



# FAIR for Water: specifications for FAIR principles and interoperability practices in the field of Water

Water4All C.4.3 team – December 2025



Co-funded by  
the European Union

<b>OUTPUT SUMMARY</b>	
<b>Project information</b>	
<b>Project Title</b>	European Partnership Water4All – Water security for the planet
<b>Project Acronym</b>	Water4All
<b>Call Identifier</b>	Horizon-CL6-2021-Climate-01-02
<b>Contract Number</b>	101060874
<b>Starting Date</b>	1 June 2022
<b>End Date</b>	31 May 2029
<b>Web-Site Address</b>	<a href="http://www.water4all-partnership.eu">www.water4all-partnership.eu</a>
<b>Coordinator</b>	ANR
<b>Management Team</b>	Benjamin Lopez, Juliette Arabi, Armelle Montrose, Claire Treignier
<b>E-Mail</b>	<a href="mailto:Water4All@agencerecherche.fr">Water4All@agencerecherche.fr</a>
<b>Phone Number</b>	+33 1 78 09 81 20 / +33 1 73 54 81 43

<b>Milestone Title</b>	
<b>Milestone Number</b>	
<b>Work Package</b>	Pillar C
<b>WP leader</b>	BRGM-DMP
<b>Nature</b>	R (Report)
<b>Dissemination</b>	PUB (Public)
<b>Editor (s)</b>	Sylvain Grellet (BRGM) and C.4.3 Team (ISPRA, FORMAS, VITO, DMP/DEP, Fraunhofer IOSB, BAFG, CNR-IT)
<b>E-Mail (s)</b>	<a href="mailto:s.grellet@brgm.fr">s.grellet@brgm.fr</a>
<b>Date of Delivery</b>	December 2025

<b>Verification by</b>	Water4All C.4.3 task
<b>Date</b>	Summer 2025
<b>Validation by</b>	Water4All C.4.3 task
<b>Date</b>	Summer 2025
<b>Review</b>	
<b>Date</b>	

## Acknowledgements

Water4All has received funding from the European Union's Horizon Europe Programme under Grant Agreement 101060874.

This document results from the contribution from a large number of stakeholders in water including Water4All partners and Water4All related networks.

Disclaimer: This document reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein

## TABLE OF CONTENT

List of acronyms.....	4
Introduction.....	5
Background to FAIR Data.....	6
The spatial data infrastructure initiatives .....	6
The Web of Data.....	1
The two initiatives brought together .....	2
FAIR initiative.....	3
Findable .....	4
F1: METADATA AND DATA ARE ASSIGNED A GLOBALLY UNIQUE AND ETERNALLY PERSISTENT IDENTIFIER. ....	4
F2. DATA ARE DESCRIBED WITH RICH METADATA (DEFINED BY R1 BELOW) .....	5
F3. METADATA CLEARLY AND EXPLICITLY INCLUDE THE IDENTIFIER OF THE DATA THEY DESCRIBE.....	6
F4. METADATA AND DATA ARE REGISTERED OR INDEXED IN A SEARCHABLE RESOURCE .....	7
Accessible .....	8
A1. METADATA AND DATA ARE RETRIEVABLE BY THEIR IDENTIFIER USING A STANDARDISED COMMUNICATIONS PROTOCOL .....	8
1.    A1.1 the protocol is open, free and universally implementable .....	8
2.    A1.2 the protocol allows for an authentication and authorisation procedure, where necessary ....	8
A2. METADATA ARE ACCESSIBLE, EVEN WHEN THE DATA ARE NO LONGER AVAILABLE .....	9
Interoperable.....	9
I1. METADATA AND DATA USE A FORMAL, ACCESSIBLE, SHARED, AND BROADLY APPLICABLE LANGUAGE FOR KNOWLEDGE REPRESENTATION .....	9
I2. METADATA AND DATA USE VOCABULARIES THAT FOLLOW FAIR PRINCIPLES .....	11
I3. METADATA AND DATA INCLUDE QUALIFIED REFERENCES TO OTHER METADATA AND DATA .....	12
Reusable .....	13
R1. METADATA AND DATA HAVE A PLURALITY OF ACCURATE AND RELEVANT ATTRIBUTES .....	13
1.    R1.1. Metadata and data are released with a clear and accessible data usage license .....	13
2.    R1.2. Metadata and data are associated with their provenance.....	13
3.    R1.3. Metadata and data meet domain-relevant community standards .....	14

