do have some data handling training in their curriculum (Database or GIS courses) but not enough effort is put into registration patterns, cross sector data handling etc. (e.g. the INPIRE directives). This makes it harder for organizations to hire and put the right profile in front of their data challenges; leading to an even more the fragmented landscape described above.
- **Legal barriers**
 - **Legal and Regulatory Frameworks:** Legal and regulatory barriers, such as data protection laws, privacy regulations, and intellectual property rights, can restrict the sharing of sensitive or proprietary data across borders.
- **Technical barriers**
 - **Data models in the production systems are too complex** and not aligned across sub-domains
 - **Fragmented Data Systems:** EU Member States often maintain separate data systems for water monitoring, management, and reporting. Lack of integration and interoperability between these systems makes it challenging to share data seamlessly across borders.
 - **Diverse Data Standards:** Different member states may use varying data standards (local, national, EU, international), formats, and protocols for collecting, storing, and sharing water-related data. Incompatibility between these standards complicates efforts to harmonize data and facilitate cross-border sharing.
 - **Technical Challenges:** Technical issues, such as outdated infrastructure, limited bandwidth, and incompatible software systems, can hinder the exchange of large volumes of data across borders. Ensuring compatibility and interoperability of data systems is essential for overcoming these challenges.

- **Data is a domain of expertise by itself** but data is often delegated to IT teams. But IT developers, architect, UX/UI experts are not often Data expert thus not inform about practices to be applied and the current legislation (especially in the environmental domain)
- **Limited access:** No access for researchers, innovators and service providers to the large data volumes of end-users (utilities, river commissions, national monitoring programs). Data Owners and Data End-users don't see themselves as part of a Data Ecosystem

User groups and Data Flows

As part of the activities to develop a Water4All Data Sharing Facility, user groups of water data have been identified to these two distinct primary groups:

- **Data professionals:** Researchers, developers, and suppliers are given the opportunity to collaborate and develop shared algorithms that can operate on a common data foundation. Access to data across borders and in a common language/format opens entirely new possibilities for creating innovative solutions and analytical models. Very large datasets can be managed and shared and made available for broader use. Overall, the Data Professionals are considered to have the capability of being data-driven (ability to deliver data and understand data shared in a FAIR/interoperable approach), hereby also being able to analyze data and create results (knowledge and tools) at the scale and complexity of the entire water systems, being full river catchments, urban areas, utilities, and industries.
- **Non-data professionals:** There are still numerous institutions working with unstructured data or data models developed locally and without consideration for data sharing. Water4All can support them in becoming more data-driven and structured by providing a universally applicable model and infrastructure. Solutions could utilize the data from unstructured sources (PDF's Excel-sheets etc.) and convert them to structured data, based on the agreed data model. Although working with data, in this context only having the capacity of being able to deal with small data sets representing minor samples of the scale and complexity of the problems.

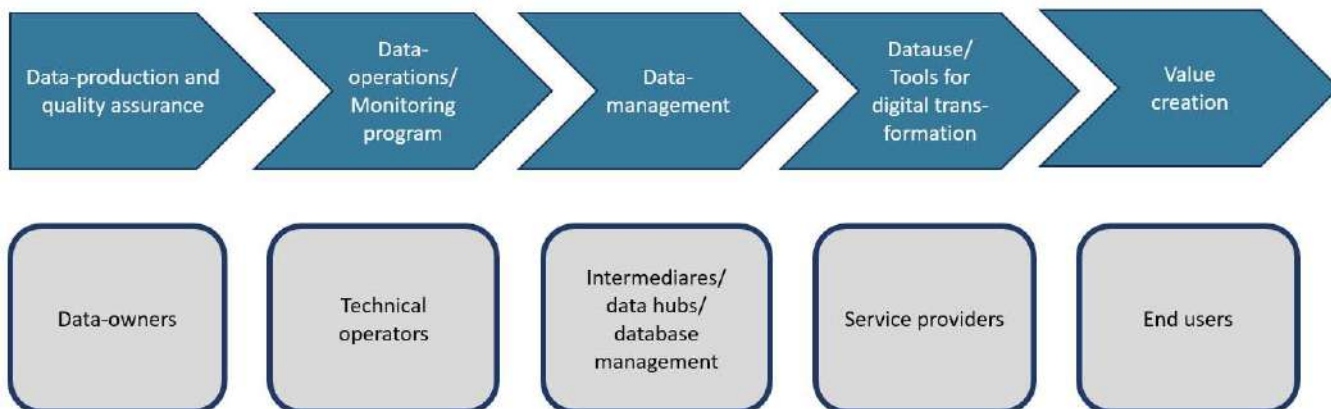
Secondary user groups could be:

- **Decision-makers:** Politicians and stakeholders are given the opportunity to discuss and make decisions on a well informed and fact-based basis rather than emotions. Water4All offers a holistic view on water issues that is not confined by borders, languages and cultures.
- **Non-experts:** Citizens, interest groups, data journalists, and similar entities can gain insight into water-related facts and thus be influenced to change behavior.

Different data flows, where data is combined across several different data productions – which at an overall level could be divided into three categories:

- **continuous flows** e.g. insight in national monitoring programs or utility process data
- **aggregated data** e.g. compilations at local or national level, eventually brought to an aggregated European level
- **project co-operation** with data produced in the context of a one time project in which various organisations do collaborate, eventually organized using the Data Space concept

– comes with different value chains for managing the data flows.



Data-Owners and Technical Operators

Relates to the original (and often local) data flows, where an organization (*Data Owner*) invests in monitoring stations with sensors installed to measure continuous flows and samples at regular or ad hoc frequency measuring selected parameters, and where an *Operator* undertakes operational responsibilities at technical level, including compiling the data produced.

Often the “raw”-data undergoes a quality assurance step which is also usually performed by the data-owner

Data management

When data flows are merged across several monitoring stations, an *Intermediary* is needed. In single stream data flows, the *Operator* eventually also becomes the *Intermediary*, however when data flows are merged across several different streams, an *Intermediary* is needed with a technical infrastructure allowing for this to be possible. This comes with huge costs, and the more different the original Data Models are, the more expensive this functionality becomes.

Often these costs are not taken into account by the Leadership level, but with this White Paper an approach is suggested which on one hand will lower costs and increase productivity of Data Management, but – importantly – will also contribute to increase harvesting the potential of data to be reused across different sources.

Furthermore, precisely for this role the organizational aspects should be carefully considered. Although few cases exist, the observations about these points to that either a co-organized approach eventually co-owned approach involving all government levels undertaking monitoring tasks is recommended to ensure adequate IT management to constitute the main focus of this structure, or, otherwise, a quite comprehensive regulatory framework to ensure an unbiased approach is needed. To be more direct, just placing the intermediary role as a unit under the national EPA may lead to various operational deficiencies.

Data use and value creation

This potential includes the perspectives of allowing researchers, domain experts and innovators to access **relevant data** representing the **relevant problems** of end-users - hereby increasing **relevant results** of these stakeholders. Eventually end-users e.g. authorities and utilities should have a substantial interest in this, but will also have to carry the investments, or parts of these.

