Draft ETSI EN 304 199 V1.0.0 (2026-03)

EUROPEAN STANDARD

Data Solutions (DATA);
Data catalogue implementation framework;
Guidelines for Data Catalogue Framework

Reference

DEN/DATA-00304199

Keywords

DATA, data collection, data interoperability, data sharing, semantic, terms, vocabulary

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI IPR online database.

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™**, **LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This draft European Standard (EN) has been produced by ETSI Technical Committee Data Solutions (DATA), and is now submitted for the combined Public Enquiry and Vote phase of the ETSI Standardisation Request deliverable Approval Procedure (SRdAP).

Proposed national transposition dates	
Date of latest announcement of this EN (doa):	3 months after ETSI publication
Date of latest publication of new National Standard or endorsement of this EN (dop/e):	6 months after doa
Date of withdrawal of any conflicting National Standard (dow):	6 months after doa

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

The present document is developed in the context of the European Commission's Standardisation Request C(2025) 4135 [i.1] addressing article 33 of the Data Act (Regulation (EU) 2023/2854).

This standardization request foresees the joint development by ETSI and CEN/CENELEC of seven deliverables, ETSI is primarily responsible for deliverables 4 and 5, while the other deliverables have a primary responsibility of CEN/CENELEC.

The present document addresses the 4th deliverable (The Data Catalogue Implementation Framework), and it aims at making the data within European data spaces visible, findable, and therefore ready for supporting the trusted data transactions concept.

It ensures that data products, including data catalogues, datasets, and their distributions, are described in a consistent way.

Its relation to the other deliverables is sequential and interdependent:

- The Harmonised standards on Trusted Data Transactions (Parts 1-3) define how to share data securely and fairly. The present document ensures that data products are structured, described, and listed in a consistent way across all sectors. It is an essential precursor to enable the automated execution of transactions (the goal of Deliverables 1-3).
- Regarding the implementation framework for semantic assets (Deliverable 5): the present document standardizes the metadata for discovery (e.g. title, publisher, availability), the subsequent deliverable on semantic assets provides the meaning of the data itself. They are complementary: the catalogue framework allows one to find a data product, while the semantic assets framework ensures its content can be correctly interpreted. Both are critical for full data and semantic interoperability.
- Regarding Quality and Maturity (Deliverables 6 and 7), a standardized catalogue is a key indicator of mature internal data governance (Deliverable 6) and is a measurable feature for assessing the maturity of a data space itself (Deliverable 7). A participant's ability to register and manage assets in a shared catalogue reflects the quality of their data governance.

1 Scope

The present document provides a framework for standardized catalogue metadata in data spaces, in support of the publication and discovery of data products and to assure findability of data products within and across data spaces.

The present document:

a) sets out the common catalogue metadata, the minimum set to be applied across all common European data spaces;

b) establishes rules on the setting out of catalogue metadata extensions, specific to a domain-specific data space or to a specific data spaces set.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found in the ETSI docbox.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents are necessary for the application of the present document.

[1] European Commission: "DCAT Application Profile for data portals in Europe".

[2] European Commission: "DCAT-AP reuse guidelines".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents may be useful in implementing an ETSI deliverable or add to the reader's understanding, but are not required for conformance to the present document.

[i.1] Commission Implementing Decision C(2025)4135 of 1.7.2025 on a standardisation request to the European standardisation organisations as regards a European Trusted Data Framework in support of Regulation (EU) 2023/2854 of the European Parliament and of the Council.

[i.2] SEMIC - Interoperable Europe.

[i.3] European Commission: "GeoDCAT Application Profile for data portals in Europe".

[i.4] mobilityDCAT-AP: "A mobility extension for the DCAT application profile for data portals in Europe".

[i.5] European Union: "StatDCAT Application Profile for data portals in Europe".

[i.6] HealthDCAT-AP.

[i.7] European Data Portal.

[i.8] EPOS-DCAT-AP: "An extension of the DCAT Application Profile for Research Infrastructures in the solid-Earth domain".

[i.9] ETSI TS 103 757: "SmartM2M; Asynchronous Contact Tracing System; Fighting pandemic disease with Internet of Things (IoT)".

[i.10] Language-DCAT-AP.

[i.11] W3C® Recommendation 22 August 2024: "Data Catalog Vocabulary (DCAT)".

[i.12] ETSI TR 104 244: "Data Solutions (DATA); Implementation of Data Catalogues with NGSI-LD and DCAT-AP".

[i.13] W3C® Recommendation 16 July 2020: "JSON-LD 1.1".

[i.14] W3C® Recommendation 25 February 2014: "RDF 1.1 Turtle".

[i.15] W3C® Recommendation 20 July 2017: "Shapes Constraint Language (SHACL)".

[i.16] Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use.

[i.17] W3C® Recommendation 15 February 2018: "ODRL Information Model 2.2".

[i.18] EN 18235-1: "Trusted Data Transactions - Part 1: Terminology, concepts and mechanisms", (produced by CEN).

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the following terms apply:

NOTE: The following terms are defined in EN 18235-1 [i.18]: data, data user, data sharing/data exchange, data licence term, data product, data producer, data provider, data rights holder, data space, data transaction, metadata, data sharing contract, trust service, trust anchor, data intermediary, data intermediation service, data catalogue, data usage consent, trust, trusted data transaction, data space participant/data space participant/ participant, data space participant role/data space participant role/participant role, party, data use, data sharing process/data exchange process.

catalogue metadata data model: data model describing the catalogue metadata used in a data space to increase the discoverability of data products within the data space and across other data spaces

data catalogue: organized inventory of data products published by one or more data providers that can be searched by data users

data product: data sharing unit, packaging data and metadata, and any associated licence term

data space: environment enabling trusted data sharing between participating parties, based on an agreed governance framework, along with an agreed set of policies, semantic models, standardized protocols, processes, and facilitating services

data space rulebook: publicly available data space register where the data space guidelines and any other governance agreement are registered and maintained

metadata: data defining and describing other data

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ACT	Asynchronous Contact Tracing
AI	Artificial Intelligence
API	Advanced Programming Interface
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
CoC	Code of Conduct
CSV	Comma-Separated Value document
CSW	Catalogue Service Web
DCAT	Data Catalogue vocabulary
DCAT-AP	Data Catalog vocabulary - Application Profile
DSO	Distribution System Operator
ECDC	European Centre for Disease Prevention and Control
EHDS	European Health Data Space
EMS	Energy Management Systems
EPOS	European Plate Observing System
EPREL	European Product Registry for Energy Labelling
ESA	Energy Smart Appliances
ETSI	European Telecommunication Standards Institute
EU	European Union
EV	Electrical Vehicle
GDPR	General Data Protection Regulation
GeoDCAT-AP	Geospatial DCAT-AP
HDAB	Health Data Access Bodies
HEMS	Home Energy Management Systems
HTML	HyperText Markup Language,
HVAC	Heating, Ventilation, Air-Conditioning
HVD	High Value Datasets
ICS	Integrated Core Services
INSPIRE	INfrastructure for SPatial InfoRmation in the European community
ISAC	Integrated Sensing and Communication
ISO	International Organization for Standardization
JSON-LD	JavaScript Object Notation for Linked Data
LDS	Language Data Space
LLM	Large Language Model
ML	Machine Learning
NAP	National Access Points
NGSI-LD	Next Generation Service Interfaces - Linked Data
NUTS	Nomenclature of Territorial Units for Statistics
ODRL	Open Digital Rights Language
PM	Post Meridiem
RDF	Resource Description Framework
SAREF	Smart Applications REFerence ontology
SEMIC	Semantic Interoperability Community
SHACL	SHApes Constraint Language
TCS	Thematic Core Services
TURTLE	Terse RDF Triple Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URLLC	Ultra-Reliable Low-Latency Communication

4 Catalogue metadata data model

4.1 Context

Data spaces [2] are environments designed to enable trusted data sharing between participants.

A fundamental requirement for realizing trusted data transactions within a data space is that participants can publish and discover data products. To meet this requirement, a data space provides guidelines consisting of:

- a catalogue metadata data model for describing data products; and
- a catalogue infrastructure for fair discoverability.

These guidelines and any other data governance agreement shall be registered in a publicly available data space rulebook.

The term catalogue metadata refers to the machine-readable representation of a description of a data product.

Data products are often described in data catalogues by a combination of dataset, dataset series, data service and distribution concepts.

Data products are usually not bound to a single data space. Other data users participating in other data spaces can take benefit from the provided data. Therefore, it is beneficial that data products can be discovered beyond the borders of a data space with a minimal publication effort. The present document sets a common basis for the metadata data model across data spaces.

To satisfy the requirements for a common metadata data model, the present document provides a reference metadata model. The present document shows how data spaces can make use of the reference metadata model to meet their needs.

The presented approach is not limited to data spaces. Any party involved in the development and implementation of data catalogues and metadata frameworks, aiming to make data assets discoverable, interoperable, and reusable, can find guidance in the presented approach.

4.2 DCAT-AP as catalogue metadata data model

To ensure interoperability between data spaces, a common data model is needed. This data model provides the formal semantic underpinning of the metadata. Having a common data model does not exclude the presence, nor the use of alternative metadata representations. However, data spaces that opt for this choice are required to provide all necessary means to make their metadata interoperable with the data model.

General provisions for a metadata data model in a data space are given in provisions 4.2.1 to 4.2.5. Provisions 4.2-6 to 4.2-9 subsequently establish DCAT-AP as the reference data model.

Provision 4.2-1: A data space shall have a common catalogue metadata data model to describe a catalogue of data products.

The catalogue metadata data model contains at minimum the metadata required for the data space to function. This includes, but is not limited to, information on the data access methods supported by the data product and licence terms determining the legal usage conditions of the data product. Depending on the needs, the catalogue metadata data model should be able to capture provenance information, including aspects such as data collection methodology, data lineage or data quality.

Provision 4.2-2: The choice of a catalogue metadata data model for a data space is part of the data space governance and shall be included in the data space rulebook.

This catalogue metadata data model is primarily used for metadata describing data products listed in the data spaces' data catalogues, but it is also used in messages exchanged during the negotiation phase of a data transaction.

Provision 4.2-3: Every data catalogue in the data space shall conform to the catalogue metadata data model of the data space.

Provision 4.2-4: Metadata exchanged to establish or to execute a trusted data transaction in a data space shall conform to the metadata data model of the data space.

The choice of a catalogue metadata data model for a data space is independent of the data format, access protocol, and access policy for metadata within the data space. It, however, shall be machine-readable and should be unambiguously transformable into RDF.