## LIST OF ACRONYMS

ACRONYM	FULL TITLE
API	Application Programming Interface
CSV	Comma-separated value
DCAT	<a href="#">Data Catalog Vocabulary (W3C ontology)</a>
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoperable, Reusable
FDMM	FAIR Data Maturity Model
FIP	FAIR Implementation Profile
GDPR	General Data Protection Regulation (Regulation (EU) 2016/679)
HTTP	Hypertext Transfer Protocol
INSPIRE	Infrastructure for Spatial Information in the European Community (Directive 2007/2/EC)
ISO	International Standards Organization
IT	Information Technology
JSON	JavaScript Object Notation
MOD	Metadata for Ontology Description and Publication Ontology
OGC	Open Geospatial Consortium
OMS	<a href="#">Observations, Measurements and Samples (OGC/ISO 19156:2023)</a>
PID	Persistent Identifier
RDA	Research Data Alliance
RDF	Resource Description Framework
REST	REpresentational State Transfer
SSN/SOSA	<a href="#">SSN/SOSA: Semantic Sensor Network / Sensor, Observation, Sample and Actuator (W3C ontology)</a>
RI	Research Infrastructure
TTL	Terse RDF Triple Language
UML	Unified Modelling Language
URI	Uniform Resource Identifier
WaterML2	<a href="https://www.ogc.org/publications/standard/waterml/">https://www.ogc.org/publications/standard/waterml/</a>
W3C	World Wide Web Consortium
XML	Extensible Markup Language

## INTRODUCTION

As mentioned in their name, FAIR principles are principles and thus can be interpreted in a number of different ways.

This document aims to provide a framework for applying them according to the approach adopted by Water4All, right from the start of the project in its task dedicated to FAIR (C.4.3).

The aim of Water4All task C.4.3 is to ensure that these exchanges of digital resources take place in accordance with FAIR principles. C.4.3 tasks and more broadly Water4All task related to digital objects shall apply this approach/philosophy directly. If a subject requires a choice to be reviewed (for justified reasons), this document will need to be updated accordingly.

Given the duration of the partnership, it is possible that international approaches to FAIR data will be adjusted. It is also possible that a new case will come to task C.4.3 attention. This document will therefore be supplemented as the project progresses.

When the document is mature the next step will be to formalize the various parts into FAIR Implementation Profiles<sup>1</sup>.

*NB1: The term "digital object" is used throughout this document.*

*This corresponds to the use of this term in the Gaia-Data project's data management plan, which reflects the current activity of the Research Data Alliance (RDA).*

*The RDA offers the following definition: "A digital object (DO) is represented by a bitstream, is referenced and identified by a persistent identifier and has properties that are described by metadata.*

*Note: As indicated we only talk about registered DOs in the context of this document.*

*Note: Properties included in metadata include discovery, contextual, schema, rights, curation and provenance information.*

*Note: A DO is said to be a dynamic DO when the information content represented in a DO is changing for some period of time or even for an indefinite duration"<sup>2</sup>.*

*This term encompasses the concepts of datasets, dataset distribution, computer code, digital resources (video, audio, pdf, etc.) and the "machine actionable" data management plans themselves.*

*The term "digital object" will be used to describe the scope of the OneWater Data project. More precise terms (e.g. dataset, data, computer code, etc.) will be used when there is a need for greater precision in relation to these types of digital objects in the sections concerned.*

*It should also be remembered that the FAIR approach is intended for both humans and machines. This crucial intention should be taken into account before reading this document and is used throughout.*

*NB2 : several equivalent initiatives aiming at clarifying FAIR principles. As there is no point in reinventing the wheel, this document builds on them (ex : French PEPR OneWater<sup>3</sup>, ...) thus applying the Reusability principle.*

---

<sup>1</sup> <https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/>

<sup>2</sup> <https://www.rd-alliance.org/sites/default/files/DFT%20Core%20Terms-and%20model-v1-6.pdf>

<sup>3</sup> <https://www.onewater.fr/fr/projets/projets-cibles>

## BACKGROUND TO FAIR DATA

The FAIR Data initiative is linked to the publication entitled "The FAIR Guiding Principles for scientific data management and stewardship"<sup>4</sup>. Since 2016, it has become a benchmark inspiring an international movement.

### The spatial data infrastructure initiatives

It should be remembered that there have been previous initiatives concerning spatial data infrastructure (whether for public authorities, research bodies, international organisations or private companies).

Since the mid-1990s, this work has been formalised within the Open Geospatial Consortium (OGC) through standards and best practices, and some of it has also been included in the ISO/TC 211 standards (ISO 191xxx series of standards).

These practices can be applied by different bodies in order to share their digital resources in an interoperable way.

In 2007, the INSPIRE Directive<sup>5</sup> was introduced in Europe on the basis of this approach. This text contains rules to be implemented for sharing geospatial and environmental data in Europe. To help public authorities implement these rules, technical documentation has been drawn up, based mainly on the OGC standards, which have been incorporated in the ISO standards.

The INSPIRE approach aims to provide three main functions:

- Search: provide a metadata catalogue describing existing data,
- View (or Find): have the option of seeing a data producer's data,
- Download (or Bind): have the option of downloading data from a producer, most of the time in order to use them in another system or calculation.

An additional function can be added: Transform, i.e. convert data from one state to another (e.g. projection into a different spatial reference frame, etc.).

All these functions are accessible via interoperable services available on the Internet. These services are accessed via web applications (e.g. data portals) or applications installed on individual computers.

These approaches (INSPIRE, OGC, ISO/TC 211) apply not only to geospatial data but also to observation data, by means of dedicated standards.

