Deliverable
D7.6 Standardisation Plan and Report. Initial

Project Acronym:	**URBANAGE**	
Project title:	Enhanced URBAN planning for AGE-friendly cities through disruptive technologies	
Grant Agreement No.	101004590	
Website:	www.**URBANAGE**.eu	
Version:	1.1	
Date:	29/07/2021	
Responsible Partner:	AIV	
Contributing Partners:	TEC, IMEC, UH, ENG, ATC, CIC	
Reviewers:	Giuseppe Ciulla (ENG) Julio García (CIC)	
Dissemination Level:	Public	x
	Confidential – only consortium members and European Commission	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004590

Revision History

Revision	Date	Author	Organization	Description
0.1	01/06/2021	Lieven Raes	AIV	Initial version
0.2	16/06/2021	Lieven Raes	AIV	Intro, chapter 3, structure improvements
0.3	28/06/2021	Lieven Raes	AIV	Chapter 4 structure, chapter 5
0.4	05/07/2021	Bert Van Nuffelen, Dwight Van Lancker, Philippe Michiels	AIV, IMEC	Chapter 3 and 4 additions
0.5	07/07/2021	Marina Klitsi, Thanasis Dalianis, Sergio Campos Cordobes, Roberto Di Bernardo	ATC, TEC, ENG	Chapter 4 and 5 additions
0.6	12/07/2021	Lieven Raes	AIV	Introduction and conclusions
0.7	13/07/2021	Lieven Raes, Julio García	AIV, CIC	Lay-out and check
0.8	15/07/2021	Julio Garcia, Giuseppe Ciulla	CIC, ENG	Review
1.0	20/07/2021	Lieven Raes, Dwight Van Lancker, Jurgen Silence	AIV	Final edits
1.1	29/07/2021	Jurgen Silence, Giuseppe Ciulla	AIV, ENG	Final version

Table of Contents

Table of Figures

Table of Tables

List of abbreviations

Abbreviation	Explanation
AI	Artificial Intelligence
AIML	Artificial Intelligence Markup Language
AIS	Automatic Identification System
BDGMM	Big Data Governance and Metadata Management
BDVA	Big Data Value Association
BIM	Building Information Model
BSP	Bulk Synchronous Parallel model
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
CKAN	Comprehensive Knowledge Archive Network
CSA	Communication Support Action
DAIRO	Data, AI and Robotics
DCAT	Data Catalog Vocabulary
DCAT-AP	Data Catalog Vocabulary Application Profile
DCMI	Dublin Core Metadata Initiative
DTC	Digital Twin Consortium
DXWG	W3C Data Exchange Working Group
E-GOV	Electronic Government
EIF	European Interoperability Framework
ENISA	European Union Agency for Cybersecurity
EO	Executive Order
ESO	European Standards Organization
ETSI	European Telecommunications Standards Institute
FPGA	Field Programmable Gate Arrays

FRBR	Functional Requirements for Bibliographic Records
GIS	Geospatial Information System
GPU	Graphics Processing Unit
HDFS	Hadoop Distributed File System
HPC	High-Performance Computing
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
IDS	International Dataspaces
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
INSPIRE	INfrastructure for SPatial InfoRmation in Europe
IoT	Internet Of Things
ISA²	Interoperability solutions for public administrations, businesses and citizens
ISO	International Organisation for Standardisation
KG	Knowledge Graphs
LDES	Linked Data Event Stream
MIM	Minimal Interoperability Mechanisms
MQTT	Message Queuing Telemetry Transport Message Queuing Telemetry Transport
NAS	Network Attached Storage
NBDIF	Big Data Interoperability Framework
NBD-PWG	Big Data Public Working Group
NBD-RA	NIST Big Data Reference Architecture
NGSI	Next Generation Service Interfaces
NISO	National Information Standards Organisation
NIST	National Institute of Standards and Technology (US)
NNEF	Neural Network Exchange Format
O&M	Observation & Measurements