Provision 4.2-5: The machine-readable data format, access protocol, and access policy for catalogue metadata within a data space are part of its data space governance and shall be included the data space rulebook.

The Data Catalogue Application Profile (DCAT-AP) [1], developed and maintained by the SEMIC [i.2] community under the Interoperable Europe initiative, provides a mature, sustainable, and standardized framework for metadata description that enables different parties to interoperate at the semantic level. Its governance is organized as a public community guaranteeing access to the data model and enabling participation by any data space.

DCAT-AP [1] is a DCAT [i.11] profile. DCAT provides the universal terminology with minimal constraints. DCAT-AP augments DCAT with minimal metadata quality expectations, profiling and implementation guidelines for Europe.

DCAT-AP [1] is a data model for sharing information about catalogues containing dataset and data service descriptions in Europe. The core classes are the classes *catalogue*, *dataset*, *distribution* and *data service*. *Dataset series* enable further organizing datasets within a catalogue. Using these entities as building blocks, data products are usually a composition of the classes' dataset, distribution and data service. As data products will evolve over time, a modular approach for the description of data products facilitates the management of the information.

Provision 4.2-6: DCAT-AP [1] shall be used as the common data model.

A minimal model documenting the mandatory information for DCAT-AP can be found here [1].

Provision 4.2-7: The catalogue metadata data model shall be based on the terminology expressed in DCAT-AP for their metadata data model. Deviations or improvements shall be aligned with the DCAT-AP community. This common semantic basis guarantees interoperability across data spaces.

EXAMPLE: In DCAT-AP, a dataset is a conceptual entity, and a distribution is a downloadable expression of a dataset. Data spaces should not merge both notions into one entity in their metadata vocabulary.

Provision 4.2-8: Data spaces shall align their catalogue metadata data model with DCAT-AP. The alignment can be either:

a) by using DCAT-AP; or

b) by using an extension of DCAT-AP; or

c) by mapping to DCAT-AP.

Dataspaces shall include the chosen approach(es) as part of the data space governance in the data space rulebook.

A data space is strictly aligned with DCAT-AP when all the metadata is exchanged according to DCAT-AP. Typically this is implemented by enforcing DCAT-AP based exchange of metadata by all the data catalogues present in the data space.

Further elaborations of the three alignment approaches are respectively provided in clauses 4.3.1, 4.3.2 and 4.3.3 of the present document.

Provision 4.2-9: The catalogue metadata data model shall be versioned to support the evolution of the data space and its alignment with new releases of DCAT-AP. The creation of the catalogue metadata data model should start from the most recent version of DCAT-AP. The catalogue metadata data model should include the version of DCAT-AP is it based upon. DCAT-AP guarantees the persistency of its versions.

Provision 4.2-10:

The data space rule book shall include:

- the catalogue metadata data model allowed according to the choice made in provision 4.2-8;
- the roles and responsibilities of the publishers of the catalogue metadata;
- the data catalogues for the metadata;
- the methodology and the KPI used to monitor the quality of the catalogue metadata.

Data product publishers shall keep the metadata of the data products up to date and ensure their quality, to assure the well-functioning of the data space.

4.3 Guidelines for using DCAT-AP as catalogue metadata data model

4.3.1 Align the catalogue metadata data model with DCAT-AP

Clause 4.3 provides the guidance to data spaces to align their catalogue metadata data model with DCAT-AP.

The easiest and most straightforward approach is to use DCAT-AP directly. Data spaces shall first explore whether this option satisfies their needs. DCAT-AP is a broad and flexible catalogue metadata data model addressing many needs already. Data space may, however, add additional guidance for their data product publishers to tailor the model for specific purposes.

4.3.2 Using DCAT-AP

The exploration can be done in various ways; here follows the description of a high-level constructive step-by-step process that incrementally creates a catalogue metadata data model for the data space strictly based on DCAT-AP.

The process starts with a data model consisting only of the DCAT-AP main classes and their mandatory attributes. This is the minimal information any DCAT-AP conformant data catalogue shall provide. (See DCAT-AP reuse guidelines [2], Step 1 (https://semiceu.github.io/DCAT-AP-reuse-guidelines/#step-1-familiarise-with-dcat-ap)).

Then, iteratively:

1) Evaluate the metadata requirements of the data space w.r.t. the data model:

   a) Note down for each class and attribute additional guidelines for the data product publishers where needed.

   b) Record any detected unclarity or potential deviation the data space wants to make.

2) Select a use case requiring to use metadata that is not yet in the data model:

   a) Add all DCAT-AP attributes and guidelines fitting the use case.

This process shall be performed until all use cases are considered. The result is an annotated DCAT-AP specification describing the catalogue metadata data model with instructions tailored to the purpose of the data space.

Having established the semantic underpinning of the catalogue metadata, the data space decides the machine-readable data format, access protocol, and access policy for the catalogue metadata (provision 4.2-5) to be used. To a large extend these are independent considerations of the data model. DCAT-AP is a Semantic Web data specification. Despite that the DCAT-AP does not enforce the data exchange in implementations to be in native Linked Data formats (RDF serializations), this is expected. To ensure cross data space interoperability, the decision on the used data format shall be validated against the guidelines for DCAT-AP implementations, and in particular the use of identifiers (such as for example https://semiceu.github.io/DCAT-AP/releases/3.0.1/#identifiers).

4.3.3 Create an extension of DCAT-AP

Based on the gap analysis described clause 4.3.2 an extension of DCAT-AP for the data space can be created.

Before deciding how to resolve these gaps, the data space shall reach out to the DCAT-AP community to determine if any identified gaps are already addressed by DCAT-AP itself or another known DCAT-AP extension. The advice and suggestions received should be incorporated. If, after this consultation, the decision is still to create an extension, follow the DCAT-AP guidelines for creating extensions (See DCAT-AP reuse guidelines [2], Steps 2, 3 and 4 (https://semiceu.github.io/DCAT-AP-reuse-guidelines/)).

The data space governance for the catalogue metadata data model shall build consensus for the proposed resolutions for each identified gap. The decisions shall be recorded and published to all data space participants.

To maintain DCAT-AP alignment the data space shall engage itself in the DCAT-AP community. A concurrent benefit of this participation is the reduction of duplicate effort for publishers operating across several data spaces.

4.4 Mapping between the chosen catalogue metadata data model and DCAT-AP

As stated in provision 4.2-8, data spaces can adopt a catalogue metadata data model other than DCAT-AP. Motivations for this choice are multiple, but typically, it has its roots in the presence of tools and agreements already available in the ecosystem that initiates the data space. Switching to a new metadata data model and the associated supportive tools it can be considered too impactful.

To ensure the interoperability between data spaces, and in particular to ensure the discoverability of data products across data spaces, the use of another metadata representation than DCAT-AP mandates the need for a mapping between the chosen metadata representation and DCAT-AP.

Provision 4.4-1: Data spaces that adopt a catalogue metadata data models other than DCAT-AP shall develop and maintain a mapping between the chosen catalogue metadata data model and DCAT-AP. The mapping methodology and any semantic loss incurred during the process shall be documented.

Mappings between two data models A and B express how information encoded in one data model is to be interpreted in the other data model. A mapping is directional from A to B when this describes a deterministic, unambiguous method for converting formatted data according to data model A into an equivalent encoding in data model B.

The implementation of a mapping can be clearer or more straightforward in one direction than the other. A typical challenge arises when a single entity in data model A corresponds to multiple entities in data model B. In such cases, the mapping from B to A is often easier to implement, as merging multiple entities into an aggregate is generally simpler than splitting a single entity into its constituent parts.

Given that in general mappings rarely are one to one, providing documentation of the mapping decisions is essential.

Provision 4.4-2: Data spaces shall provide at least a mapping from the chosen catalogue metadata data model to DCAT-AP.

This direction enables the possibility that data products from the data space are discoverable by other data spaces. Enabling the mapping in the other direction eases the use of external data products by the data space.

Provision 4.4-3: The mapping shall cover the minimum DCAT-AP; all its main classes and their mandatory attributes.

To ensure that the data products from the data space are effectively discoverable by other data spaces, data space are expected to implement the mapping.

Provision 4.4-4: The Data Space should provide and maintain a conversion tool that implements the mapping.

EXAMPLE: The geospatial data space using the INSPIRE ISO XML metadata maintains a GeoDCAT-AP extension to DCAT-AP, which includes:

- The extra elements specific to geospatial data beyond the core DCAT-AP information.
- Guidelines for providing quality metadata.
- An XSLT transformation providing a reference implementation of a mapping from INSPIRE ISO XML.

Creating a mapping may follow a similar process as described before on the use of DCAT-AP:

1) First, the catalogue metadata data model is compared to the minimal DCAT-AP model:
    a) Note down for each class and attribute the mapping that needs to be performed.
    b) Record any detected unclarity or potential blocking point in the conversion.
2) Select a non-addressed class:
    a) Add all DCAT-AP attributes and guidelines fitting the use case.

This process shall be performed until all the metadata of the chosen data model is processed. The result is an annotated comparison where for each class and attribute in the catalogue metadata data model a conversion guideline to DCAT-AP is formulated.

5 Compliance validation

Data spaces shall ensure the compliance of the exchanged data product metadata with the made data space agreements (i.e. provision 4.2-1).

Provision 5-1: Data spaces shall provide the necessary means for data space participants to validate the conformity of the shared data product metadata with respect to the data space catalogue metadata data model.

When the data space adopts DCAT-AP as catalogue metadata data model, these necessary means may be provided by the DCAT-AP community. When the data space adopts the mapping approach, the data space may rely on the chosen metadata data model community.

Multiple, complementary testing methods exist to verify the compliance with DCAP-AP. Some target directly the metadata data model; others test whether the exchanged metadata is compliant with the metadata data model. These methods are complementary to each other in order to create evidence of the correct intended behaviour.

As part of the governance of the catalogue metadata data model, data spaces shall provide the means to discuss on the detected compliance issues.

Compliance of the metadata data model

Provision 5-2: Data spaces shall share the compliance assessment of the catalogue metadata data model (e.g. including the compliance assessment in the catalogue metadata data model).

Provision 5-3: The compliance with DCAT-AP should be provided by adding references to the corresponding DCAT-AP terms.

The inclusion of these references in the DCAT-AP usage guidelines for the data space is not only valuable for the compliance check, but also aids in future updates to align with newer versions of DCAT-AP.

Testing whether one data model is compliant with another data model is, in general, a process involving human activity. Each aspect of data model has to be compared, and this activity is also part of the creation process of metadata data model. That information is thus valuable knowledge for the compliance assessment.