The approach promoted by INSPIRE has undeniably enabled the sharing of data that was previously inaccessible. Many countries outside Europe are very impressed by the effort and momentum around the

---

<sup>4</sup> 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>

<sup>5</sup> [https://knowledge-base.inspire.ec.europa.eu/index\\_en](https://knowledge-base.inspire.ec.europa.eu/index_en)

Directive (which in turn has provided input for the OGC and ISO standards, and also enabled many open-source implementations to emerge and/or be improved).

However, this approach has also shown its limitations. These methods were a good strategy for sharing data in an interoperable way, but were confined to more advanced users with a fairly intimate knowledge of the technical specifications to be implemented (e.g. standards such as CSW, WFS and SOS).

One of the main barriers was exposure of data in accordance with a data standard via solutions such as the OGC WFS. These approaches can be fairly complex to configure and use if no support is provided to the user (e.g. for requesting this type of service).

## The Web of Data

In parallel with the work on spatial data infrastructure, the World Wide Web Consortium – W3C (founded in 1994, the same year as the OGC) has spearheaded a considerable collaborative effort to provide a set of technologies, standards and best practices that form the basis for representing, publishing and sharing data via the web in a standardised way. Both public and private bodies have been involved in this effort.

The solutions proposed by the W3C, better known as "Semantic Web Technologies" (see "the Semantic Web Layercake"<sup>6</sup>), are embodied in four principles, known as the "Linked Data Principles" and described by Tim Berners-Lee<sup>7</sup>:

- Use Uniform Resource Identifiers (URIs) as names for things,
- Use HTTP URIs so that people can look up those names,
- When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL),
- Include links to other URIs so that they can discover more things.

The Web of Data results from the adoption of these principles by data providers around the world. It is not intended to replace the existing "Web of Documents", but rather to supplement it in order to achieve the vision of the Web as proposed by Tim Berners-Lee<sup>8</sup>.

In addition to these principles, Tim Berners-Lee also suggested a 5-star deployment scheme for the notation of Linked Open Data<sup>9</sup>. The aim of this approach is to encourage data providers to follow these best practices, which will help improve the visibility and reusability of their data across the Web.

In the same vein, the W3C's "Data on the Web" Working Group published its first working draft in February 2015 ("Data on the Web Best Practices"). These W3C recommendations were designed to ensure that data were discoverable and understandable by both humans and machines. They became a W3C Recommendation in January 2017 (<https://www.w3.org/TR/dwbp/>).

---

<sup>6</sup> <https://www.w3.org/2007/03/layerCake.svg>

<sup>7</sup> <https://www.w3.org/DesignIssues/LinkedData.html>

<sup>8</sup> <https://www.w3.org/History/1989/proposal.html>

<sup>9</sup> <https://5stardata.info/en/>

## The two initiatives brought together

Following on from these activities, the OGC and W3C began working together in 2015<sup>10</sup> after a workshop on common data sharing issues in 2014<sup>11</sup>.

These closer ties led to "cross-fertilisation" between the two standardisation initiatives.

This resulted in the "Data on the Web Best Practices" being updated as W3C/OGC "Spatial Data on the Web Best Practices" (<https://www.w3.org/TR/sdw-bp/>). It also formalised GeoSPARQL (the SSN/SOSA ontology) and supplemented the DCAT cataloguing work initiated at the W3C under its "Dataset Exchange" Working Group.

By highlighting some of these best practices (BP = Data on the Web Best Practices, SBP = Spatial Data on the Web Best Practices), it can be seen that they overlap with the initiatives encountered in Open Data and Open Science, and therefore with FAIR criteria (as will be seen below).

- BP1: Provide metadata,
- BP9, SBP1: Persistent URIs / Globally unique ID / Globally unique and persistent identifier,
- BP10: Use persistent URIs as identifiers within datasets,
- BP14: Provide data in multiple formats,
- BP15/SBP10: Reuse vocabularies (code list registries, ontologies), preferably standardised ones,
- BP17: Provide bulk download,
- BP18: Provide subsets for large datasets,
- BP19: Use content negotiation for serving data available in multiple formats,
- BP22: Provide an explanation for data that is not available,
- SBP2: Make it indexable by search engines,
- SBP13: Expose data through APIs.

These closer ties also triggered a shift in the approach to OGC web services.

Indeed, despite their major value to society, by the mid-2010s it was clear that environmental geographical data were still being underused by the web community. This could be explained by:

- the widespread use of resource-oriented architectures (REST APIs) based on identifying resources with URIs and using the HTTP protocol to interact with them,

---

<sup>10</sup> <https://www.w3.org/press-releases/2015/spatial/>

<sup>11</sup> <https://www.w3.org/2014/03/lgd/>

- the increasing maturity of protocols and development frameworks such as OData<sup>12</sup> and Swagger/OpenAPI<sup>13</sup>.

The analysis of this finding and the cooperation between the OGC and W3C led to a real paradigm shift in the specification of technical standards for the interoperability of environmental geographical data.

The OGC web services specifications then gradually moved away from the architectural styles of the 1990s/2000s (inherited from XML-RPC and SOAP) and evolved towards a more web-oriented approach, aiming for behaviours conforming to the REST architectural style<sup>14</sup> and providing resource representations more in tune with the expectations of developer communities (e.g. JSON instead of XML).