OASC	Open Agile Smart Cities
OASIS	Organization for the Advancement of Structured Information Standards
OAuth	Open Authorization
OGC	Open Geospatial Consortium
OIDC	OpenID Connect
ONNX	Open Neural Network Exchange
openMI	open Modelling Interface
OSLO	Open Standards of Linked Organisations
OWASP	Open Web Application Security Project
OWL	Web Ontology Language
PFA	Portable Format for Analytics
PMML	Predictive Model Markup Language
PSI	Public Sector Information (directive)
RDF	Resource Description Framework
SAML	Security Assertion Markup Language
SAREF	Smart Appliances REFerence ontology
SDO	Standardisation Organisation
SKOS	Simple Knowledge Organization System
SLA	Service Level Agreement
SOAP	Simple Object Access Protocol
SOS	Sensor Observation Service
SOSA	Sensor Observation Sample and Actuator
SPARQL	SPARQL Protocol And RDF Query Language
SSN	Semantic Sensor Network
SSO	Single sign-on
TD	Thing Description
TSDB	Time Series Database

UML	Unified Modeling Language
W3C	World Wide Web Consortium
WFS	Web Feature Service
WMS	Web Map Service
WoT	Web Of Things
WSDL	Web Service Description Language
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformations

1 Executive Summary

There are many reasons to use standards for building complex ICT solutions like Digital Twins. Many of these reasons lie in robustness, interoperability and transferability. In some well-proven domains, it is obvious not to reinvent the wheel, and some standards are considered no brainers (e.g. HTML or WMS, WFS for geospatial representation).

On the one hand, with standards, we mean core IT standards like DCAT, WMS, CityGML. These standards are subject independent. On the other hand, data standardisation can be more considered as tasks related to giving meaning to data to realise semantic interoperability. Examples are the semantic web standards, SAREF, NGSI-LD and also Inspire data specifications. This report covers both.

Chapter 3 focuses on those standards that are no brainers and needs to be well considered to build a state of the art solution usable in different circumstances of the three pilots and beyond. However, the real motivation for using standards is that the **URBANAGE** consortium believes that innovation also lies in how disruptive technologies can only cooperate successfully when they cooperate in a well-designed way, implying the re-use of standards.

The **URBANAGE** approach is to start supply driven by studying relevant standardisation bodies in web, geospatial, smart city, IoT, Big Data and data-infrastructure fields. Starting from generic worldwide active organisations like W3C, ISO and OGC, we also looked ad more EU oriented standardisation organisations and standard related organisations. The latter is less focused on the standard development itself but on promoting end-user oriented and solution-related best practices. Also, the EU itself has a tradition of supporting standardisation organisations. Some of them, like ISA², CEN and ETSI, have a close relationship with the EU and offer interesting standards worth scrutinising in the Digital Twin context. Starting from these organisations supply, relevant standards were investigated in the relevant domains of the **URBANAGE** architecture. In the IoT domain, Linked Data Event Streams looks very promising. Metadata is a second category with well-established standards in the open and geospatial fields. The standards by ISO², W3C and Inspire (EC) are quite well integrated and well established.

In the AI field, standardisation initiatives are under development. It is less clear how and which standards will be applied. The use of AI in the project pilots is not yet well known at writing this document. The Big Data related architecture and standards will be based on the promising NBDIF interoperability framework. The proposed security standards like SAML and OAuth2.0, and OpenID are quite generic standards for identification (authentication and authorisation). For data visualisation and presentation, including ergonomics and human-system interaction, the ISO 9241 standards will be applied. The next step of the methodology is finding out if some of the used crucial standards need any adaptation to make them better usable in **URBANAGE** and the standardisation community. Depending on the standardisation body, membership is required to be able to contribute. A membership inventory has been made. Half of the consortium partner, including most technical and pilot partners, is a member of at least one relevant organisation. Only a few partners have experience with the standardisation work itself. Based on the experience of participating in standardisation initiatives of some of the SDO's it became clear that an end to end standardisation process often exceeds the project duration. For that reason, a long-term commitment of the partner participating in a (new) standardisation initiative is necessary. In that context, **URBANAGE** will focus on the core standards itself and more data and ontology related standards (e.g. NGSI-LD) and the implementation of concepts like the OASC MIM's (Minimal Interoperability Mechanisms).