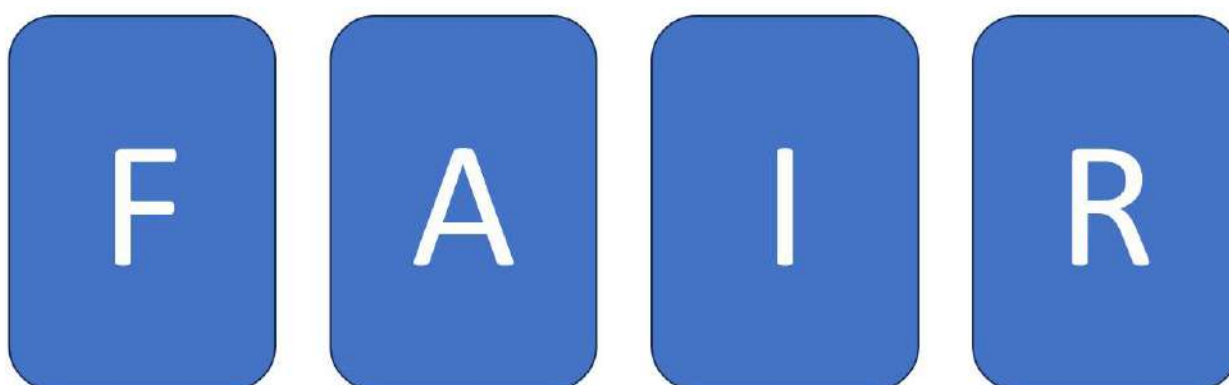
E.g. when it comes to make publicly generated data, e.g. national monitoring data, accessible the public entities eventually will have to finance forming of an *Intermediary* with the necessary, data knowledge and technical infrastructure to receive and exhibit data from different sources.

The FAIR-principles

The dynamic around FAIR principles stem from the 2016 publication ‘The FAIR Guiding Principles for scientific data management and stewardship’⁴. Since then, there is an international incentive towards applying those principles in the global scientific community.

As mentioned in the reference literature⁵, the “*principles emphasise **machine-actionability** (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data*”.

At its core of this approach are the FAIR principles, which are data which meet principles of findability, accessibility, interoperability, and reusability (FAIR) for both Human AND machines.



The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

More information about the FAIR-principles can be found in Annex.

Making data HUMAN understandable

The FAIR principles aim to enhance the usability and effectiveness of data sharing efforts across different domains, including scientific research, industry, and public administration. As Water4All operated cross-country, cross-

⁴ Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

⁵ <https://www.go-fair.org/fair-principles/>

languages, cross-cultures and cross domain, we also need to put an extra focus on the human aspect and make data **understandable**.

Making data understandable involves ensuring that users can comprehend the meaning, context, and significance of the data. Achieving understandability involves the following practices:

- **Clear Documentation:** Provide detailed documentation that explains the structure, content, and semantics of the data. Include descriptions of variables, units of measurement, data collection methods, and any assumptions or limitations.
Metadata of the shared observable properties should be described according to the principles of “Observation and measurements” and displayed in shared vocabularies.
- **Standardized Formats:** Use standardized data formats and conventions to organize and present the data in a consistent and interpretable manner. The standardized formats should be shared and used across member states and domains. In Water4All we use the standards described in OMS (Observation, Measurements and Samples)
- **Contextual Information:** Provide contextual information, such as background knowledge, metadata, or links to related resources, to help users interpret the data within its broader context. This could include references to relevant literature, research articles, or explanatory materials.

A down to earth example of the FAIR principles in use:



An Italian guy and a Chinese girl wants to communicate...

Findable:

the first thing they need is how to reach each other. They could look in the phone book or search the internet. In order to succeed they also need to be listed there - making each other findable.

Accessible:

When they call the number there has to be a working infrastructure that actually make the phone ring in China when the Italian guy dials the number. They also both need the proper equipment (a phone) that understand the technical protocols. Also, if passwords or something like it is needed, they need to know the key to be able to access.

Interoperable:

The Italian guy does not speak Chinese – and the Chinese girl does not speak Italian. So even if they managed to both Find and Access each other – they cannot communicate. They are not interoperable. So either they learn each other’s languages or the use a common language e.g. English. If they do not know all the word, they look up in the dictionary that translates Chinese into English and Italian into English and vice versa.

Reuseable:

We all know the first versions of Google Translate where you could get something completely strange (and sometimes funny). The reason for that is that the words was just translated one by one without any context.

In order to actually reuse the information received in the phone call, they also need a clear definition of the words and the phrases should be understood in the right context, as described in the next chapter.

The “Observations, measurement and samples” concept

In our effort to make data understandable in the shared context across member states, we use the “Observations, measurements and samples” concept, which is originally described as ISO-standard (ISO 19156) and updated in 2023.

The concept is a generic model used when we need to know SOMETHING about SOMETHING/SOMEPLACE. The concept does not only apply for water properties but can be used for all kinds of data and domains.

E.g. if we want to know the waterflow of a river, we define what waterflow is, the Unit of Measure and how it is measured. Once defined, we can use that single description of the property “waterflow” for all rivers.

Following the concept, we all share the same definition of what waterflow is and its Unit of Measure – meaning that we are defining the shared language and making the results UNDERSTANDABLE.

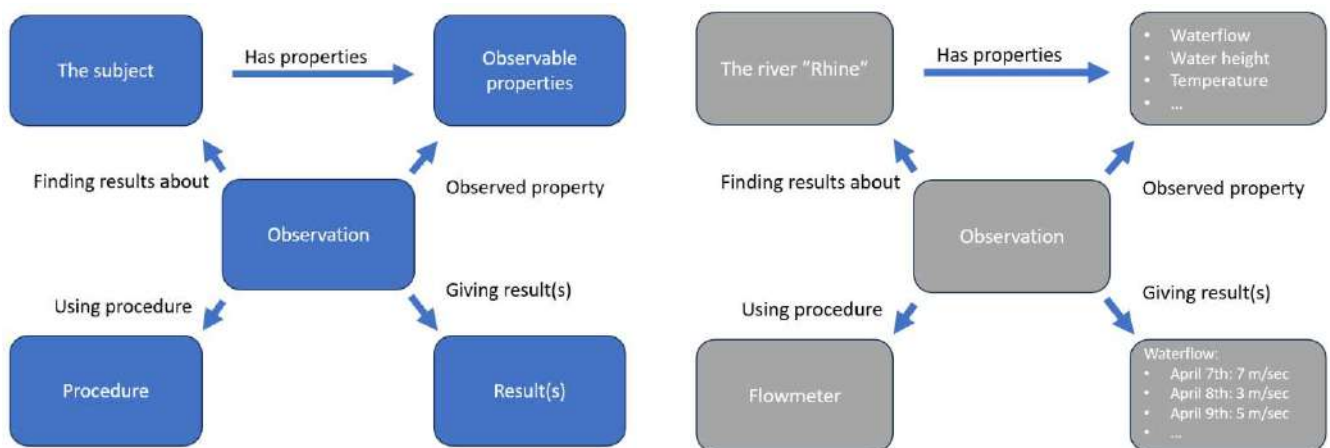


Figure 1 the concept to the left and an example of use to the right

The “Feature of Interest”

The “Feature of Interest” is the “object” we want to know something about. It could e.g. be the river “Rhine” – or part of it. Therefore, we place some sensors or other equipment in the river to give us the results.