The result of this shift is that the digital resources that were to some extent "hidden" behind the old OGC web services approach can now be seen directly on the Web of Data, making them more visible and therefore more reusable.

However, it is too early to consider previous practices as "dead", as many communities and information systems were built and still function in this way. This means that any project based on FAIR Data must consider both approaches and, when possible, support the move towards the latest practices in order to maximise the FAIRification of data.

## FAIR initiative

As mentioned at the beginning of this section, the FAIR Data initiative is rooted in the publication by Wilkinson *et al.* (2016) and the creation of the Research Data Alliance in 2013.

However, before analysing the FAIR principles and defining how Water4All intends to apply them, the brief historical overview above serves to emphasise that these initiatives need to be seen as an ongoing effort to better share digital objects for both humans and machines. Water4All is positioned along this continuum.

Moreover, certain individuals/organisations are often very active in several of these bodies and initiatives (W3C, OGC, INSPIRE, RDA, etc.), which gradually leads to consistency in the approaches. They are therefore not antagonistic but complementary, since they share the same objective. Speaking of interoperability or FAIR is then roughly the same.

Thus, the pros/cons approaches for why interoperability/FAIR data matters also meet. The following image summarizes EU Water4All partnership vision on the topic.

---

<sup>12</sup> [https://en.wikipedia.org/wiki/Open\\_Data\\_Protocol](https://en.wikipedia.org/wiki/Open_Data_Protocol)

<sup>13</sup> [https://en.wikipedia.org/wiki/OpenAPI\\_Specification](https://en.wikipedia.org/wiki/OpenAPI_Specification)

<sup>14</sup> [https://ics.uci.edu/~fielding/pubs/dissertation/rest\\_arch\\_style.htm](https://ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm)



- Science: research reproducibility, discovery, access, reuse of data & code
- Public monitoring: enable cross boundaries work (international > continent > country > and below)
- IT: not reinventing standards (*ex: how to exchange a timeseries*)
  - ⇒ Rationalization of technologies (less technical debt), methods, recruited profiles and training
  - ⇒ Save of time & money, coherence
  - ⇒ Enterprise Architecture

- Inconsistencies in data & code exchange approaches > silos
  - Within & between administrative levels
  - Also withing organizations information system parts
- ⇒ Many barriers to F.A.I.R.
- ⇒ Risks of miss-interpretations thus un-informed decisions
- ⇒ Need for tons of converters accross systems
- ⇒ Loss of time and money (often hidden costs)
- ⇒ No Virtual Research Environment, Digital Twins/Shadows, AI, Data Spaces dynamically connected to data\*
- \* unless one wants to spend a lot of money for each reinvention of standards

Figure 1 – Water4all vision on why interoperability/FAIR data matters<sup>15</sup>

FAIR principles can be accessed and described in detail on the Go-FAIR website<sup>16</sup>. It is important to keep in mind that FAIR principles do apply to exchange with humans AND machines. And also that FAIR principles apply to the notion of digital objects (see note 'NB1' in the Introduction).

Each FAIR principle being a principle there is a need to map each of them to most recent norms, standards and best practices from the communities (ISO, OGC, W3C, INSPIRE, RDA) and FAIR APIs (OGC).

Water4All decisions relating to FAIR Data will be formalised in the form of FAIR Implementation Profiles (FIPs) for each digital resource using the Go FAIR FIP Wizard. This will provide a framework for setting up the OneWater Data Platform.

## FINDABLE

The first step in re(using) data is to find them. Metadata and data must be easy to find and accessible to both humans AND machines.

## F1: METADATA AND DATA ARE ASSIGNED A GLOBALLY UNIQUE AND ETERNALLY PERSISTENT IDENTIFIER

### Analysis

Each digital object exchanged in Water4all can be regarded as a resource in the sense of the Web/Web of Data. The goal is then to enable a technical organisation to be put in place to resolve and supply these resources, taking account of the fact that organisations and representations (e.g. of metadata and data) can change over time.

<sup>15</sup> « Global water security, climate change and the role of digital solutions in enhancing water resource management and resilience worldwide - Water4All vision » Aquatech 2025, Amsterdam. <https://brgm.hal.science/hal-05006009>

<sup>16</sup> <https://www.go-fair.org/fair-principles/>

The most future-proof solution, derived from the W3C's Spatial Data On The Web recommendations, is to use HTTP URIs to identify and link resources<sup>17</sup>. This then enables humans and computer programs to access them.

HTTP URIs provide a simple decentralised way of creating unique identifiers, as each owner of a domain name (or any entity to which it delegates management) can create new URIs. Good practice dictates that these domain names should not be correlated with the names of the bodies they represent, as these names can change over time (there is a very long list of public bodies whose names have changed).

These HTTP URIs serve not only as identifiers for declaring resources, but also as links for looking them up in different representations (the principle of content negotiation)<sup>18</sup>.

At the time of writing this document, the recommendations from the INSPIRE Directive (INSPIRE Metadata Guidelines<sup>19</sup>) had not yet been updated to join this initiative resulting from the W3C's "Spatial Data On the Web" Working Group.