Complementary automated checks should also be applied, e.g. using the SHACL [i.15] shapes associated with the distinct profiles in order to identify differences. For compliance checks, each detected difference should correspond to a valid extension of DCAT-AP introduced by the data space's metadata model.

Compliance through implementation of a compliant catalogue metadata exchange

A metadata exchange occurs when the provider (typically the data product owner) and the receiver (typically the data catalogue) transfer data using an agreed-upon protocol and an accepted format. The exchange is considered DCAT-AP compliant if the payload - the data being transferred - conforms to the DCAT-AP specification.

DCAT-AP provides a SHACL [i.15] representation of the data model. The DCAT-AP community is also provided with an operational environment to perform a compliance check of the exchanged data using the provided SHACL [i.15] representation.

Provision 5-4: The compliance of exchanged data could be validated using the DCAT-AP SHACL [i.15] representation.

While a sample-based check provides insight into expected general compliance, a systemic and integrated check applied to all exchanged data ensures a continuous validation.

6 Considerations for creating a catalogue metadata data model

This clause considers topics that require attention when establishing the data space catalogue metadata data model.

Controlled vocabularies (codelist) usage

The use of shared controlled vocabularies (also called code lists) for the value spaces of properties is encouraged to increase interoperability.

DCAT-AP has identified four expressions to make the usage intentions of a codelist in the data model and its dependent data models clear. (see for example https://semiceu.github.io/DCAT-AP/releases/3.0.1/#expected-usage-of-controlled-vocabularies).

Data Product terms of use

The catalogue metadata data model supports the expression of key information of a data product such as terms of usage, legal and commercial terms, consent and authorizations, etc. DCAT-AP offers a variety of metadata elements that can support this need. There is the ability to mention the accessibility level (whether or not the data is public accessible or has a restriction in place) but also to express the licence or individual rights that apply. (see for example https://semiceu.github.io/DCAT-AP/releases/3.0.1/#legal-information).

For a machine-readable representation of the applicable terms of use, the Open Digital Rights Language (ODRL) is recommended. ODRL is a policy expression language that provides a flexible and interoperable information, vocabulary, and encoding mechanisms for representing statements about usage (i.e. permissions, prohibitions, and obligations) of content and services (see ODRL Information Model 2.2 [i.17]).

Multi-profile access to data product metadata

Data product catalogues acting in data spaces facilitate mainly, but not only, the discoverability of the data products. This makes its target audience diverse in nature, ranging from human interfaces to support for machine interaction.

To handle this diversity, catalogues are recommended to implement content negotiation, i.e. the ability to retrieve depending on the users' requirements the most appropriate response for a request of information about a data product. Typically, the content negotiation by profile serves both human-readable (i.e. HTML) and machine-readable (e.g. RDF in serializations in TURTLE [i.14] and JSON-LD [i.13]) representations of their catalogue data.

Support for Artificial Intelligence and Machine Learning

To support the European AI ecosystem and ensure alignment with relevant regulatory frameworks (e.g. the European AI Act), the catalogue metadata data model needs to include capabilities to assess the suitability of data products for Artificial Intelligence (AI) and Machine Learning (ML) purposes.

Data spaces are encouraged to provide metadata attributes that allow AI agents and developers to evaluate data suitability for model training, testing, and validation:

- Data Quality: Metrics regarding the accuracy, completeness, and consistency of the data.
- Bias and Representativeness: Information on the diversity of the dataset and potential known biases to ensure ethical and fair AI outcomes.
- Lineage and Provenance: Detailed history of the data collection, transformations, and original sources to ensure traceability and trust.

It is suggested that data spaces operating according to clause 4.3 of the present document link DCAT-AP metadata with specialized AI-related metadata. This linking is essential to enable the automated discovery and ingestion of data by ML pipelines, reducing manual intervention in the data preparation phase.

7 ETSI EN 304 199 Management Framework

7.1 Motivation and objectives

The Management framework is designed to ensure that the present document evolves in a manner that is both reliable for implementers and responsive to technological and regulatory advancements. Its core objectives are to support:

- **Stability & Market Deployment:** To create standards that enable interoperability, foster investment, and ensure stable deployment and an effective and smooth evolution of technologies.
- **Evolution & Innovation:** To efficiently incorporate technical innovations, market requirements, and evolution of the regulatory requirements.

- **Relevance & Timeliness:** To develop standards swiftly to keep pace with the fast-moving telecom/IT sector, ensuring they are relevant and adopted in the UE and benefit of a global relevance.

Provide a clear, structured process to incorporate necessary changes based on user experience, technological advancements, and changes in foundational standards.

ETSI has a well-established, proven, extensive, long-term, stable framework for the development and maintenance of standards.

Clause 7 is intended to give some basic indications and some specific information applicable to the present document, and does not replace neither modify the procedures and guidelines as defined on the ETSI website.

7.2 Governance

The present document is developed and maintained according to the ETSI procedures by the ETSI Technical Committee Data Solutions, following the well-established ETSI Consensus-Driven approach.

It is also subject to a period of review by the ETSI membership and a formal approval by the EU National Standard Bodies.

The overall process is described in https://www.etsi.org/standards/standards-making?highlight=WyJhZmZlY3RlZCJ.

7.3 Release cycling and long-term maintenance

Releases are feature-driven and based on the completion of work items. The process is fully transparent and requires ETSI members to submit a Work Item Proposal which includes a detailed scheduling. The next steps are a technical review by TC Data Solutions and a formal approval by the ETSI NSB group. Then, TC Data solution will add it to its work program.

The stability of ETSI guarantees a proper long-term maintenance of the present document, subject to market needs and ETSI Members contributions.

7.4 Versioning and editorial quality

ETSI versioning is linked to the document releases identification.

The format is ETSI EN 304 199 VX.Y.Z where:

- Z is a sequential number indicating non-technical changes (such as editorial changes) e.g. from 2.2.1 to 2.2.2
- Y is a sequential number indicating technical changes (such as a new feature or a feature modification) e.g. from 2.2.2 to 2.3.0
- X is a sequential number indicating a new Release e.g. from 2.3.0 to 3.3.0)

The versioning system for the ETSI standards is described in https://portal.etsi.org/Services/editHelp/How-to-start/Document-procedures-and-types/Version-numbering-system for further details.

ETSI editHelp! provided full support to the publication with guides and drafting rules, and a proactive process of revision, assuring the formal quality of the present document (https://portal.etsi.org/Services/editHelp).

7.5 Accessibility

All the ETSI published standards are publicly accessible via the ETSI website.

Annex A (informative): Examples of a DCAT-AP Extension

A.1 EPOS-DCAT-AP

European Plate Observing System (EPOS) is a pan-European research infrastructure that integrates data, services, and equipment across the solid-Earth sciences - including seismology, volcanology, geodesy, geomagnetism, and geological modelling. Its aim is to enable multidisciplinary research by making diverse resources findable, accessible, interoperable, and reusable (FAIR).

To support this, EPOS developed EPOS-DCAT-AP [i.8], an extension of DCAT-AP that provides a domain-specific metadata model to describe the wide variety of assets managed by the EPOS community. While DCAT-AP offers a robust, generic framework for describing datasets and services, it does not fully capture the complexity of EPOS resources - which include not only datasets but also services, software, equipment, facilities, and related publications.

EPOS-DCAT-AP was therefore created to extend DCAT-AP while remaining fully interoperable with it. The extension reuses the DCAT-AP core structure and vocabulary but adds new elements to better reflect the specificities of the EPOS infrastructure.

Key features of the extension include:

- Additional classes, such as epos:Equipment, epos:Facility, epos:SoftwareApplication, and epos:Publication, to describe assets beyond standard datasets and services.
- Extended properties and constraints, for example:
  - epos:dynamicRange, epos:resolution, and epos:samplePeriod for the *Equipment* class, capturing technical instrument characteristics.
  - Extended use of dcat:publisher and dcat:contactPoint to align with schema:Organization and schema:ContactPoint.
  - epos:hasQualityAnnotation to capture dataset quality information.
- Integration of additional vocabularies, such as schema.org and hydra, to describe APIs, service parameters, and operations.
- Controlled vocabularies defined using SKOS, for consistent categorization of resource types, scientific domains, and instrumentation.

This extension enables EPOS to harvest and integrate metadata from its distributed Thematic Core Services (TCS) into a central catalogue managed by the Integrated Core Services (ICS). It demonstrates how DCAT-AP can be specialized for a specific community while maintaining semantic and structural compatibility with the European data catalogue ecosystem.

EPOS-DCAT-AP thus serves as a practical example of extending DCAT-AP to meet domain-specific needs in a structured and interoperable manner.

A.2 Language-DCAT-AP

Large Language Models (LLMs) such as BERT, GPT-3, ChatGPT, and GPT-4 represent one of the most transformative breakthroughs in artificial intelligence in recent years. These models are trained on massive volumes of language and multimodal data, often encompassing dozens or even hundreds of terabytes. However, within Europe, there is a major imbalance: while English benefits from abundant data resources, most other European languages remain severely under-resourced. This imbalance limits the ability of European LLMs and AI systems to reflect the continent's linguistic and cultural diversity, highlighting the urgent need for a coordinated effort to collect, organize, and share multilingual data across all European languages.

To address this challenge, the Language Data Space (LDS) was established as one of the 14 official EU Data Space projects. Its core objective is to develop and deploy a European platform and marketplace for the collection, creation, sharing, and re-use of multilingual and multimodal language data. Implemented as a procurement action (CNECT/LUX/2022/OP/0026), the LDS provides a governance framework and technical infrastructure based on openness, transparency, and collaboration among stakeholders from industry, research, public administration, cultural associations, NGOs, and citizens.

To facilitate semantic interoperability, the LDS created and adopted a common metadata model LanguageDCAT-AP [i.10], as an extension of DCAT-AP tailored to the needs of the language data community. It combines DCAT-AP properties for generic concepts with the META-SHARE ontology to describe language-specific data concepts.

Key features of the extension include:

- Re-using DCAT-AP properties, classes and recommendations as much as possible.
- Minimize to-the-extent possible mandatory properties (trying to balance between the consumer's requirements and the provider's wishes).
- Adding community properties: types and subtypes of dataset/language resource: for instance, dcat:Dataset is used for all data, but also lrType property is used with a controlled vocabulary to distinguish between "corpora" and lexical/conceptual resources (e.g. terminological glossaries, thesauri, gazeteers, morphological lexia, etc.); GDPR aspects, content (e.g. annotation types, model functions, etc.).
- "Changing" some data types: e.g. free text properties with range rdf:langstring instead of rdf:Literal.
- Mandatory/Recommended controlled vocabularies:
    - keep where possible, e.g. "dct:spatial" (EU Vocabularies Continents, EU Vocabularies Countries, EU Vocabularies Places), "dcat:mediatype" (IANA Media Types);
    - for language data specific properties, use own vocabularies (e.g. size units, model functions, standards, etc.) and request for additions in the EU vocabularies and, then adopt the EU vocabulary.