Sometimes a sample taken from a river could be the feature of interest for the lab, but it always represents the place in the river where it was taken.

Observable properties

All physical objects have properties that we can measure. And objects of the same type share the same properties. All human beings have body temperature, blood pressure, weight etc. All rivers have Waterflow, Temperature, Water level etc.

Each observable property is given a solid semantic definition, a unit of measure, an observation procedure (description of how to measure) and maybe a constraint for the result.

Observation

When the actual observation or measurement takes place at a certain point in time or period, we are getting results within the allowed range for a certain observed property, following an agreed procedure and with of a given quality (raw, validated, ...)

Procedure

For each observable property we can define how to perform the observation and follow a specific procedure. Sometimes there are different ways to get a result, but when we do the concrete observation, we point of which of the different procedures we used.

Result

The result is the product of the observation. It could be 7. In the context of Figure 3 it is understood that it is m/sec on the 4th of April 10.07 am, on station 12 (geo position 55.66707: 12.58056), following procedure XYZ

Theme/subject	Observable property				Result		Metadata about observation		
River	Property	Unit Of Measurement	Constraint	Procedure	Date 09.04.23	Date 27.07.23	Sensor/station	Agent	Location
	Water level	Meter above sea level	0<	Hydrological Monitoring Protocols	8.5	6.3	Data logger 12	Danmarks miljøportal	55.7604° N 12.4617° E
	Water flow	Liters per second (L/s)	0-10	Hydrological Monitoring Protocols	575	324	Flow meter 1	Danmarks miljøportal	55.7604° N 12.4617° E
	Temperature	Degree Celsius	-5 - 50	Field Measurement Techniques	7	18	Temperature logger	Danmarks miljøportal	55.7604° N 12.4617° E
Lake	Property	Unit Of Measurement	Constraint	Procedure	Date 09.04.23	Date 27.07.23	Sensor/station	Agent	Location
	Water level	Meter above sea level	0<	Monitoring and Assessment of Water Resources	22	17	Station 46	Ministero della Transizione Ecologica	45.6045° N 10.5199° E
	Oxygen level	Milligram/liter	0-15	Monitoring and Assessment of Water Resources	358	412	Sensor 23	Ministero della Transizione Ecologica	45.6045° N 10.5199° E
	Temperature	Degree Celsius	-5 - 50	Field Measurement Techniques	10	22	Sensor 23	Ministero della Transizione Ecologica	45.6045° N 10.5199° E
Groundwater	Property	Unit Of Measurement	Constraint	Procedure	Date 09.04.23	Date 27.07.23	Sensor/station	Agent	Location
	Water level	Meter above reference point	0<	Groundwater Monitoring Procedures	7.5	7.1	Groundwater level sensor	Bundesministerium für Umwelt	51.1657° N 10.4515° E
	Temperature	Degree Celsius	-5 - 50	Groundwater Monitoring Procedures	2	5	Sensor 3	Bundesministerium für Umwelt	51.1657° N 10.4515° E
	Salinity	Parts per thousand	0 - 30	Groundwater Monitoring Procedures	2	3	Sensor 4	Bundesministerium für Umwelt	51.1657° N 10.4515° E

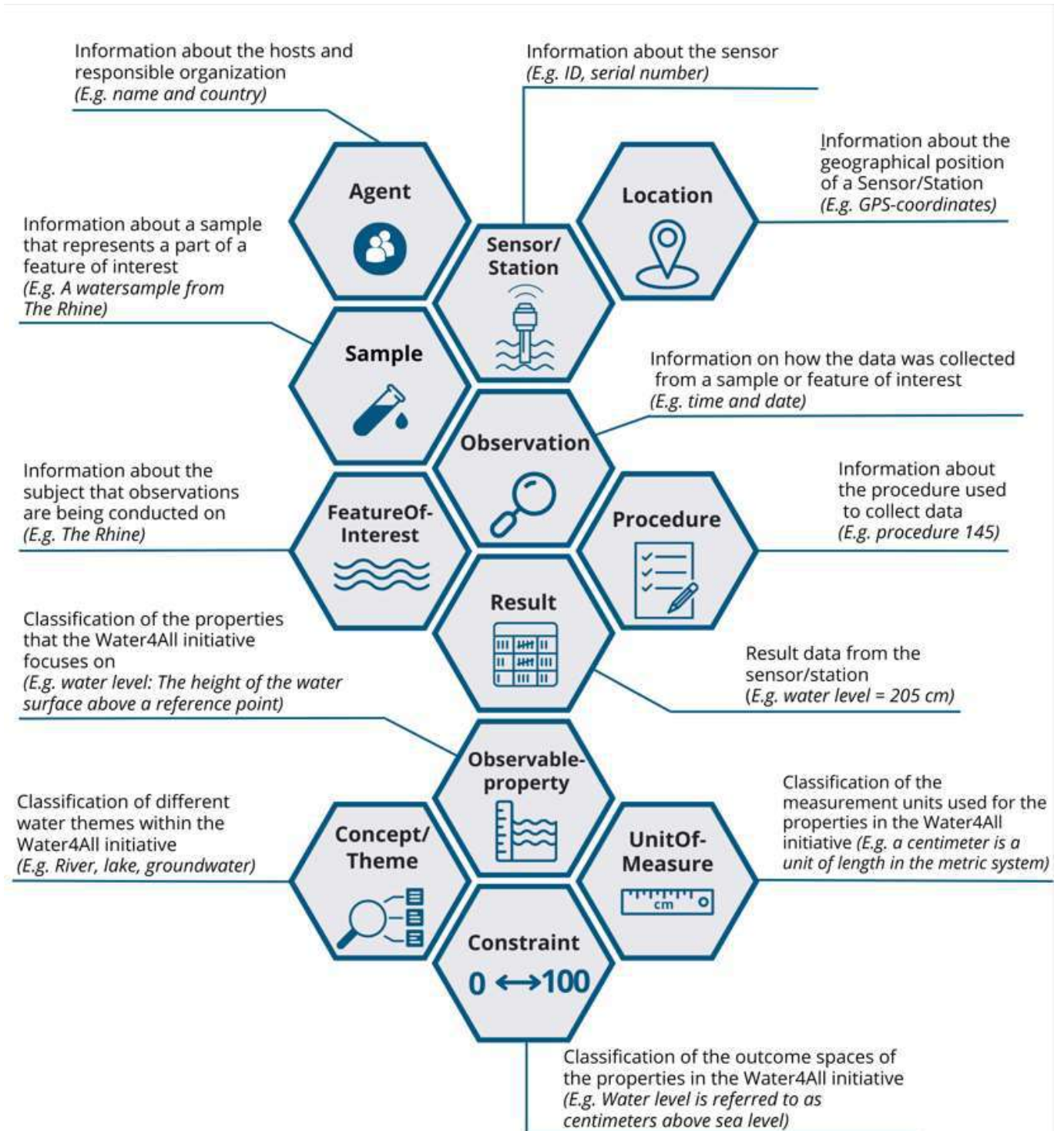
Figure 2 How data from different domains is collected in a harmonized way

The above figure shows how different properties from different domains is described in a uniform way. The same goes for the results after observation/measuring.