### **Water4All recommendation**

To meet criterion F1, Water4All recommends assigning HTTP URIs to each digital object as far as possible. The domain name and the naming rule will be defined according to the level of detail concerned and the type of digital object.

## **F2. DATA ARE DESCRIBED WITH RICH METADATA (DEFINED BY R1 BELOW)**

### **Analysis**

The meaning of the term "metadata" is often overloaded by several different interpretations which, although complementary, result in the use of different standards and best practices.

These readings are as follows (using the terms from the French "Passport for Open Science"<sup>20</sup>):

- Documentary metadata provide more specific information about things like the institute and the individuals who produced the data, conditions of use and access, the dataset's persistent identifier, the identifier of the publications and software code linked to the data, etc.

These metadata describe the datasets, their distribution and the services used to access them. The reference standards in this case are: W3C:DCAT and OGC/ISO 19115. The use of W3C:DCAT for this purpose is part of Best Practice 2<sup>21</sup>.

This approach can also be supplemented by a description of the process that led to the creation of the dataset (use of PROV-O in ontology or LI\_Lineage in ISO 19115).

---

<sup>17</sup> <https://www.w3.org/TR/sdw-bp/#globally-unique-ids>

<sup>18</sup> Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. <http://linkeddatabook.com/editions/1.0/>

<sup>19</sup> <https://github.com/INSPIRE-MIF/technical-guidelines/tree/main/metadata>

<sup>20</sup> <https://www.ouvrirlascience.fr/open-science-research-data/>

<sup>21</sup> <https://www.w3.org/TR/dwbp/>

The metadata can also be supplemented by a description of the dataset structure. This case is often covered by the concept of Feature Catalogue (ISO 19110) and/or provision of a data model/ontology. These data models/ontologies are then in turn described by metadata.

- *Scientific metadata* (author's note: or structuring data) provide information about the data, in particular: protocol and context in which they were obtained, time references, settings of instruments used, analysis tools and software, etc. using the controlled vocabulary of the research field.

Such metadata describe how a measurement result was obtained. The reference standards are W3C:SSN/SOSA (there are other observation ontologies but SSN/SOSA has the advantage of being more consistent with ISO 19156) or OGC/ISO 19156 (Observations, Measurements & Samples).

To take this idea further, it is also necessary to consider that what is "data" for one use may be regarded as "metadata" for another: rainfall observation data may be regarded as scientific metadata for water quality observation data.

### **Water4All recommendation**

To meet criterion F2, Water4all recommends:

- for "metadata" describing the datasets, their distribution and the services used to access them, implement W3C:DCAT and its extension (for the ontology approach) and OGC/ISO 19115 (for the UML approach);  
in the case of "metadata" describing data models and ontologies, examine work on the extension of W3C:DCAT (e.g. MOD Ontology) and the use of ISO 19110;
- for "metadata" describing how a measurement result was obtained, implement SSN/SOSA (for the ontology approach) and OGC/ISO 19156 (for the UML approach).

The objectives in terms of "richness of data" will also have to be specified. An example of this is the exercise carried out by the French biodiversity center (PNDB<sup>22</sup>) on this subject<sup>23</sup>.

## **F3. METADATA CLEARLY AND EXPLICITLY INCLUDE THE IDENTIFIER OF THE DATA THEY DESCRIBE**

### **Analysis**

The standards mentioned in F2 natively include references (identifiers) to the datasets they describe.

### **Water4All recommendation**

By applying the standards identified in F2, F3 will be covered.

---

<sup>22</sup> <https://www.pndb.fr/en/home>

<sup>23</sup> <https://www.pndb.fr/fr/outils-services/partager-les-metadonnees#Quelles%20sont%20les%20m%C3%A9tadonn%C3%A9es%20minimales%20%C3%A0%20partager?>

## F4. METADATA AND DATA ARE REGISTERED OR INDEXED IN A SEARCHABLE RESOURCE

### Analysis

It is key to clarify the concept of searchable resource for this criterion.

If a digital object is to be searchable, its descriptive information must be indexed in a search system.

Two complementary approaches should then be followed:

- A°/ The portal approach: descriptive information about the digital object must be declared in a catalogue of digital objects. This is traditionally found in spatial data infrastructure under the terms data catalogues and metadata (logic to be extended to include all the digital objects that Water4All FAIR activities should cover),

This portal approach is also found with semantic artefacts (Linked Open Vocabularies – LOV, OntoPortal, etc.).

- B°/ The indexing engine approach: for several years now, the crawlers of the main search engines have been able to index resources with a certain degree of semantics (as opposed to indexing free text) based on the Schema.org vocabulary. To facilitate the finding of digital objects, many data portals have set up descriptions of their digital resources using Schema.org. There is also mapping between DCAT and Schema.org<sup>24</sup> (as an example, the INSPIRE Geocatalogue uses this).

In conjunction with criterion F1 above, this means that regardless of the approach adopted by the user (human or machine) to find a digital object, they will be redirected to the same place (to the same HTTP URI).

### Water4All recommendation

Water4All recommends implementing the cataloguing mechanisms corresponding to the digital objects identified with the communities (A°/ approach) by:

- using the most appropriate semantic and technical standards (catalogue query APIs) from the standards and best practices baseline (ISO, OGC, W3C, INSPIRE, RDA),
- providing interfaces for humans AND machines,
- and guaranteeing indexing using crawlers (B°/ approach).