A.3 mobilityDCAT-AP (NAPCORE)

The National Access Point Coordination Organisation for Europe (NAPCORE) coordinates European National Access Points (NAPs) for mobility/ITS data, with the aim of improving interoperability and harmonising standards and access across borders. It is co-funded under the Connecting Europe Facility and acts as the coordination mechanism for NAPs as the backbone of European mobility data exchange.

mobilityDCAT-AP [i.4] is a NAPCORE application profile that extends DCAT-AP for describing mobility and ITS-related datasets and services on NAPs and other mobility data portals. It provides a structured, interoperable and harmonised way to describe and exchange metadata, improving cross-border and cross-sector discoverability. An authoritative online specification and implementation resources (serializations, SHACL [i.15], examples) are maintained openly.

DCAT-AP is cross-domain and generic; mobility portals and NAPs require domain-specific semantics, tighter constraints, and selected controlled vocabularies (e.g. ITS themes, conditions for access/usage) to enable consistent discovery and automated exchange across Member States. mobilityDCAT-AP was therefore developed by the NAPCORE Metadata Working Group as a DCAT-AP-conformant extension tailored to mobility, aligning closely with DCAT-AP editors to ensure interoperability.

The specification reuses DCAT-AP v3 and adds/adjusts classes, properties, constraints, and vocabularies (some reused from geoDCAT-AP v3), while also de-scoping some optional DCAT-AP terms to minimize ambiguity. A non-exhaustive list of notable changes follows (mobilityDCAT-AP [i.4]).

Added / emphasized classes

- dcat:DatasetSeries (to reflect DCAT-3 notions).

- Mobility-specific/supportive classes such as Mobility Data Standard, Quality Annotation, Assessment, and selected organizational/contact entities; supportive location/contact classes align with SEMIC and vCard/Core Location patterns.

Property additions and refinements

- On dcat:Dataset: dcat:inSeries, foaf:page, dct:source.
- On dcat:Distribution: foaf:page, dct:language, dct:modified, dct:issued.
- On Mobility Data Standard: mobilitydcatap:specificContentModel.
- Obligation levels of some DCAT-AP properties are tightened, e.g. making dcat:record and dcat:distribution mandatory to ensure complete catalogue/record/dataset/distribution chains.

Controlled vocabularies

- Mandated/selected CVs include: EU Authority Table "Frequency" (replacing an earlier mobility list), Mobility Theme (with links to applicable EU delegated regulations via dcatap:applicableLegislation), and Conditions for Access and Usage (extended value set). Spatial references align with NUTS codes.

Conformance & validation

- The spec provides RDF serializations and SHACL [i.15] shapes for automated validation; guidance and a "minimum profile" and example populations are included. A form-based metadata generator UI is also available for implementers.

In summary, mobilityDCAT-AP is a domain-specific DCAT-AP extension for mobility/ITS that standardizes how NAPs and mobility portals describe datasets and access, ensuring semantic consistency, machine-readability, and cross-border interoperability within the European mobility data.

A.4 HealthDCAT-AP

The European Health Data Space (EHDS) Regulation introduces specific obligations for the description and cataloguing of datasets made available for the secondary use of health data. In particular, Article 77 requires Health Data Access Bodies (HDABs) to provide a public, standardized and machine-readable catalogue describing available datasets and their key characteristics (including, among others, source, scope, main features, the nature of the data in the dataset, and the conditions under which the dataset can be made available).

In this context, HealthDCAT-AP is an application profile conformant with DCAT-AP, designed specifically to describe health datasets within the Health-Data@EU/EHDS setting. HealthDCAT-AP is explicitly defined as an extension of DCAT-AP v3.0: it reuses the core DCAT-AP classes (e.g. *dcat:Catalog, dcat:CatalogRecord, dcat:Dataset, dcat:Distribution, dcat:DataService*) and introduces additional classes, properties, constraints and controlled vocabularies to address health-specific requirements without compromising interoperability with the wider DCAT-AP ecosystem.

HealthDCAT-AP extends the generic baseline by introducing an explicit access-level model that categorizes datasets into non-public, restricted, and public tiers. This distinction drives differentiated metadata requirements, enabling the standard to handle complex health scenarios such as mediated access for personal data versus conditional access for non-personal data. This is reinforced by specific governance properties, including the mandatory inclusion of a Health Data Access Body (HDAB) and refined publisher semantics that identify trusted data holders in alignment with European Health Data Space (EHDS) regulations.

The standard significantly enhances dataset characterization to address the specific needs of health analysis. It introduces mandatory health categories and detailed properties for population context, such as minimum and maximum typical age, population coverage, and specific magnitudes like record counts and unique individuals. Furthermore, it strengthens semantic interpretability by including properties that explicitly state the coding systems and code values used within the data.

To manage the high sensitivity of the domain, HealthDCAT-AP integrates specialized vocabularies for machine-readable governance. It utilizes DPV for indicating personal data and legal bases, DQV for quality annotations, PROV-O for provenance, and ODRL for usage policies. Additionally, it implements a practical pattern for variable documentation using CSVW-based entities (Table, Column) linked to synthetic or anonymous sample distributions, improving discoverability without exposing sensitive underlying data.

Conformance is enforced through a layered validation framework that builds upon DCAT-AP prerequisites. The extension provides specific SHACL [i.15] shapes tailored to each access level, supporting the automated validation of structural constraints and controlled vocabularies. This ensures that datasets, particularly non-public ones, strictly adhere to the mediated-access model by requiring the presence of HDAB-related distributions and precise governance metadata.

A.5 AI-Ready Data

While DCAT-AP serves as the essential backbone for the discovery of datasets, it often needs to be complemented by specialized metadata standards and practices that describe the technical nuances and the internal structure required by Machine Learning (ML) workflows and AI pipelines.

Through the use of dct:conformsTo or dcat:accessService properties, a DCAT-AP record can point to more granular metadata structures:

- MLDCAT-AP: An extension of DCAT-AP specifically designed to describe Machine Learning assets. It extends the catalogue level to include descriptions of ML models, software, and specific performance metrics, allowing for the discovery of models alongside the data used to train them.
- Dataset Cards and Model Cards: Essential transparency tools that provide a standardized way to document the "biography" of a dataset or model, including its intended use, limitations, and ethical considerations.
- CSVW (CSV on the Web): A crucial tool for describing the internal content and fields of tabular datasets. It defines column types and semantic meanings, allowing AI agents to parse the data without ambiguity.
- Data Package: A container format used to wrap data files with their corresponding schema and metadata, ensuring that the structural description and provenance travel with the data itself.

Annex B (informative): Use Cases examples

B.1 Cognitive traffic management with Cellular sensing and Smart City data

B.1.1 Overview

A city integrates real-time "sensing-as-a-service" data from its 6G cellular network with its existing static and dynamic urban data. Using a DCAT-AP-based catalogue as the discovery and integration layer, the city's Traffic Management System creates a self-healing, proactive urban traffic ecosystem that minimizes congestion, prioritizes emergency response, and optimizes the flow of all road users (vehicles, public transport, cyclists, pedestrians) by fusing real-time 6G sensing data with rich contextual smart city data.

NOTE: The intention here is to show an example of a potentially relevant use case. Some hypothetical extension and data that are used in the description of this use case are not necessary part of current DCAT AP framework, neither are intended to be a pre-emption of its evolution.

B.1.2 Actors

Data Providers:

- 6G Network Operator: Provides anonymized and aggregated sensing data (e.g. user equipment location, speed, spatial density) and network data (e.g. handover rates, signal strength).
- City Municipality: Provides city datasets (roads, traffic lights, points of interest, etc.).
- Traffic management system: Provides real-time traffic signal status, variable message signs, and traffic sensor data, etc.
- Emergency Services: Provides real-time location and status of ambulances, fire trucks, and police cars, current Emergency Departments load an availability.
- Event Organizers: Provide event schedules, locations, and expected attendance.

Data Consumers:

- AI-Powered Traffic Management System: The primary consumer that ingests all data to make control decisions.
- City Dashboard & Operator: For human-in-the-loop monitoring and override.
- City traffic Mobile App: Receives filtered information for public alerts and route suggestions.

Enabling Technologies:

- 6G Cellular Network: Provides Ultra-Reliable Low-Latency Communication (URLLC), integrated sensing and communication (ISAC), and massive connectivity.
- DCAT-AP-based City Data Catalogue: Serves as the basis for all data assets, describing their provenance, format, location, and indicating its semantics links.
- AI/ML Analytics Engine: Processes the fused data for pattern recognition, prediction, and optimization.

B.1.3 The Role of DCAT-AP and its Extensions

DCAT-AP provides the standardized metadata framework that solves the core challenge of data findability across the independent data spaces of the various actors. It creates a federated catalogue where any data consumer can discover and understand available data assets, making integration seamless and scalable.

Core DCAT-AP for Discovery:

- Each dataset (e.g. "Real-time UE Density Map," "Hospital Locations," "Traffic Signal Inventory") is registered with a common vocabulary as a dcat:Dataset.
- Standardized metadata includes dct:title, dct:description, dct:publisher, dct:issued, and dct:modified and ensure basic findability and understanding.
- The access endpoint (dcat:accessURL or dcat:downloadURL) provides the location to Traffic management System to fetch the live data.

Critical Extensions for Enhanced Semantics:

- STAT-DCAT-AP [i.5] (for spatial data): Crucial for geo-referencing. It specifies the spatial coverage (dct:spatial) of each dataset (e.g. the city boundary, a specific district), enabling consumers to find all data relevant to a specific geographic area.
- GeoDCAT-AP [i.3]: Further enriches spatial descriptions for interoperability with GIS systems.

Provenance & Quality Extensions: Future extensions could include:

- Data Freshness: A property like xsd:duration to indicate "updated every 500 ms" for 6G data vs. "updated every 5 minutes" for legacy sensors, allowing consumers to find data streams that meet their latency requirements.
- Accuracy/Uncertainty: Metadata describing the confidence level of the 6G sensing data (e.g. location accuracy of ±50 cm).
- Licensing & Usage Rights: Clearly defines how sensitive data like UE density can be used (anonymized, aggregated-only).
- The Traffic Management System discovers necessary data by querying the DCAT-AP compliant catalogue for datasets with specific properties (e.g. theme: "traffic", spatial: "City-Centre", temporalResolution: "<1s"). The catalogue returns the end points, and the Traffic Management System can automatically subscribe to the relevant data streams.