Building blocks in the observations, measurements and samples concept

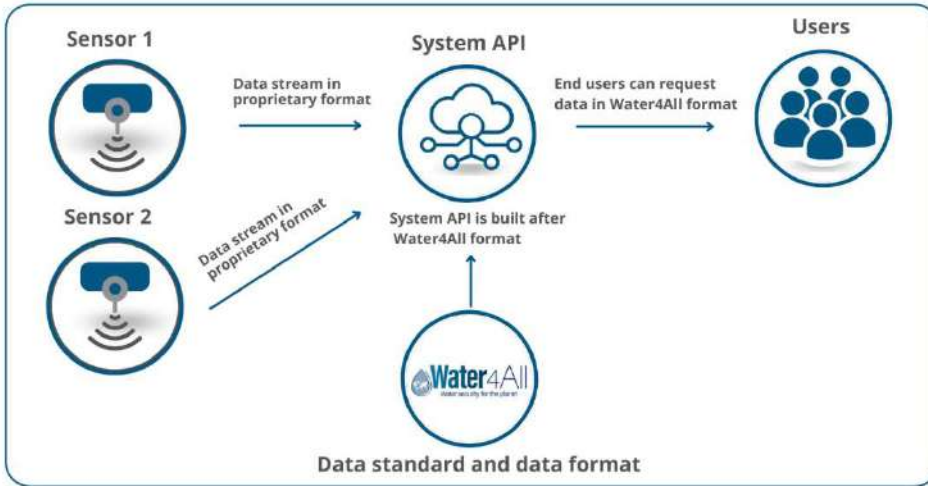
The “observation, measurements and samples” concept can be broken down to single elements (here called building blocks). Each of these building blocks embraces specific functionality and structure.

When we use the concept, our common language and common structures in our data sharing effort is described. And use of the concept ensures that exchanged data fits together and is widely understood.

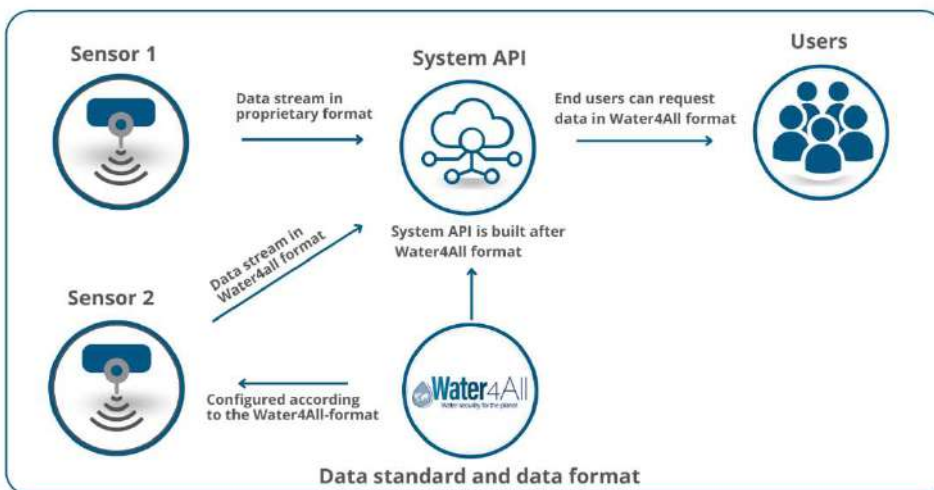


Use of the concept

The following section proposes a way to progressively add more FAIRness within Sensor based system.



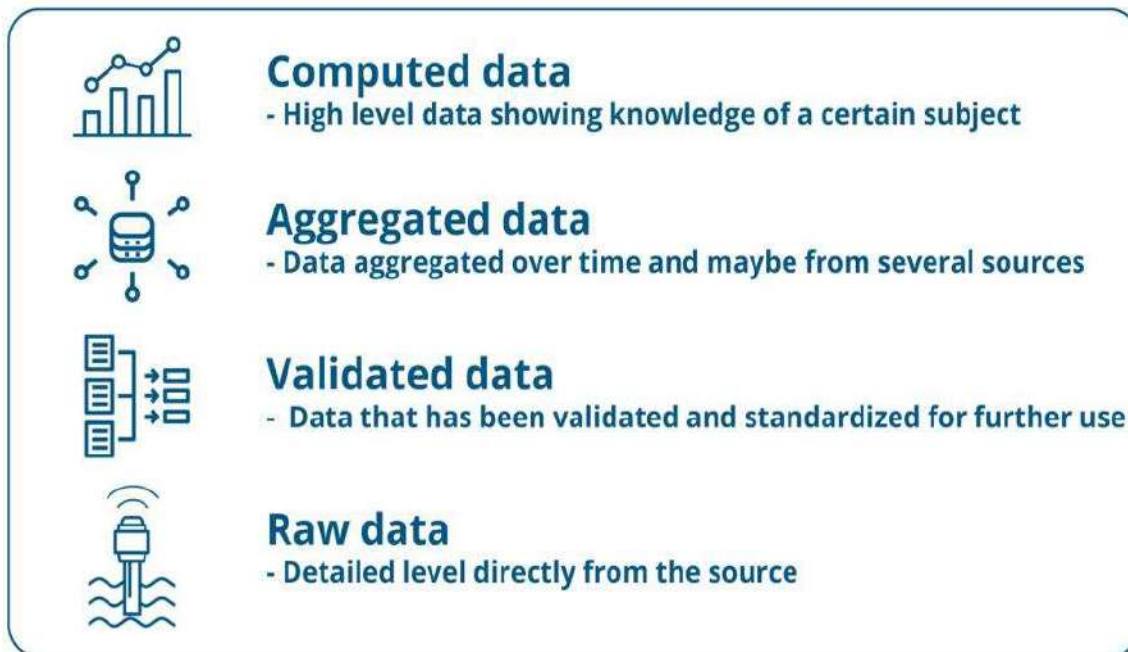
Many of the existing local applications and sensors do not adhere to the vocabularies, standards, and rules established in international FAIR Water data Standards and the best practices Water4All is recommending. Therefore, in order to exchange data in a FAIR Water data approach that follows common rules, data can be translated locally into the international FAIR Water data Standards and best practices recommended by Water4All using lookups in a shared observable property catalog.



Over time, the local systems can be built to directly utilize the core Water4All vocabularies. In both scenarios, access to the agreed-upon standards in a central property catalog is essential.⁶

⁶ The term 'Sensor' can be taken loosely in the case of that figure. Most of the time it's not at the Sensor level that one can configure/add some more FAIRness (sensor manufacturers are not haven't yet being convinced to change their legacy systems). In a sensor system, it's at the 'aggregator' level (the system that aggregates data from the sensors) that one often add FAIRness (ex: Observations, measurements and samples) etc. We should not make people believe that's the sensor output stream that will change. Maybe, that schema and the one before should have an extra box materializing that 'aggregation' level

Different levels of data sharing



Many of the API's, based on sensor standards, are describing, and exchanging data on a detailed sub-domain level. Basically, they are describing raw data, produced by a sensor or some human observations.