---

<sup>24</sup> <https://ec-jrc.github.io/dcat-ap-to-schema-org>

# ACCESSIBLE

## A1. METADATA AND DATA ARE RETRIEVABLE BY THEIR IDENTIFIER USING A STANDARDISED COMMUNICATIONS PROTOCOL

### 1. A1.1 the protocol is open, free and universally implementable

#### **Analysis**

The term protocol can be understood in several ways, depending on the context.

For the purposes of this exercise, it will be restricted to exchanges of data in the software application layer on the Internet.

Two main use cases for accessing a digital object were identified:

- 1°/ The user knows the address of this object/resource and can therefore point to it using an HTTP URI (example of a link to a hydrogeological unit <https://data.geoscience.fr/id/hydrogeounit/121AS01> ),
- 2°/ The user does not know the address of the object or does not know exactly which object to point to. In this case the user must query the APIs (either directly or via a user interface) to find the answer to their questions (based on filters).

In case 1°, the protocol used must be Hypertext Transfer Protocol (HTTP), which is THE Internet protocol, as it is standardised, open and documented.

In response to case 2°, there is an underlying shift at international level towards simple, documented APIs (OpenAPI<sup>25</sup> or even OData<sup>26</sup> in some cases) that are open, standardised and use architectures such as REST. REST architectures are based on HTTP verbs.

#### **Water4All recommendation**

To deploy solutions based on the HTTP protocol and RESTful FAIR APIs defined by international standards and best practices (e.g. OGC APIs). In each case, these APIs will also need to be described in accordance with international best practices (e.g. OpenAPI, OData).

The additional tools deployed to meet other criteria (e.g. F4) must comply with the approach for this criterion A1.1.

### 2. A1.2 the protocol allows for an authentication and authorisation procedure, where necessary

#### **Analysis**

Following the same approach as for criterion A1.1, it may be added that:

---

<sup>25</sup> <https://swagger.io/specification/>

<sup>26</sup> <https://www.odata.org/>

- regarding HTTP, Hypertext Transfer Protocol Secure (HTTPS) is used to manage a secure connection to the data stream via authentication certification,
- in addition, regarding distributed/service-oriented architectures, FAIR APIs such as those of the OGC can be deployed in network environments where authentication and authorisation layers are present.

### **Water4All recommendation**

The implementation choices (HTTPS and OGC APIs) enable criterion A1.2 to be met.

## **A2. METADATA ARE ACCESSIBLE, EVEN WHEN THE DATA ARE NO LONGER AVAILABLE**

### **Analysis**

This good data management practice would avoid having to deal with an HTTP 404 error message in a scenario in which an external system had found a digital object of interest to it and pointed to it.

Taking this idea further, this practice should also be applied to the digital objects described by these metadata (e.g. the datasets, or even an individual data in a dataset described by the metadata).

Indeed, rather than deleting these digital objects and therefore providing an HTTP 404 code in response to a direct call to this object if the metadata had previously been removed (i.e. in the example given earlier, calling the HTTP URI <https://data.geoscience.fr/id/hydrogeounit/121AS01> without going via the metadata), it would be better to provide a minimal representation of it by marking it as "frozen", "deprecated" or "replaced by" depending on the case.

### **Water4All recommendation**

In response to criterion A2, Water4all recommends to:

- maintain the metadata representation of a digital object even if this object no longer exists (withdrawn) or is inaccessible (embargoed) as long as possible,
- also provide, where possible, a minimal representation of the digital object if it has been withdrawn by giving information on its status (e.g. "frozen", "deprecated", "replaced by", etc.).

## **INTEROPERABLE**

### **I1. METADATA AND DATA USE A FORMAL, ACCESSIBLE, SHARED, AND BROADLY APPLICABLE LANGUAGE FOR KNOWLEDGE REPRESENTATION**

#### **Analysis**

Just before the 2000s, the international communities began formalising knowledge in their area in UML models. This path was the basis for the OGC and ISO 191xx standards, the initiative of the INSPIRE Directive, other standards in the public & scientific communities (e.g. WMO/OGC WaterML2.0 through the OGC/WMO Hydro Domain Working Group, IUGS CGI/OGC GeoSciML) and European environmental reporting.

For several decades, this work was considered to be the only formal representation of data (including environmental data).

There is now a move to "transcribe" this knowledge into Semantic Web notations (RDF representation of knowledge).

This initiative linked to the movement started by the W3C (Data on the Web Best Practices<sup>27</sup> and then Spatial Data on the Web Best Practices<sup>28</sup>) led to:

- a description of the catalogue (dcat:Catalog), metadata record (dcat:CatalogRecord), dataset (dcat:DataSet) and service (dcat:DataService) using DCAT (or GeoDCAT) ontologies. A GeoDCAT\_AP application profile was even defined for data portals in Europe.
- a representation of Observations and Samples using the W3C:SOSA ontology and I-ADOPT for observed variables. This formalisation of observed variables according to I-ADOPT was also used for the area of water in Theia/OZCAR<sup>29</sup>.
- the structuring of controlled vocabularies using the W3C:SKOS ontology, among others, and applying the rules defined in the publication "Ten simple rules for making a vocabulary FAIR"<sup>30</sup>. Examples include initiatives such as Theia/OZCAR and EnvThes.
- and also use of ontologies derived from the "transcription" initiatives described at the beginning of this section (e.g. HY\_Features for surface hydrography<sup>31</sup>, GeoSciML Lite ontology for geological features, etc.).