B.1.4 Scenarios

Scenario A: Real-Time Emergency Vehicle Priority & Jam Avoidance

1) Trigger:
   - An ambulance is dispatched from a hospital. Its onboard system automatically publishes its real-time location, destination, and priority level to a city dataset.
2) Data Fusion:
   - The Traffic Management System discovers the "Active Emergency Vehicles" dataset via DCAT-AP.
   - It simultaneously ingests the "Real-Time UE Velocity" map from the 6G network, which shows a traffic jam forming around on the ambulance's planned route.
   - It cross-references this with the "Traffic Light Network" dataset.
3) AI Analysis & Action:
   - The AI predicts the ambulance's time of arrival at each intersection and the jam's location.

- It calculates an optimal "green wave" corridor, dynamically adjusting the traffic lights ahead of the ambulance.
- Concurrently, it sends instructions to variable message signals on the approach to the jam, instructing vehicles to clear the left lane.

4) Result:
   - The ambulance encounters a pre-cleared path, reducing its response time.

Scenario B: Predictive Traffic Management for a Major Concert

1) Trigger: An "Event" dataset is updated in the DCAT-AP catalogue with details of a stadium concert starting at 8:00 PM at a given date.
2) Data Fusion:
   - The Traffic Management System analyses historical 6G sensing data from similar events to model crowd arrival/departure patterns.
   - On the day, it fuses the real-time 6G UE density data with the event location and public transport schedules.
3) Predictive Analysis & Action:
   - The AI predicts a 90 % probability of a major jam on the main access road by 7:45 PM.
   - Proactively, the system:
     - Adjusts traffic light cycles on feeder routes to slow the inflow of vehicles.
     - Sends push notifications to public navigation apps suggesting alternative routes and highlighting public transport options.
     - Alerts public transport operators to prepare for higher demand.
4) Result:
   - The jam is mitigated before it even fully forms. Overall traffic flow in the city is less disrupted.

Scenario C: Dynamic Resolution of a Non-Emergency Traffic Jam

1) Trigger:
   - The 6G network's sensing data detects an anomalous, growing cluster of stationary and slow-moving UEs on a highway off-ramp.
2) Data Fusion & Diagnosis:
   - The Traffic Management System checks the "Roadworks" dataset via DCAT-AP. There are no scheduled works.
   - It checks "Traffic Signal Status" - all signals are operational.
   - It analyses the shape and movement of the UE cluster, identifying it as a bottleneck caused by a poorly timed traffic light at the end of the off-ramp.
3) AI Analysis & Action:
   - The AI runs a micro-simulation and identifies that increasing the green time for the off-ramp lane by 20 % for the next 15 minutes will dissolve the jam.
   - It implements this change automatically.
4) Result:
   - The jam dissolves without human intervention, improving air quality and driver satisfaction.

B.1.5 Considerations

In this use case, the adoption of a structured semantic framework for DCAT-AP is key to unlocking data findability. It moves the system from a collection of siloed, reactive components to a single, intelligent, and predictive organism that can dynamically optimize the complex ecosystem of urban mobility for the benefit of all citizens and services. Some of the benefits are listed below:

- Data Findability: DCAT-AP provides a common language for describing data, making it possible to automatically discover and access relevant datasets across the independent data spaces of the network operator, municipality, and emergency services. Without this, data remains locked in silos and cannot be fused for a holistic view.
- Enhanced Public Safety: Faster emergency response times save lives, enabled by the system's ability to instantly find and integrate real-time vehicle location with traffic and road data.
- Optimized Traffic Flow: Reduced congestion, shorter travel times, and lower fuel consumption/emissions are a direct result of being able to discover and utilize real-time sensing data and correlate it with historical pattern.
- Proactive vs. Reactive Management: The system predicts and prevents problems rather than just responding to them, finding and querying event data and historical models in advance.
- Resource Efficiency: Dynamic use of existing infrastructure (roads, traffic lights) reduces the need for costly new construction.
- Data-Driven Decision Making: City planners use the fused historical data for long-term urban planning, relying on the catalogue to find all relevant past datasets for analysis.
- The DCAT-AP framework ensures that as new data providers (e.g. connected vehicle fleets, drone delivery services) emerge, their data can be immediately discovered and understood, allowing for seamless future integration and new services offering.

B.2 Cross-Border Pandemic Preparedness with Asynchronous Contact Tracing and Smart Health Data

B.2.1 Overview

Multiple European countries deploy the ETSI-standardized Asynchronous Contact Tracing (ACT) [i.9] system to monitor environmental samples (waste water, air filters) for pandemic pathogens. Each national health service operates its own ACT Control Service, generating valuable data on pathogen detection.

To overcome data silos and enable a proactive, pan-European health response, these national datasets are registered in a central, federated catalogue based on DCAT-AP and hosted on a platform like data.europa.eu [i.7]. This catalogue acts as the discovery layer, allowing public health agencies, researchers, and international bodies (e.g. ECDC) to find, understand, and integrate ACT data with other contextual datasets (e.g. population mobility, hospital capacity). This creates a powerful "immune system" for the continent, enabling early detection of cross-border health threats and data-driven policy decisions.

B.2.2 Actors

Data Providers:

- National Health Service (Country A): Provides the ACT Control Service dataset, containing anonymized detection alerts (BSSID, Geohash location, TEST-RESULT, TEST-TIME).
- National Health Service (Country B): Provides a similar ACT Control Service dataset.
- Catalogue Host: European Data Portal (data.europa.eu) - Hosts the central federated DCAT-AP catalogue.

- Eurostat: Provides contextual datasets like "Population Density by Region" and "Cross-Border Commuter Flows".
- National Meteorological Services: Provide weather data (temperature, humidity) that can influence virus survival.

Data Consumers:

- European Centre for Disease Prevention and Control (ECDC): The primary consumer. Ingests fused ACT data from multiple countries to model pathogen spread and issue early warnings.
- National Public Health Agencies: Monitor for imported risks from neighbouring countries.
- Research Institutions: Use the federated data for epidemiological studies.
- Travel App Providers: Receive filtered information to provide health risk alerts to travellers.

Enabling Technologies:

- ETSI ACT Standard (ETSI TS 103 757 [i.9]): Provides the interoperable method for data collection and messaging.
- oneM2M Suite (www.oneM2M.org): Provides the communication architecture and framework for the collection and the sharing of ACT data, and the ability to consume these data for the Data consumers, either natively or via interworking.
- DCAT-AP-based European Data Catalogue [i.7]: Serves as the foundational metadata layer for all data assets, describing their provenance, format, location, and semantics.
- STAT-DCAT-AP [i.5] / GeoDCAT-AP [i.3]: Critical extensions for geo-referencing, allowing consumers to find all data relevant to a specific geographic area (e.g. the Rhine watershed, the Lombardy region).
- HealthDCAT-AP [i.6], is the base for all the health information.
- AI/ML Analytics Engines: Process the fused, cross-border data for pattern recognition and predictive modelling.

B.2.3 The Role of DCAT-AP and its Extensions

DCAT-AP provides the standardized metadata framework that solves the core challenge of data findability across the independent data spaces of various national health services. It creates a federated catalogue where any authorized consumer can discover and understand available ACT datasets, making cross-border integration seamless.

Core DCAT-AP for Discovery:

- Each national ACT Control Service dataset (e.g. "French National ACT Alerts", "German Waste-Water Pathogen Monitoring") is registered as a dcat:Dataset.
- Standardized metadata (dct:title, dct:description, dct:publisher, dct:issued) ensures basic findability and understanding. The dct:spatial property would indicate national coverage.
- The dcat:accessURL points to the live API endpoint of the national ACT Control Service, from which the ECDC can fetch data.

Critical Extensions for Enhanced Semantics and Interoperability:

- STAT-DCAT-AP/GeoDCAT-AP: These are crucial for this use case. They allow the precise geo-referencing of each ACT detection event using the Geohash system (as mandated by the ACT standard). A consumer can query the catalogue for all datasets with dct:spatial overlapping "Border Region X," finding ACT data from both Country A and Country B.
- HealthDCAT-AP Essential for information such as Vaccinations, Intensive Care Units locations an occupancy, etc.

- Provenance & Quality Extensions: Future DCAT-AP extensions could include:
    - Data Freshness (xsd:duration): Indicates "updated every 6 hours," allowing the ECDC to find and prioritize the most timely data streams.
    - Accuracy/Uncertainty: Metadata describing the confidence level of the ACT test (e.g. test-result: 75 meaning a 75 % contamination level).
- Other Extensions: A dcat:theme like "Public Health" or "Pandemic Preparedness" would allow consumers to easily find all relevant health datasets.

The ECDC discovers necessary data by querying the European Data Portal [i.7] catalogue for datasets with specific properties (e.g. theme: "health", spatial: "AlpineRegion", temporalResolution: "< 24h"). The catalogue returns the endpoints for the relevant national ACT services, and the ECDC's system can automatically subscribe to these cross-border data streams.

B.2.4 Scenarios

Scenario A: Early Detection of a Cross-Border Health Threat

1) Trigger:
    - The ACT system in Country A detects a sharp increase in pathogen concentration in waste-water samples from a city near the border with Country B. This alert, with its Geohash location, is automatically published to the national ACT dataset, which is registered in European Data Portal [i.7].
2) Data Fusion & Discovery:
    - The ECDC's monitoring system periodically queries the European Data Portal [i.7] catalogue for new ACT alerts with a high TEST-RESULT value.
    - It discovers the new alert from Country A and, using the GeoDCAT-AP metadata, also automatically finds the ACT dataset from the neighbouring region in Country B.
    - It ingests this data and cross-references it with Eurostat's "Cross-Border Commuter Flow" data, also discovered via DCAT-AP.
3) AI Analysis & Action:
    - The AI model predicts a high probability of the pathogen spreading to the city in Country B within 3-5 days, based on commuter patterns.
    - The ECDC issues a pre-emptive early-warning alert to the public health agency of Country B.
    - The health agency in Country B increases wastewater testing frequency in the at-risk area and prepares local hospitals.
4) Result: A potential outbreak in Country B is mitigated or detected at the earliest possible stage, thanks to the findability and integration of cross-border data enabled by DCAT-AP.