The starting point, however, could be a "higher" level of data sharing. We might need to develop special API's that are able to share data on both aggregated and computed level and rules on HOW TO aggregate and compute data.

The current standards on "Observation, Measurements and Samples" could serve as a standard for sharing data on a higher level if we document and share how it is done.

Key elements of a European Water Data Ecosystem

The key elements of a European Water Data Ecosystem could include:

- A strategic review of the current situation e.g. how to address barriers for harmonizing data at source
- Require all water data producers to use same format for location descriptions for monitoring stations
- Require all water data producers to use shared vocabulary allowing for immediate cross-border use
- Require national monitoring data to be FAIRified and available for free
- Ensure all new regulatory initiatives to use shared codelists and unique identifiers of parameters
- Undertake a strategic review of interoperability and reusability of biological parameters
- Adopt a shared parameter list for chemicals and update existing regulations accordingly
- Adopt a shared list of observable properties e.g. chemical parameters and taxonomi
- Develop operational guidelines on how build a common approach to design of datamodels
- Initiate a broad communication campaign to promote the vision and recommendations
- Promote use of a common metadata data catalogue in order to make findability of data more efficient

In the next section, a number of recommendations are listed envisaging how to initiate a Europe-wide process towards building a European Water Data Ecosystem.

Recommendations

In order to achieve the commonly shared and understandable eco-system for the water domain, there are a list of issues we need to agree on, listed in a sequence considered the easiest mentioned first.

- **Organizational and political recommendations**
 - **Demonstrate business value of sharing data.** It all starts and ends with business value. If one cannot see a clear business case in investing in data, it is not going to happen. Identification and showcasing of best practices in data sharing creating value throughout specific use-cases will provide valuable insights.
 - **Identify barriers to sharing data - organizations.** Identify barriers to sharing data across the water domain between organizations. Especially vested interests and perceptions of own uniqueness (leading to customized, but also sub-optimized IT solutions) constitute significant barriers for data sharing, beginning with willingness to share and placing a greater, common good at the core of matters.
 - **Agree on common principles of observations and measurements.** principles backed up with core models for surface and ground water features to constitute the core of Best Practices in Data Management, eventually towards a common standard or regulatory steps to require use of this in an EU context.
 - **Governance process:** The key stakeholders within data sharing should do this following agreements on ownership rights, service levels of the intermediary and user rights.
- **Cultural recommendations**
 - **Identify incentives for sharing data between researchers.** Though the Frascati manual and the associated data management plans have been in place for some time within European and national funded research projects, it appears that many researchers put little emphasis on sharing data. What are the incentives and what are the barriers for sharing data from a researcher point of view.
 - **Identify barriers to sharing data between different sub-domains** in the water sector. Identify barriers to data sharing apparent across the water sector. E.g., the geological surveys of Europe have developed a number of data standards but how interoperable are these data compared with data collected by, e.g., sub-domains related to lakes, rivers and streams or wastewater.
 - **Build on joint, cross-cutting strategies:** Expand cooperation between DG Digital – data spaces – DG Environment – protection measures – and DG Research – research – to exploit synergies
 - **Analyze interoperability problems related to biological data:** this is a particularly fragmented field, making interoperability very difficult. A feasibility study could initiate a relevant, but difficult process. At the end, however, this may be essential for managing biodiversity and for cross-cutting use of data within all biosphere-terrain data
- **Technical recommendations**
 - **Common datasets and API catalogue :** to facilitate the discovery of available Water datasets all across Europe even if not FAIR in the first place. This would help the EU Water community find relevant data and access to it. While also helping identify those who could be helped to achieve FAIRness. Common data model: The core data model shall be based on the Observation, Measurement and Samples principles and all exchange of data shall meet the specifications derived from the common model. E.g. data from a sensor on a detailed level would be defined according to Observations, measurements and samples

- **Common classifications/vocabularies/Observable property catalogue**: It is essential that all parties are using the same and shared vocabularies. E.g. Observable properties, describing of semantics, Units of measure, procedures, types and roles.
- **Establish EU codelists for observable properties** – e.g. chemical substances and taxonomi – to enable that rules and regulations can be digitalized and make data models and data sharing easier – and make use of the same classification system for chemical parameters mandatory. Aligning content is feasible as many organizations and countries do already provide such codelists (ex : DK DEP parameter-list, FR Sandre’s PAR reference dataset, ...). However, a proper unique EU reference point is clearly missing.
- **Emphasize Technical Coherence**: Focus on sharing software components, compliant with the above-mentioned standards that could be reused across organizations.
- **Align use of metadata**: harmonize use of metadata, and ideally reduce the number of fields of metadata used for better interoperability – simplicity of data models and data systems are a way to ensure interoperability.
- **Develop, implement and enforce a unanimous approach to site descriptions**: without the same way to describe sites including types of monitoring stations, data from different sources cannot be merged.
- **Resource recommendations**
 - **Ensure proper training to FAIR Water Data**: there is an equally important need to, in parallel, train people within organizations about FAIR Water Data Sharing and provide them with tooling that also contribute to a water data ecosystem

Annex 1

About Water4All

Access to clean water is a fundamental human right, yet millions of people worldwide lack this basic necessity. Water Pollution levels are at levels making achievement of good ecological status more a fantasy than realistic. Floods are occurring with increasing frequency and damages.

Recognizing the importance of addressing these issues, the European Union (EU) has launched several initiatives to ensure that all citizens have access to safe and clean drinking water, prevent floods and droughts etc. One such initiative is the Water4All Partnership, a comprehensive European endeavor aimed at improving water quality and availability across Europe.

The Water4All Partnership is a collaborative effort involving various stakeholders, including governmental bodies, non-governmental organizations (NGOs), research institutions, and industry partners. Launched by the European Union, the Partnership aims to address water-related challenges such as water scarcity, pollution, and inadequate infrastructure. The Water4All gathers more than 90 partners representing the whole Research and Innovation (R&I) chain. Water4All runs its activities across five thematic and two cross-cutting themes of its Strategic Research and Innovation Agenda (SRIA), with a preliminary budget of 112 M€, including EU contribution.

The Water4All Partnership focuses on delivering sound knowledge to address global water challenges, and recognizes in its "Open Science Open Data Strategy"⁷ the significance of sharing knowledge and data to achieve its objectives. In line with the Horizon Europe Model Grant Agreement (Article 17-Communication, Dissemination, Open Science and Visibility), the partnership emphasizes the application of open science, open data practices, and adherence to the FAIR principles (Findability, Accessibility, Interoperability, and Reuse), while considering the principle of "as Open as Possible, as Closed as Necessary" to respect GDPR sensitivities and allow for intellectual property rights protection when required.⁸

In Subtask C.4.3: Toolbox for managing water related data, the sharing of water-related data will be implemented through the development of interoperable interfaces for finding and accessing data thus contributing to a vision of a digitalized European Water Sector based on Open Data. These interfaces will support the compilation and storage of, as well as access to products associated with Water4All activities. The Water4All Partnership has chosen an approach to see its efforts regarding data sharing as being part of a wider European community and has accordingly decided to compile experiences on Data Management and Data Sharing as a source for a wider European Dialogue.