It should be noted that in 2024, both practices still coexist because the uses, tools and habits of users in the communities converge but did not merge.

For example:

- many users (humans AND software) are unfamiliar with ontologies and only know how to manipulate data using representations from the world of databases or flat files that do not allow information to be represented according to Semantic Web standards or "simple" APIs (OpenAPI or even OData),
- and conversely, there are working methods that only capitalise on pre-existing ontologies without considering the work already carried out to formalise the area of water, albeit in UML notation.

A number of initiatives are attempting to reconcile the two (e.g. OGC Rainbow<sup>32</sup>, OGC building blocks). Another example is the work being carried out to update the W3C SSN/SOSA ontology. This has been done in such a way that it is the reference (canonical OWL representation) for the ISO/OGC 19156:2023 "Observations, measurements & samples" standard. Further more, links between these practices are therefore possible.

Nevertheless, the current situation will most likely persist for many years as it takes time to prepare a transition from one practice to another, and Water4All cannot afford to exclude either of them in its work.

---

<sup>27</sup> <https://www.w3.org/TR/dwbp/>

<sup>28</sup> <https://www.w3.org/TR/sdw-bp/>

<sup>29</sup> Coussot, C., Braud, I., Chaffard, V., Boudevillain, B., Galle, S., 2024. Implementing a new Research Data Alliance recommendation, the I-ADOPT framework, for the naming of environmental variables of continental surfaces, Earth Science Informatics, accepted.

<sup>30</sup> <https://doi.org/10.1371/journal.pcbi.1009041>

<sup>31</sup> e.g. [https://www.opengis.net/def/schema/hy\\_features/hyf/HY\\_Catchment](https://www.opengis.net/def/schema/hy_features/hyf/HY_Catchment)

<sup>32</sup> <https://www.ogc.org/resources/rainbow/#1-4-fair>

### **Water4All recommendation**

To meet criterion I1, Water4all will formalise knowledge on digital objects using formal descriptions in UML and ontologies (OWL or SKOS).

It will use both the UML models and ontologies mentioned in the analysis and will supplement them where necessary by contributing to the standardisation approach and international best practices (OGC, W3C, RDA, INSPIRE, etc.) especially on new arising topics in the community.

- In the case of the UML, the models from the communities will be reused if they conform to the approach of the ISO 191xx series (managed by ISO/TC 211<sup>33</sup>). If this is not the case, these models will have to be "ported", both in terms of IT notations and domain modelling.

They will also have to comply with a certain open-world assumption<sup>34</sup>. This term, usually associated with the Semantic Web, symbolises the act of looking at pre-existing models before reinventing the wheel. This intellectual mechanism can also be applied to UML modelling, given the number of existing business standards in 2024 (see the first paragraph of the Analysis in this section).

- In the case of ontologies, those from the W3C (and the W3C/OGC collaboration), or from working groups such as the RDA, will be preferred. The use of Semantic Web modelling languages (OWL, SKOS) for ontologies (or other semantic artefacts) will also be encouraged, along with re-use and alignment with appropriate existing ontologies/thesauruses appropriate

Water4All will do its utmost to create links between the two approaches: also using and generating the corresponding schemas (JSON-Schema, JSON-LD, XSD, etc.)

## **12. METADATA AND DATA USE VOCABULARIES THAT FOLLOW FAIR PRINCIPLES**

### **Analysis**

The term "vocabulary" can be understood in two ways:

The broadest meaning, which also embraces the concept of ontology. This has already been described in detail in the analysis of criterion I1.

And more restrictively, corresponding to the concept of controlled vocabulary. Regarding this concept, the publication "Ten simple rules for making a FAIR vocabulary" has already been mentioned.

- Concerning the first meaning, although they are indexed by search engines, they are not systematically offered in ontology portals (Linked Open Vocabularies – LOV, OntoPortal, etc.) or approaches such as the FIP Wizard<sup>35</sup>, because these initiatives are firmly rooted in the Semantic Web.

As mentioned in criterion I1, work is being carried out to ensure that the standards derived from the approaches are regarded as FAIR. Members of Water4All task on FAIR have the necessary connections

---

<sup>33</sup> <https://www.iso.org/committee/54904.html>

<sup>34</sup> [https://en.wikipedia.org/wiki/Open-world\\_assumption](https://en.wikipedia.org/wiki/Open-world_assumption)

<sup>35</sup> <https://w3id.org/fip/wizard>

with the groups responsible for standardising water standards and, more broadly, with the OGC groups in charge of this activity.

- Regarding controlled vocabularies, as mentioned in I1, the communities have already generated numerous vocabularies. Not all are currently regarded as FAIR, but some of them are being worked on to bring them into compliance.