Scenario B: Data-Driven Research for Pandemic Origins

1) Trigger:
    - A research institute begins a study on the environmental origins of a new virus variant.
2) Data Discovery:
    - A researcher uses the data.europa.eu [i.7] portal to search for all "Pathogen" related datasets. Using facets for geography (dct:spatial) and theme (dcat:theme), they quickly discover the ACT datasets from multiple European countries, as well as meteorological data.
3) Federated Analysis:
    - The researcher accesses all these datasets via their standardized dcat:accessURLs.

- By correlating historical ACT detection data with weather patterns, they identify specific environmental conditions that correlate with the emergence of the variant.

4) Result:

   - The research provides valuable insights for predictive modelling of future pandemic risks, a feat only possible through the findability and interoperability provided by the DCAT-AP catalogue.

B.2.5 Considerations

In this use case, the adoption of the structured semantic framework of DCAT-AP on a platform like European Data Portal [i.7] is the key to unlocking data findability across national borders. It moves the European health ecosystem from a collection of isolated, reactive national systems into a single, intelligent, and predictive organism:

- Cross-Border Data Findability: DCAT-AP provides a common language for describing data, making it possible to automatically discover and access relevant ACT datasets across different countries and languages. Without this, national ACT data remains locked in silos.
- Interoperability: oneM2M provides an interoperable data sharing between sensors and applications and across data lakes, providing an API/protocols interworking framework and supporting semantic interoperability. DCAT-AP then makes these interoperable data streams discoverable.
- Enhanced Early-Warning Systems: Faster detection of cross-border health threats saves lives, enabled by the system's ability to instantly find and integrate real-time environmental data from multiple sources.
- Proactive vs. Reactive Health Policy: Public health bodies can predict and prevent problems rather than just responding to them, by finding and querying international data in advance.
- Resource Efficiency: Dynamic use of existing data (from national ACT systems) creates a powerful supranational tool without the need for a duplicate, centralized data collection infrastructure.
- Futureproofing: The DCAT-AP framework ensures that as new countries deploy ACT systems or new data providers (e.g. private lab networks) emerge, their data can be immediately discovered and integrated into the European health data space.

B.3 Interoperability of Energy Smart Appliances with SAREF and Smart Grid Data

B.3.1 Overview

The European Code of Conduct for Energy Smart Appliances (CoC-ESA) creates a framework that stimulates manufacturers to bring interoperable devices (e.g. heat pumps, washing machines, EV chargers) to the EU market. These devices adjust their energy consumption to empower end-users with automated savings and participation in flexibility markets, while also supporting grid stability. Acting as intelligent coordinators, Energy Management Systems (EMSs) optimize flexibility from multiple appliances. They automatically schedule operation during the lowest electricity prices or highest renewable energy generation and respond to grid signals like congestion alerts from Distribution System Operators (DSOs). A core CoC-ESA requirement is the semantic mapping of device data and capabilities to the ETSI SAREF ontology and its energy flexibility extension, SAREF4ENER.

A DCAT-AP-based catalogue overcomes data silos across manufacturers by registering SAREF-mapped device datasets. It serves as a discovery layer, enabling Energy Management Systems and grid operators to find, understand, and integrate real-time flexibility from millions of household appliances. This effectively creates a virtual power plant for data-driven optimization of energy use, renewable integration, and grid resilience. Moreover, by making this catalogue accessible to end-users, it delivers on the promise of interoperability, enhancing transparency and building trust for those who have invested in smart appliances.

B.3.2 Actors

Data Providers:

- Energy Smart Appliance Manufacturers (CoC-ESA Signatories): Provide SAREF-mapped datasets describing their devices' capabilities, real-time status, and available flexibility (e.g. " Flexibility Offer", "Flexibility Profile", "Real-Time Appliance Status").
- Home/Building Energy Management Systems (CoC-ESA Signatories): Provide devices with SAREF compliant activation plans that contain optimized operational schedules based on flexibility profiles.
- Distribution System Operators (DSOs): Provide grid status data, including current load, congestion alerts, and local renewable energy generation forecasts.
- Energy Suppliers/Aggregators: Provide datasets on energy prices (e.g. dynamic tariff schedules) and aggregate flexibility requests.
- European Product Registry for Energy Labelling (EPREL): Provides the authoritative product database, which can be linked via DCAT-AP to the dynamic SAREF datasets.

Data Consumers:

- Home/Building Energy Management Systems (CoC-ESA Signatories): The primary consumer. Discovers and ingests flexibility signals from registered appliances to optimize local energy use and cost.
- Prosumers: individuals who own energy smart appliances and actively participate in the energy ecosystem by consuming, generating (for example, through renewable sources), and storing energy (such as in batteries). They also seek access to their own energy data.
- Distribution System Operators (DSOs): Consume aggregated flexibility data to manage local grid constraints and prevent congestion.
- Energy Aggregators: Discover available flexibility from device fleets to participate in energy markets.
- Smart Charging Service Providers: Integrate appliance data to optimize electric vehicle charging without overloading the household circuit.

Enabling Technologies:

- ETSI SAREF (Smart Applications REFerence ontology) and SARE4ENER (extension for energy flexibility): Provides the unified semantic framework for interoperability, enabling all devices to describe their capabilities and data in a standardized language that can be seamlessly exchanged with Energy Management Systems to optimize energy usage.
- European Code of Conduct for Energy Smart Appliances (CoC-ESA): Defines the policy and technical framework, mandating the use of SAREF for semantic interoperability.
- DCAT-AP-based Energy Data Catalogue: Serves as the foundational metadata layer for all data assets, describing the SAREF-mapped datasets, their provenance, and access endpoints.
- AI/ML Optimization Engines: Process the integrated data to calculate optimal flexibility schedules and grid-balancing actions (e.g. Energy Management Systems).

B.3.3 The Role of DCAT-AP and its Extensions

DCAT-AP provides the standardized metadata framework that solves the core challenge of data findability across the independent data spaces of various appliance manufacturers, grid actors and end-users. It creates a federated catalogue where any authorized energy management system can discover and understand available flexibility resources, making integration seamless.

Core DCAT-AP for Discovery:

- Each manufacturer's dataset (e.g. "Brand X Heat Pump Flexibility Profile," "Brand Y Smart Appliance Status") is registered as a dcat:Dataset.
- Standardized metadata includes dct:title, dct:description, dct:publisher, and dct:issued to ensure basic findability.
- The dcat:accessURL provides the endpoint for the HEMS to fetch the live or forecasted device data.

Core SAREF role for semantic findability and interoperability:

- SAREF Integration: This is the crucial semantic role. DCAT-AP records would link to the SAREF ontology (dct:conformsTo), indicating that the dataset's content is semantically described using SAREF concepts (e.g. saref:Device, saref:Power, s4ener:FlexibilityProfile).

DCAT-AP Extensions for Enhanced Semantics:

STAT-DCAT-AP [i.5] / GeoDCAT-AP [i.3]: Important for grid management. They specify the spatial coverage (dct:spatial) of device fleets (e.g. a neighbourhood, a city), enabling DSOs to find all flexibility resources relevant to a congested grid area.

- **Provenance & Quality Extensions:** Future extensions could include:
  - **Data Freshness (**xsd:duration**):** Indicates "updated every 5 seconds" for real-time status vs. "updated daily" for static capability data.
  - **Reliability/Accuracy:** Metadata describing the confidence level of a flexibility forecast.
  - **Licensing & Usage Rights:** Clearly defines how device data can be used (e.g. for grid balancing only, not for marketing).

B.3.4 Scenarios

Scenario A: Cost-Optimized Appliance Scheduling by an Energy Management System

1) Trigger:
   - A homeowner loads laundry into a CoC-ESA compliant washing machine, selects "Eco & Cheap" mode, and sets the latest desired finish time. She also connects her electric vehicle to a CoC-ESA compliant charger, specifying the latest time by which the car should be fully charged.
2) Data Discovery:
   - The Energy Management System (EMS) identifies the capabilities of both the washing machine and the EV charger through the DCAT-AP catalogue, recognizing them as SAREF-compliant devices capable of deferring their start times.
   - Additionally, the EMS locates the "Dynamic Electricity Prices" dataset provided by the energy supplier and optimizes appliance operation based on real-time pricing information.
3) Optimization & Action:
   - The EMS determines the optimal start times for both the wash cycle and EV charging, scheduling them to coincide with periods of lowest electricity prices, often when renewable energy generation is highest, or when home battery storage is available.
   - The EMS sends activation plans to the washing machine and EV charging station, providing specific instructions for when each appliance should begin its operation.
4) Result:
   - Both the laundry and EV charging are completed at minimal cost, fully aligned with the end-user's preferences and requirements. At the same time, the household's energy consumption is intelligently shifted to support broader grid stability.

Scenario B: Dynamic Response to a Grid Congestion Alert

1) Trigger:

   - A Distribution System Operator (DSO) detects a potential overload on a local transformer due to high air conditioning use on a hot afternoon. It publishes a "Congestion Alert" dataset to the DCAT-AP catalogue.

2) Data Fusion & Discovery:

   - An Aggregator's system discovers the DSO's alert via DCAT-AP, querying by location and theme.
   - Simultaneously, it queries the catalogue for all SAREF-mapped "HVAC Flexibility" datasets within the affected geographic area.
   - It discovers datasets from multiple manufacturers' heat pump and air conditioner fleets.

3) AI Analysis & Action:

   - The Aggregator's AI calculates the required power reduction and purchases the flexibility from the available device fleets.
   - Signals are sent to the participating HEMS units, which slightly and temporarily adjust the thermostat setpoints for thousands of compliant air conditioners.

4) Result:

   - The grid congestion is alleviated without any consumer noticing a significant comfort loss, preventing a blackout.

B.3.5 Considerations

In this use case, the adoption of the structured semantic framework of DCAT-AP is the key to unlocking the value of the CoC-ESA's SAREF mandate. It moves the energy system from a collection of siloed, "dumb" appliances into a dynamic, intelligent, and responsive network:

- Device and Data Findability: DCAT-AP provides a common language for describing SAREF-mapped data, making it possible for a HEMS to automatically discover and integrate all compatible appliances in a home, regardless of brand. Without this, interoperability remains a manual, point-to-point challenge.
- Grid Stability and Resilience: The ability to discover and aggregate flexibility from millions of small devices provides a powerful, cost-effective tool for balancing the grid, especially with high volumes of intermittent renewable energy.
- Consumer Empowerment and Savings: Automated discovery and optimization enable consumers to reduce their energy bills effortlessly by allowing their systems to respond to market and grid signals.
- Interoperability by Design: SAREF provides the semantic interoperability, while DCAT-AP provides the discovery mechanism. Together, they ensure that interoperability is not just a technical feature but an operational reality.
- Futureproofing and Scalability: The DCAT-AP framework ensures that as new manufacturers sign the CoC-ESA or new device types (e.g. bi-directional EV chargers) emerge, their data can be immediately discovered and integrated into the energy data space, scaling the system effortlessly.