This White Paper on Best Practices in Data Management is therefore first of all aimed a wider European Discussion and is not only for internal purposes, despite the actual work will strive to be in accordance with the suggestions of the White Paper.

⁷ <https://research-and-innovation.ec.europa.eu>

⁸ [European Commission \(2021\). Model Grant Agreement V1.1](#)

As shown here, the Water4All’s Strategic Research and Innovation Agenda (SRIA) includes five themes in the areas of circular economy, ecosystems, sustainable water management, water and health and infrastructures, as well as two cross-cutting themes.



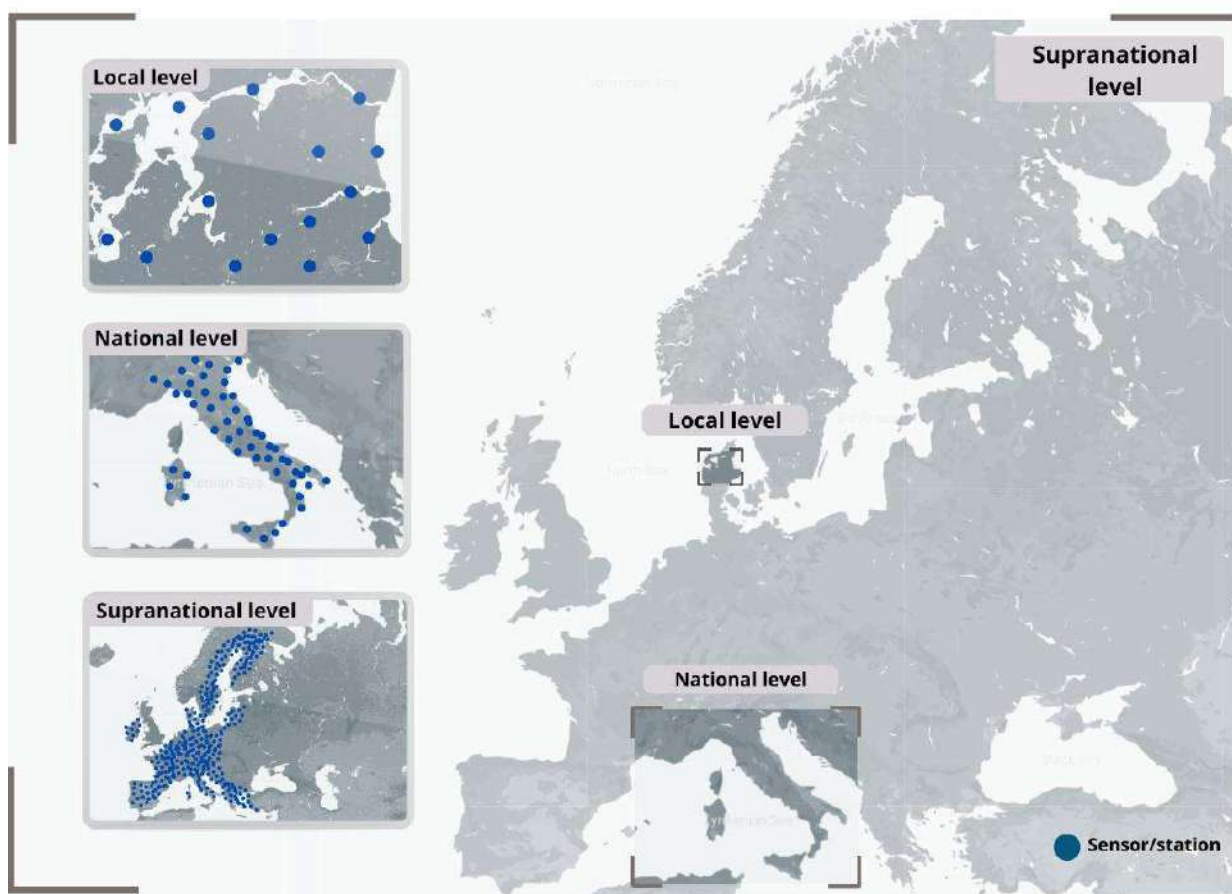
Digitalization (The digital revolution “big data, AI, IoT”) is considered a key enabler for addressing the vast challenges within these five themes:

Annex 2

A holistic view of the water domain

Water4All aims to elevate knowledge about water (in the broadest sense) to the EU level. Until now, individual countries have worked with local-level observations and measurements for local purposes, often at the municipality or regional level and frequently in local, proprietary formats.

Water4All seeks to harness all these local measurements and elevate them to both a national and supranational level.



Therefore, there is a need to draw on all these (existing) data sources from a holistic perspective and view all the local measurements in a uniform manner. Water4All establishes the framework for how we collectively should understand water data. The framework “Observation, Measurements and Samples” offers a systematic approach to document agreed issues like topics, properties semantics etc. of the water domain – or any domain for that matter. All those features - rivers, aquifers, Industrial sites etc - need also to be shared using the appropriate Feature model

When all parties share data according to these agreements, we will be able to use the data across borders and domains and achieve a holistic view on EU-level.

In the following, an example of use of this approach is shown. It should be noted that there are several IT platforms at national level that do collect these data in a structured way.

An example of use⁹

A scientist, working with the topic of waterflow and water level in the river “Rhine” needs data from the river to do her work.

1. On a map, she points out the Rhine and the theme “waterflow/waterlevel”
2. Pointing out the Rhine automatically gives a geospace that can be used as search criteria.
3. The application searches all stations (also cross country) within the border of the geospace that also fulfills the capability of measuring “waterflow/waterlevel” and points them out on a map.¹⁰
4. On the map she can filter out active/inactive stations in a certain period of time
5. By selecting the stations¹¹ she is interested in, she can retrieve data from these stations in the selected period of time. All stations is also a possibility
6. Maybe she needs data from another topic like rainfall, which she can get in the same manner.
7. With all these data in hand she can continue her scientific work.



Figure 3 Selection of all station with the ability to observe waterlevel. The big difference between this scenario and the current situation is that it is now possible to retrieve data from already existing stations/sensors without caring about borders or ownership of the data.

⁹ A number of IT platforms at national level exists that do collect these data in a structured way, it would be relevant to map these.

¹⁰ Should also be stations mentioning that their FeatureOfInterest is the Rhine (or a segment, watercourse linke, watercourserl ink sequence pertaining to the Rhine).

¹¹ Water4All is as part of the work with a Data Sharing Facility testing by use of prototyping with SensorThings API to make data at a station level accessible – which doesn't imply that Water4All will build a permanently working system for this.