#### **Water4All recommendation**

To meet criterion I2, and in addition to the approach already described for meeting criterion I1, Water4all will

- contribute to the drive to make pre-existing standards FAIR within its area of activity (water and the models concerned),
- identify controlled vocabularies that meet the needs expressed in Water4all projects and tasks, help bring those that are not FAIR into compliance and create new vocabularies if a need arises in an area that is not already covered.

## 13. METADATA AND DATA INCLUDE QUALIFIED REFERENCES TO OTHER METADATA AND DATA

#### **Analysis**

This criterion highlights two facts:

- Digital objects are linked (i.e. references to them are created),
- And the relationships between them are "qualified". This can mean that the semantics of the relationship between two digital objects is specified (e.g. hasResult from an Observation to its result in SSN/SOSA) or may even carry information itself (e.g. saying that a piezometer is monitoring a hydrogeological unit is one thing, specifying when this association was made, by whom and the uncertainty associated with this relationship is additional information).

It will be possible to establish permanent links if the digital objects have globally unique and persistent identifiers (see criterion F1). The choice for criterion F1 is to rely on HTTP URIs.

Then, according to the data models used (depending on the case), the relationships will be "just" semantised associations or qualified associations. It will then be necessary to see thematically to what extent an association needs to be refined (carrying additional attributes).

#### **Water4All recommendation**

The modelling choices made by Water4all (F2, I1 and I2) and the use of HTTP URIs will enable criterion I3 to be met.

Attention will also be paid to linking controlled vocabularies as much as possible, both with internal links and to external controlled vocabularies.

Formal alignments between ontologies and vocabularies will also be structured (e.g. SSSOM<sup>36</sup>)

---

<sup>36</sup> <https://mapping-commons.github.io/sssom/spec-intro/>

# REUSABLE

## R1. METADATA AND DATA HAVE A PLURALITY OF ACCURATE AND RELEVANT ATTRIBUTES

### 1. R1.1. Metadata and data are released with a clear and accessible data usage license

#### **Analysis**

The standards mentioned in criteria I1 and I2 can be used to link to licences.

Licence information will be formally represented (with a URI).

#### **Water4All recommendation**

The approach proposed in criteria I1 and I2 will enable criterion R1.1 to be validated.

In addition, and following the approach of the Theia/OZCAR information system, Water4All will:

- strongly recommend that data providers specify a licence,
- and recommend the CC-BY 4.0 licence whenever possible.

### 2. R1.2. Metadata and data are associated with their provenance

#### **Analysis**

The term provenance can be open to interpretation.

For example, documenting provenance may involve answering the following questions:

- What? title, description, measured variable + unit, object of interest
- When? time period
- Where? spatial extent + coordinate system
- Who? contacts – different contact roles
- How? level of data quality (raw, validated, developed), measurement protocol, instrument used
- Some readings may also limit it to "how"

Regarding the ontologies identified above, these concepts are already covered:

- Dataset metadata (dcat:DataSet or in the ISO 19115 sense)
  - By design, the properties of these models provide provenance information for the dataset described. More specifically:

- 19115: Lineage: provides information describing the provenance of the dataset, as well as the steps and algorithms used.
  - DCAT: also mentions the use of the PROV-O ontology to provide more details in the description of provenance.
- Observation (meta)data (W3C:SSN/SOSA or ISO/OGC: Observations, Measurements and Samples):
  - The model can already provide "provenance" information with regard to the observation result.

#### **Water4All recommendation**

The modelling choices made by Water4all (F2, I1 and I2) will enable criterion R1.2 to be met.

### 3. R1.3. Metadata and data meet domain-relevant community standards

#### **Analysis**

The context of the exchange of digital resources on "water" in various EU Countries, across Europe and internationally shows a dichotomy between organisations/information systems that:

- do their best to apply FAIR for Water practices identified in this document, and even drive forward the standardisation effort,
- partially implement these practices, often limiting themselves to the cataloguing part and the web-based map display service,
- redefine ad hoc technical and semantic standards locally.

In these three situations, the bodies in charge consider that, within their respective systems, they comply with the standards of their (local) community. This raises real concerns about integrating these sub-communities in terms of FAIR practices.

Help is needed to ensure that the international initiatives currently in place around water data (e.g. OGC/WMO WaterML2, OGC/ISO 19156, etc.) are applied within the various projects and organizations involved in Water4all while encouraging their wider adoption by identifying the obstacles to this adoption.

#### **Water4All recommendation**

The technical conclusions reached for each of the above criteria will help with the establishment and implementation of a FAIR for Water ecosystem.

In order to help reduce the level of community fragmentation with regard to the practices of exchanging digital resources and to achieve FAIR for Water in practice, the technical actions will need to work in conjunction with other activities within Water4All (ex: task C.4.1 with Research Infrastructure, the direct link with the funded projects, ...).



[water4all@agencerecherche.fr](mailto:water4all@agencerecherche.fr)  
[www.water4all-partnership.eu](http://www.water4all-partnership.eu)

Grant Agreement n° 101060874



**Co-funded by  
the European Union**