B.4 Mapping DCAT-AP to NGSI-LD in Environmental Monitoring

B.4.1 Overview

This use case exploits the combined use of the Next Generation Service Interface - Linked Data (NGSI-LD) and the Data Catalogue Vocabulary - Application Profile for Europe (DCAT-AP) in an environmental monitoring use case. The goal is to ensure that the metadata describing datasets and data services in the NGSI-LD-based data space remain interoperable while supporting real-time context information exchange. In this way catalogue metadata expressed with DCAT-AP can be made accessible through NGSI-LD APIs and, conversely, dynamic resources managed by NGSI-LD brokers can be surfaced as DCAT-AP compliant catalogue entries. More detailed information can be obtained from ETSI TR 104 244 [i.12].

NGSI-LD focuses on context information management, delivering a resource-oriented API with native support for geospatial queries, temporal queries and publish-subscribe mechanisms. By weaving together both specifications a data space operator can present a unified catalogue in which static files, live data streams and complex entity collections sit side by side, each described with DCAT-AP semantics yet accessible through NGSI-LD operations.

B.4.2 Description of the Use Case

The use case describes an environmental monitoring network that collects water-quality and ecosystem data for a coastal area and exposes it through an NGSI-LD platform, with metadata published to an open data portal using DCAT-AP mappings. Sensors and models provide parameters such as nutrients, dissolved oxygen, chlorophyll, and temperature for a defined marine or lagoon zone, complemented by additional analytical and forecast datasets. The NGSI-LD broker stores both the observation entities and DCAT-AP-style catalogue, dataset, distribution, and data service entities so that an open data portal (for example, CKAN-based) can harvest DCAT-AP metadata and offer standardized discovery and access.

Actors

- **Organization A:** Manages the NGSI-LD broker, the data collection processes, and implements metadata transformation.
- **Organization B:** Administers the open data portal and publishes DCAT-AP-compliant catalogues.
- **Generic Users:** Various stakeholders (policy makers, technical researchers, commercial entities, general public) who discover and access datasets.

B.4.3 Scenarios

Scenario 1: Publishing Monitored Environmental Data

Continuous measurements (physical parameters such as temperature and salinity; chemical parameters such as dissolved oxygen and nutrients; biological parameters such as chlorophyll and species abundance) are collected from a distributed network of automated instruments, periodic field campaigns, local information systems, users' applications.

Each measurement is represented as an NGSI-LD Entity, containing temporal timestamps, geographic coordinates, and parameter values.

At the catalogue level, an NGSI-LD Dataset Entity is created with DCAT-AP-aligned attributes: multilingual title and description, spatial coverage polygon, temporal coverage, and thematic classification. A Distribution Entity describes how the data can be accessed: direct NGSI-LD API queries, downloadable file exports in standard formats (JSON, CSV), and media types. A Data Service Entity defines the operational characteristics of the live API endpoint.

An automated data integration pipeline, which uses specific connectors to detect Catalog, Dataset and Distribution Entities without focusing on the continuous temporal measurement for performance reasons, works in the NGSI-LD broker to transform them into the open data portal's internal structure. The portal exposes the catalogue through both human-readable web interfaces and machine-readable DCAT-AP RDF exports, enabling cross-border discovery and federation with other European data initiatives.

This approach maximizes data reusability through standardized semantics; provides both real-time and static access modalities; ensures regulatory compliance with open data directives; supports automated harvesting by downstream portals that can precisely process the long temporal sequences on their own, based on knowledge contained in the metadata.

Scenario 2: Integrating External Reference Datasets

Beyond locally collected data, the monitoring initiative identifies external authoritative datasets that provide essential context but are managed by other organizations: satellite-derived products, regional model outputs, historical reference data, or specialized laboratory results.

For each external source, an NGSI-LD Catalog Entity and an NGSI Dataset Entity are created with appropriate DCAT-AP metadata (title, abstract, keywords, spatial bounds, temporal extent, update frequency, licensing terms). Distribution Entities are configured to reference the external source's access points (landing page URL, API endpoint, file download link) without duplicating the actual data.

These external Dataset Entities are linked to a Catalogue structure as locally produced datasets, creating a unified logical view. The automated pipeline publishes metadata for both internal and external datasets to the open data portal with consistent formatting, ensuring users can discover the complete ecosystem of available information.

Users browsing the portal see an integrated catalogue where both local observations and external reference data are indexed using common filters (theme, geographic region, temporal period). When accessing external datasets, users are directed to the authoritative source through documented access URLs while benefiting from harmonised DCAT-AP metadata and cross-referenced relationships.

This aims at reducing data fragmentation across provider boundaries; maintains data sovereignty by keeping external data at authoritative sources; enables federated discovery across heterogeneous data ecosystems; supports integrated analysis workflows that combine multiple data sources.

B.4.4 Strategies

A catalogue that adheres to DCAT-AP may be exposed through one of two complementary strategies. The first strategy treats the catalogue as a read-only knowledge base whose resources are serialized as NGSI-LD Entities. In this configuration the NGSI-LD broker does not store the datasets themselves; it simply offers an alternative API that mirrors the SPARQL or CSW endpoint traditionally used by catalogue portals. Implementers will generate one NGSI-LD Entity of type Catalogue for each dcat:Catalog, one Entity of type Dataset for every *dcat:Dataset*, and so forth, thereby preserving the original URI identifiers and the semantic relationships that connect them. Because NGSI-LD Entities support the *observedAt* temporal attribute, catalogue maintainers can attach timestamps to every property, enabling consumers to query for records updated after a given date without re-harvesting the whole catalogue. Geospatial coverage, when present, is encoded as GeoProperty values so that geographical filters may be executed server-side through the *georel* operator.

The second strategy extends the catalogue paradigm to the live domain. Here the NGSI-LD broker acts as the system of record for the underlying data objects, for example traffic sensors emitting hourly air-quality measurements or energy meters recording consumption data. Each Entity or Entity collection held by the broker is linked to a DCAT-AP dataset in the catalogue. The Distribution entities then point back to the broker by providing an *accessURL* that resolves to an NGSI-LD query endpoint, while the *dataService* element references an NGSI-LD API description with the *conformsTo* link set to the NGSI-LD specification URI. Temporal properties in NGSI-LD are used to calculate the *dcat:temporalCoverage* interval, thereby signalling to catalogue users that the dataset grows over time and is best accessed through incremental queries rather than static downloads.

The following clause introduces guidelines and examples for implementing DCAT-AP using NGSI-LD. The objective is to be able to describe the metadata related to NGSI-LD information available through the NGSI-LD Context Broker using the conceptual classes defined in the DCAT-AP model.

B.4.5 Examples

This clause presents practical examples of how DCAT-AP catalogue metadata is encoded as NGSI-LD entities in the environmental monitoring use case. The listings demonstrate core DCAT-AP classes - Catalog, Dataset, and Distribution - following the NGSI-LD information model. Each example illustrates how catalogue properties such as title, description, spatial and temporal coverage are mapped to NGSI-LD Property, Relationship, or GeoProperty attributes, while preserving semantic alignment with the DCAT-AP vocabulary through appropriate @context declarations.

The listing B.4.5.1 shows a Catalog Entity representing a marine satellite data service. It includes descriptive properties (title, description, homepage) expressed as NGSI-LD Property attributes with simple string values.

The listing B.4.5.2 demonstrates a Dataset Entity describing a biogeochemical analysis and forecast product for a Mediterranean marine region. Key features include:

- Thematic classification through Relationship attributes linking to controlled vocabularies
- Spatial coverage expressed as a GeoProperty with a Polygon geometry
- Temporal coverage captured in a JsonProperty containing start date, end date, and status
- Link to the Distribution (the technical access)
- Logical link to the actual data entities

Listing B.4.5.3 presents a Distribution entity that specifies how the dataset can be accessed. Distribution entities typically include access URLs, download URLs, format specifications, and service endpoints, enabling data consumers to retrieve or query the underlying environmental observations.

All examples reference a shared @context combining the DCAT-AP JSON-LD [i.13] context and the NGSI-LD core context, ensuring that property names are correctly resolved to their semantic definitions in both specifications. This dual-context approach enables NGSI-LD brokers to serve catalogue metadata that remains compatible with DCAT-AP harvesting tools and SPARQL endpoints used by open data portals.

Listing B.4.5.1: Snippet of the NGSI-LD Catalog Entity

```
    {
        "id": "urn:ngsi-ld:Catalog:CopernicusMarine",
        "type": "Catalog",
        "title": {
            "type": "LanguageProperty",
            "languageMap": {
                "en": "Copernicus Marine Service",
                "fr": "Service Marine Copernicus"
            }
        },
        "homepage": {
            "type": "Property",
            "value": "https://marine.copernicus.eu/"
        },
        "description": {
            "type": "LanguageProperty",
            "languageMap": {
                "fr": "Fournit des informations de référence gratuites, ouvertes, régulières et 
systématiques sur l'état, la variabilité et la dynamique de l'océan bleu (physique), blanc (glace 
de mer) et vert (biogéochimique) à travers l'océan mondial et les mers régionales européennes.",
                "en": " Provides free, open, regular and systematic reference information on the 
state, variability and dynamics of the blue ocean (physics), white ocean (sea ice) and green ocean 
(biogeochemistry) across the global ocean and European regional seas."
            }
        },
        "publisher": {
            "type": "Property",
            "value": "Copernicus Marine Service"
        },
        "language": {
            "type": "Property",
            "value": ["en", "it"]
        },
  // THE GLUE: Linking to the Datasets
```

```
  "dataset": {
    "type": "Relationship",
    "object": [
      "urn:ngsi-ld:Dataset:Med-SST-Analysis-2024",
      "urn:ngsi-ld:Dataset:Med-Chlorophyll-2024"
    ]
  },
        "@context": [
            "https://example.org/jsonld-contexts/dcat-compound.jsonld",
            "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.9.jsonld"
        ]
    }
```