Observations, measurements and samples (OMS)

In our effort to make data understandable in the shared context across member states, we use the “Observations, measurements and samples” concept, which is originally described as ISO-standard (ISO 19156) and updated in 2023. It is widely used in e.g.:

- WaterML2 series of standards
- INSPIRE Data Specifications (Observation and Measurements guidelines and several INSPIRE themes)
- WMO Core Metadata Profile
- Many other domain standards (Soil, LandInfra, AirQuality, Geology...)
- Many information system (regional, national, continental and international), and several tools (open-source or no). A quick list around the water domain
 - NR-Can, GSC, USGS, US EPA, CUASHI, SDSC, BRGM, UK CEH, DELTARES, GRDC, BaFG, Univ Tartu, Fraunhofer IOSB, DataCove, NIWA/LAWA (NZ), Federation University (Australia), BoM Australia, CSIRO, ...
 - WMO, UNESCO, ...

Kisters, 52°N, Aquatic Informatics, ...

Towards Data Models and related Documentation

The Observations, Measurements and Samples Principles as described above will subsequently constitute the framework for the design of one of the core part of Water4All data models related activities. Already significant work has been undertaken regarding formulating an internationally agreed approach, e.g. in the course of the OGC/WMO Water Quality Interoperability Experiment (IE): <https://github.com/opengeospatial/WaterQualityIE>.

The Water Quality IE builds on the OGC Semantic Baseline:

- Observations, measurements and samples : OGC/ISO 19156:2023
- OGC /WMO WaterML2 Part 3 : Surface HydroFeatures
- OGC/WMO WaterML2 Part4 : GroundWaterML2

As those three standards are also reused within major institution and organisations information systems, Water4All proposes to build on them.

Regarding Data Models, local data models shall inevitably be made according to local needs, however certain degrees of harmonization should be made across water segments, i.e. groundwater and surface water, quantity and quality – thus making interoperable data models. One also has to keep in mind that when several local data need to appear in various local implementations, then they are good candidates to be added to the international standards Semantic Baseline. That’s actually how the three standards mentioned right above were made.

For subsequent building of the technical interfaces between distributed datasets, the documentation following each local solution will become important in order to smoothen the process. Accordingly, international guidelines (endorsed at an appropriate, political level) should not only outline a common, basic foundation but also outline formats for documentation. Also, agreed approaches to metadata should be described. Sharing documentation of local solution could be useful to help data providers map to standards and enrich standards. Otherwise, the point of interoperability is to define the ‘external’ contract between system regardless of the local solution in place

Annex 3

The FAIR principles

(<https://www.go-fair.org/fair-principles>)

In 2016, the '[FAIR Guiding Principles for scientific data management and stewardship](#)' were published in *Scientific Data*. The authors intended to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the [FAIRification process](#).

[F1. \(Meta\)data are assigned a globally unique and persistent identifier](#)

[F2. Data are described with rich metadata \(defined by R1 below\)](#)

[F3. Metadata clearly and explicitly include the identifier of the data they describe](#)

[F4. \(Meta\)data are registered or indexed in a searchable resource](#)

Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

[A1. \(Meta\)data are retrievable by their identifier using a standardised communications protocol](#)

[A1.1 The protocol is open, free, and universally implementable](#)

[A1.2 The protocol allows for an authentication and authorisation procedure, where necessary](#)

[A2. Metadata are accessible, even when the data are no longer available](#)

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

[I1. \(Meta\)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.](#)

[I2. \(Meta\)data use vocabularies that follow FAIR principles](#)

[I3. \(Meta\)data include qualified references to other \(meta\)data](#)

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).

Annex 4

The principles of “Internet of Water”

In the US a program called “InternetOfWater”¹² (see internetofwater.org) states the following principles for harmonizing water data management, which very well together with what Water4All White Paper is based on:

- Water data are essential for efficient, equitable, sustainable, and resilient water planning, management, and stewardship.
- Modern data infrastructure increases the usefulness of water data and enables its broadest possible application.
- Data equity is necessary for water equity; modern data infrastructure should be implemented and governed so that data are usable by and for overburdened communities.
- All water data produced for the public good should, by default, be findable, accessible, interoperable, and reusable (FAIR) for public use or authorized users.
- Security and privacy risks associated with sharing data can be mitigated using mechanisms for tiered access for authorized users.
- Commonly accepted data, metadata, and exchange standards should be adopted by water data producers to promote interoperability, efficiency, sharing, equity, and secondary uses of data
- Control and responsibility over data are best maintained by data producers.
- Data producers are responsible for sharing data of known quality and documenting essential metadata; data users are responsible for determining whether data are appropriate for specific purposes and uses.
- Federated, distributed systems of interoperable public water data generally provide scalability and flexibility to meet the diverse needs of data producers and users.

¹² www.internetofwater.org

Annex 5

European initiatives focusing on water data sharing

In the following a non-exhaustive list of European initiatives focusing on water data sharing is presented:

Name of Initiative	Activities	Information
Stars4Water	The STARS4Water Metadata Portal is an open access, flexible and modular environment, easily adaptable to the different needs of the River Basin Organisations. It aims to unlock existing data and information sources for water resources management and climate change adaptation, including new developments at EU and global level and lessons learnt from previous studies on data collection and data standards.	https://stars4water.eu/
EOSC	The EOSC Association works to advance Open Science in the service of creating new knowledge, inspiring education, spurring innovation and promoting accessibility and transparency.	https://eosc.eu/
EMODNet	The European Marine Observation and Data Network (EMODnet) is the European Commission (EC) in situ marine data service of the EC Directorate-General Maritime Affairs and Fisheries (EC DG MARE) and funded by the European Maritime Fisheries and Aquaculture Fund. Established in 2009, EMODnet plays a pivotal role as a trusted source of in situ marine environmental and human activities data and data products, serving a diverse user base across various sectors.	https://emodnet.ec.europa.eu/en
Aqualnra	The AqualNFRA project aims to develop a virtual environment equipped with FAIR multi-disciplinary data and services to support marine and freshwater scientists and stakeholders restoring healthy oceans, seas, coastal and inland waters. The AqualNFRA virtual environment will	https://aquainfra.eu/

	enable the target stakeholders to store, share, access, analyse and process research data and other research digital objects from their own discipline, across research infrastructures, disciplines and national borders leveraging on EOSC and the other existing operational dataspace.	
GSEU	To develop the geological data infrastructure - building on the existing EGD - to provide permanent access to and dissemination of the data and information services developed under the project and beyond, targeting a wide range of stakeholders, with the specific aim of enabling further innovation and strengthening the market uptake of innovative solutions;	https://www.geologicalservice.eu/
ICT4Water	The ICT4Water cluster is a hub for EU-funded research and innovation projects on ICT applied to water management.	http://www.ict4water.eu/
SWAN Forum IGU	A collaborative effort of global, SWAN utilities working together to address the challenges faced by the water sector. Our mission is to develop a digital ecosystem that supports interoperable applications and consistent asset data structures, simplifying digital solution adoption and procurement for water utilities worldwide. By focusing on asset data structure and interoperability reference architecture, we aim to create a more flexible, innovative, and efficient water sector.	Interoperable Utility Group - SWAN Forum (swan-forum.com)
WaterVerse	WATERVERSE is an EU-funded project aimed at developing a Water Data Management Ecosystem (WDME) for making data management practices and resources in the water sector accessible, affordable, secure, fair, and easy to use.	https://waterverse.eu/

Annex 6

EU solution for regulating chemical substances in the environment.