Listing B.4.5.2: Snippet of the NGSI-LD Dataset Entity

```
    {
        "id": "urn:ngsi-ld:Dataset: Med-SST-Analysis-2024",
        "type": "Dataset",
        "title": {
            "type": "LanguageProperty",
            "fr": "Mer Méditerranée - Biogéochimie - Analyses et prévisions",
            "en": " Mediterranean Sea - Biogeochemistry - Analysis and Forecasts"
        },
        "source": {
            "type": "Property",
            "value": "https://marine.copernicus.eu/"
        },
        "description": {
            "type": "Property",
            "value": "Données biogéochimiques en mer Méditerranée, résolution 4km"
        },
        "dataProvider": {
            "type": "Property",
            "value": "Copernicus Marine Service (CMEMS)"
        },
        "license": {
            "type": "Property",
            "value": "CC-BY-4.0"
        },
        "theme": [
            {
                "type": "Relationship",
                "datasetId": "urn:ngsi-ld:Dataset:Theme:WayterNutrients",
                "object": "urn:ngsi-ld:Theme:WaterNutrients"
            },
            {
                "type": "Relationship",
                "datasetId": "urn:ngsi-ld:Dataset:Theme:EnvironmentalRisks",
                "object": "urn:ngsi-ld:Theme:EnvironmentalRisks"
            }
        ],
        "spatial": {
            "type": "GeoProperty",
            "value": {
                "type": "Polygon",
                "coordinates": [
                    [
                        [
                            3.5772046,
                            43.278991
                        ],
                        [
                            3.8483749,
                            43.4466945
                        ],
                        [
                            3.902072,
                            43.457236
                        ],
                        [
                            3.5772046,
                            43.278991
                        ]
                    ]
                ]
            }
```

```
        },
        "temporal": {
            "type": "JsonProperty",
            "json": {
                "status": "active",
                "endDate": "2025-12-04T00:00:00Z",
                "startDate": "2024-01-16T00:00:00Z"
            }
        },
// Link to the Distribution (the technical access)
  "distribution": {
    "type": "Relationship",
    "object": [
      "urn:ngsi-ld:Distribution:Med-SST-NGSI-API"
    ]
  },

        "@context": [
            "https://example.org/jsonld-contexts/dcat-compound.jsonld",
            "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.9.jsonld"
        ]
    }
```

Listing B.4.5.3: Snippet of NGSI-LD Distribution Entity

```
    {
  "id": "urn:ngsi-ld:Distribution:Med-SST-NGSI-API",
  "type": "Distribution",
  "accessURL": {
    "type": "Property",
    "value": "https://my-broker.example.org/ngsi-ld/v1/entities?type=SeaSurfaceTemperature"
  },
  "format": {
    "type": "Property",
    "value": "NGSI-LD"
  },
  "description": {
    "type": "Property",
    "value": "Direct NGSI-LD API access to the Sea Surface Temperature entities."
  },
  // Inverse link to the Dataset (optional but useful for bidirectional navigation)
  "servesDataset": {
    "type": "Relationship",
    "object": "urn:ngsi-ld:Dataset:Med-SST-Analysis-2024"
  },
        "@context": [
            "https://example.org/jsonld-contexts/dcat-compound.jsonld",
            "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.9.jsonld"
        ]
    }
```

In the last lines of the examples the link to the @context https://example.org/jsonld-contexts/dcat-compound.jsonld is essential. It maps NGSI-LD terms (like accessURL or Dataset) to the official DCAT-AP RDF URIs (dcat:accessURL, dcat:Dataset). With this context, a DCAT harvester (software that collects open data) can technically read your Context Broker and understand that it is acting as a Data Catalog.

B.5 Semantic Adaptation and Automated Federation of High-Value Datasets (HVD)

B.5.1 Overview

The Commission Implementing Regulation (EU) 2023/138 (HVD Implementing Act) [i.16] defines specific technical and legal obligations for datasets with high socio-economic potential. Catalogues allow users and computer programs to search for datasets based on their metadata records. To ensure that the dataset metadata are discoverable and compliant across the Union, the generic DCAT-AP model needs to be adapted into a specialized profile. This use case illustrates how this semantic adaptation enables the automated federation of metadata from local providers to the European Data Portal (data.europa.eu), ensuring that the "High-Value" status and its associated legal requirements are preserved throughout the entire harvesting chain.

B.5.2 Actors

Data Providers:

- Public sector bodies and public undertakings: Those responsible for metadata who perform the initial semantic adaptation by tagging their datasets according to the HVD requirements.
- National Data Portals: The national nodes that act as aggregators, harvesting metadata from local sources and validating its compliance before further federation.

Data Consumers:

- European Data Portal (data.europa.eu): The central federator that consumes national metadata and exposes it to the public, providing a unified view of all European HVD.
- Information Re-users, Citizens, and Public Administrations: Stakeholders who consume aggregated metadata to create transnational cross-border services or conduct research, relying on the metadata to guarantee the technical and legal quality of the data products.
- Policy Makers and Regulators: Use the federated metadata to monitor the implementation of the Open Data Directive across Member States.

B.5.3 Enabling Technologies

- Metadata Harvesting Protocols (e.g. OAI-PMH, SPARQL, POST, GET): The technical "pipelines" that allow for the automated movement of metadata between different administrative levels.
- Shapes Constraint Language (SHACL) [i.15]: The validation technology used to enforce the HVD adaptation, ensuring that every federated record contains the mandatory legal tags.
- Linked Data URI Strategy: Use of persistent URIs to reference the HVD Implementing Act and thematic categories.
- Catalogue Service Web (CSW): Defines common interfaces to discover, browse, and query metadata about data, services, and other potential resources. This location service is used to collect metadata from datasets classified as INSPIRE-like HVD.

B.5.4 The Role of DCAT-AP Adaptation

In this scenario, DCAT-AP is not just a descriptive tool but a regulatory enforcement mechanism. The "Adaptation" involves profiling the standard DCAT-AP to include mandatory legal and technical attributes:

- Legal Referencing: The adaptation mandates the use of the dcatap:applicableLegislation property to point specifically to the HVD Regulation URI.

- Thematic Alignment: It restricts the dcat:theme property to the specific HVD categories (e.g. Geospatial, Statistics, Mobility) using European controlled vocabularies.
- Technical Compliance: It enforces the description of APIs and bulk downloads through dcat:DataService and dcat:Distribution, ensuring the technical "machine-readability" required by law is explicitly documented.

For the categories of geospatial, earth observation and environment, and mobility, additional one can use the extension GeoDCAT-AP. It is an extension of DCAT-AP for describing geospatial datasets, dataset series, and services. It provides an RDF syntax binding for the union of metadata elements defined in Regulation (EC) No 1205/2008.

B.5.5 Scenario: Automated Compliance through Metadata Federation

- Step 1 - Adaptation at Source: A regional statistical office publishes a census dataset. Using the adapted DCAT-AP profile, they include the mandatory HVD legislative tag and describe the access API.
- Step 2 - National Harvesting: The National Data Portal's harvester visits the regional catalogue. It automatically recognizes the HVD tag and "pulls" the metadata into the national repository.
- Step 3 - European Federation: data.europa.eu harvesters collect the metadata from the National Portal. Because the HVD adaptation is consistent, the European portal automatically promotes the dataset to the "High-Value" global collection.
- Result: The dataset is discoverable as an HVD at the European level without any manual reporting or data re-entry by the original provider.

B.5.6 Considerations

This use case demonstrates that the adaptation of DCAT-AP is the cornerstone of the European data federation strategy. The primary considerations are:

- From Manual to Automated Reporting: Compliance with European Directives is achieved as a "by-product" of good metadata management, eliminating the need for separate administrative reports.
- Semantic Consistency: The federation process ensures that the legal status (HVD) assigned at the local level remains trustworthy and verifiable at the European level.
- Regulatory Certainty: By adapting DCAT-AP with specific technical requirements (APIs, formats), the metadata provides re-users with an immediate guarantee of the dataset's quality and terms of use.

Annex C (informative): Change history

Date	Version	Information about changes
September 2025	V0.0.0	First initial container with initial draft introduction and scope clauses.
September 2025	V0.1.0	Built with a structure similar to EN 303 760. Added clauses for: • Getting started • Use and instantiation of DCAT-AP metadata • Extension of DCAT-AP • Contribution to the SEMIC DCAT-AP suite of profiles Guide to the management of this EN and its development and evolution
September 2025	V0.1.1	Few fixings, addition of initial headed for use cases annex, initial identification of use cases
October 2025	V0.2.0	Addition of a use case for cognitive traffic management with Cellular sensing and Smart City data
October 2025	V0.3.0	Restructured format end improve content to consider the SEMIC suggestions in DATA(25)000128 Addition of a use case in DATA(25)000127 on interoperability of energy smart appliances. Editorials and improvements suggested in DATA(25)000130 Revision marks are expressed against version_0.1.1
October 2025	V0.3.1	Editorial fixings Revision of the use case for cognitive traffic management with Cellular sensing and Smart City data Initial incomplete text for clause 4.2
October 2025	V0.4.0	Introduction of a new use case on Cross-Border Pandemic Preparedness using ETSI ACT standard and data.europa.eu and revision of the energy use case on Smart Appliances and Smart Grid (DATA(25)000147) New annex providing examples of DCAT-AP extensions from DATA(25)000143 Improvement of introduction, initial content for terms and acronyms, text updates from (DATA(25)000141r1)
November 2025	V0.5.0	Full revision (draft) based on DATA(25)000157. Inclusion of comments in DATA(25)000153 respect to clauses 1-3 (the other are superseded by DATA(25)000157. Editorial upgrades
November 2025	V0.6.0 V0.6.1	Editorial upgrades (including DATA(25)000176. Revision of the use case on interoperability of energy smart appliances DATA(25)000182 Inclusion of the mapping case of NGSI-LD in Environmental Monitoring in DATA(25)000183
November 2025	V0.6.2	Inclusion of the revision of mapping case of NGSI-LD in Environmental Monitoring in DATA(25)000183r1 Editorial revisions
December 2025	V0.6.3	Revision of mapping case of NGSI-LD in Environmental Monitoring in DATA(25)000190r1 Introduction of the data space rulebook concept Revision and integration of the test section 5 (now compliance) Editorial corrections
February 2026	V0.6.4	Implementation (only accepted comments) of Data(26)000001r1, DATA(25)000198, DATA(26)000025 (late), DATA(26)000023 (late), emails comments (late). Editorial corrections. To be checked at the meeting of 6th February 2026 against comments
February 2026	V0.6.5	Merging with SEMIC/EC feedback. RM version showing resulting changes with respect to 0.6.3. Editorial corrections. To be checked at the meeting of 6th February 2026 against comments
February 2026	V0.6.6	Endorsed at the meeting of 6th February 2026 against comments to be sent to Public Enquiry and National Vote

History

Version	Date	Status	
V1.0.0	March 2026	SRdAP process	EV 20260615: 2026-03-17 to 2026-06-15