Problem

In a highly digitalized world, the EU-legislation on chemical substances in the environment is still words written on paper. In the many EU directives and appendixes covering the monitoring of chemical substances, the substances are listed unstructured as plain text, without any unique identifier, sometimes in a table, sometimes abbreviated and randomly with external references.

Translating directives into national regulations, surveillance programs, and ultimately into laboratory contracts and analyses is a time-consuming and error-prone process. There is a risk that important details may be lost in translation, leading to errors and varying interpretations of which substances to measure and how to measure them.

At the national level this makes it in essence close to impossible to track the monitoring of chemical substances across directives and thus across different media in our environment. Do we monitor the same substances in our groundwater as in our drinking water? It is a simple question, it will take hours to find the answer, and you will most probably get it wrong!

Most toxic chemical substances float freely across media and across borders, but regulation works in silos on both the EU-level and National level. Being able to track how toxic chemicals spread, from groundwater into human tissue is extremely time consuming or outright impossible, because the data gathered has almost no interoperability.

Everyone in the environmental sector in the European Union recognize the problem. There are many regional, national solutions in every country, but the real solution lies in finding a common unified approach.

In fact, when the marine researchers in Denmark had found their solution on how to identify and register chemicals in the seabed, they unintentionally created another problem, for the broader use of the data.

This will often prompt the valid question: "Isn't the real issue that no one can agree on adopting one of the many existing classifications already available?".

Any classification of chemicals is designed with specific purposes in mind. For example, the CAS Registry, created by the American Chemical Society (ACS), was developed to meet chemical industry needs. If the need arises for the Danish EPA to add a newly identified degradation product of a pesticide to the CAS Registry, it would not be possible.

Likewise, the European List of Notified Chemical Substances (ELINCS) was created as part of the EU's regulatory framework for managing chemicals. Its main purpose was to document and regulate substances notified for marketing within the European Economic Area (EEA). A degradation product of a pesticide does not belong to the list.

To illustrate why we cannot use a list created for the regulation of commercial activities, the list shows some of the shortcomings of REACH/EC-numbers:

- The EC numbers in the list are not associated with CAS numbers or chemical unique identifiers. You can't even look up e.g. "Trifluoroacetic acid" (TFA), one of Europe's currently most investigated chemical substances in environmental monitoring.
- The EC numbers do not cover all chemical substances, incl. degradation products which are monitored in the environment.
- The EC numbers are not associated with monitoring data (wastewater, surface water, groundwater, bathing water, drinking water, etc.), but an EU parameter list will be able to make the connection.
- The EC numbers are not linked to the unique IDs used in the WISE reporting.
- The EC numbers are not linked to the EQS found in the Water Framework Directive or other EU legislation.
- The EC numbers do not contain popular names of substances, e.g. PFOS, which is widely used.
- The EC numbers do not contain references to the parts of EU legislation in which they are regulated with limit values or the like.
- The EC numbers are not digitized and therefore cannot be used immediately. As they are accessible in a digital format, they will have to be incorporated into national lists manually, risking errors.

The lack of a common unified approach to the problem has a cost. The administrative burden of decoding regulation is found on all levels of the value chain, state level, regional level, municipalities and in private companies. These dispersed efforts, by highly skilled chemists, have a direct impact on quality and most important on the lack of interoperability of the data produced. IT systems can't interact, reporting becomes manual processes, workload increases exponentially with every new directive.

There is a need to create a solution for the classification of chemicals and a corresponding digitized version of all regulation in environmental analysis across all media, from seabed analysis to livestock feed, to human blood samples.

Solution

The Danish "Parameter List" exemplifies a unified approach. The solution caters for all EU and National regulations across all the different media and for the whole value chain. From national agencies to municipalities to private enterprises.

The Danish parameter list assigns a unique code to all physical, chemical, and microbiological parameters reported in Danish environmental regulation. In the list, the parameters can be associated with metadata securing their unambiguously description. I.e. chemical parameters are added information reg. exact chemical structure, so there is no doubt about the identity of the chemical.

The parameters can also be added into groups (e.g. pesticides, metals, priority substances, river basin specific pollutants or parameters included in the EQS for dioxins and dioxin-like compounds) and be part of one or many groups.

Specific parameters can also be associated with threshold values in the parameter list, e.g. environmental quality standards (EQS) or drinking water standards. It is also possible to link the threshold values and the related parameters, to specific analytical quality requirements as defined in the directives, e.g. Commission Directive 2009/90/EC of 31 July 2009. This enables a digitization of environmental legislation.

Standardization of parameters involved in environmental regulation in the EU, especially the unambiguous naming and identification of substances and substance groups, has a wide range of benefits outlined below.

It is worth highlighting here that standardization across the EU will ensure that legislators, officials, and citizens have a more comparable basis for assessing and addressing problematic issues such as the prevalence of PFAS in the EU surface waters, because the measurements are based on identical compounds.

The potentials for the EU and the EU countries using a shared unique digital code system within environmental monitoring and regulation are many:

Enhanced data quality

The quality of environmental data in European countries will increase due to use of the unique code unambiguously describing the identity of a parameter. There will be no doubt about the identity of a parameter. Thus, the quality of data reported to the EU, e.g. via WISE, will also increase. Centralized grouping, e.g. of priority substances, will also decrease the risk of errors at national level.

Digitized environmental legislation can elucidate the analysis quality requirements associated with specific measurements and thus contributes to improving data quality.

Efficient workflows

The unique code can be coupled to additional information for instance associated with analysis (NORMAN network) and legislation (ECHA), which will contribute to fewer manual errors and more efficient workflows. Digitized environmental legislation and centralized digital grouping, e.g. of priority substances, will also contribute to more efficient workflows at national level.

Increased interoperability

Data will be easier to compare across countries due to harmonized data. For instance, it becomes obvious which country measure the linear PFOS and which measure the sum of linear and branched PFOS for the parameter "PFOS".

Nuanced overview

It will be possible to draw a more detailed and accurate pollution picture cross border. In principle, a unique code can be created for any environmental relevant parameter, thus making it possible to share knowledge and create an overview beyond the mandatory EU-WISE reporting system. For example, it would be possible to easily display which pesticides contribute to the ambiguous parameter "sum of pesticides" in which countries.

Increased knowledge exchange

A common platform displaying environmental data based on unique codes improves the international exchange of information on new environmental substances, acting as an early warning and input system for national and EU watch-lists.

Recommendation

In essence, the proposal, which has already been successfully implemented in Denmark, is to create a classification system tailored to environmental regulation. This would allow for the digitization of regulations and promote interoperability of environmental chemical analysis results across the Union. Paving the road to automating many processes, including reporting to the EU.

The Danish EPA and the Danish Environmental Portal has jointly developed the Parameter List. What began as a modest IT investment quickly became crucial component driving everything else forward.

Except from a continuous strong governance by the Danish EPA in creating the content of the list, there have been no implementation activities amongst users. Technology suppliers, municipalities, regulators and laboratories quickly spotted the obvious benefits and what we have seen, has been a self-driven pull-effect.

There is ample reason to believe that an implementation of a Parameter List on EU-level, would be an even bigger game changer, than it has been in Denmark.