During the next two years between this initial report and the final report, standards will be part of the strategic technical and functional design discussions and analysis. D7.7, the final standardisation plan and report will reflect these discussions in terms of approach, methodology and activities.

2 Introduction

The main goal of this deliverable is to present an overview of the different relevant standards for realising an interoperable Digital City Twin Solution to link the technical work developed in the project to relevant standardisation work of the concerned technical committees. The objective is to ensure that **URBANAGE** is on the right page with regards to future standardisation. The document will be used to find out where contributing to the standardisation community is helpful for **URBANAGE** and the SDO's.

In chapter 3, the relevant standardisation organisations, including their families of standards and norms, are presented. Next to the relevant well-established SDO's and relevant sector organisations contributing to norms and standards are taken into account.

Chapter 4, dives deeper into the standards themselves and what these standards can mean for the **URBANAGE** Digital City Twin solutions.

Chapter 5, contains the first version of the standardisation plan. It gives a first overview based on the current business, functional and technical insights on what standards will be used, what standards might be used and where **URBANAGE** can cooperate with existing SDO working groups and standardisation committees.

Finally, in chapter 6, we will draw the first set of conclusions, including the needed steps during the **URBANAGE** project, and an overview of open questions we will scrutinise to prepare the D7.7 Standardisation Plan and Report (final version).

3 Relevant Digital Twin standardisation and standardisation related organisations

3.1 W3C – World Wide Web Consortium

The W3C (World Wide Web Consortium) is an international community led by Tim Berners Lee. W3C's mission is to lead the web to its full potential. The W3C has different families of standards overarching many aspects of today essential web technology.

- Web Design and Applications;
- Web of devices;
- Web architecture;
- Semantic Web;
- XML technology;
- Web of services;
- Browsers and Authoring tools.

It makes no sense to discuss in the frame of **URBANAGE** the W3C supported generic web technology. Nevertheless, four domains are of particular interest for the development of Digital Twin data technology.
XML technology is of major importance for data use and exchange in a Digital Twin, including XML, XML namespaces, XML schema, XSLT.

Closely related is the **web of services** standards use to allow data interchange and interoperability. Next to XML, it comprises technologies like HTTP, SOAP, WSDL, SPARQL, and others.

The **Semantic Web** is the third group of standards to support a "web of data". The goal is to enable a computer to do more useful work with data by linking the data in a meaningful way. This technology is very promising for digital Twins where different data sources have to communicate. Used technologies and standards are RDF, SPARQL, OWL and SKOS.

Web of devices (Web of Things). The use of web technology is devices, and the ability to link devices to the web (like sensors) anytime and anywhere is of major importance for a Digital Twin. Recent W3C recommendations are Web of Things (WoT) [1] Architecture, the Web of Things (WoT) and the Thing Description (TD).

A related domain of particular interest for a digital Twin and of specific importance is the metadata domain based on W3C standards as Dublin Core related and related to RDF and other semantic web concepts.

3.2 OGC – Open Geospatial Consortium

The Open Geospatial Consortium (OGC), an international voluntary consensus standards organisation, originated in 1994, encouraging the development and implementation of open standards for geospatial content and services, sensor web and Internet of Things, GIS data processing and data sharing. The OGC has, comparable to the W3C, several widely used standards in the geospatial domain. Examples of these commonly used open standards are WMS and WFS to show geospatial referenced maps and vector features on a map. In the Smart City domain, the standards in the fields mentioned below are of particular interest.

- 2D and 3D visualisation standards;
- IoT standards;
- Model interaction standards.

Just likes it makes no sense to discuss generic W3C web standards, the use of OGC GIS standards like WMS, WFS and even KML are obvious. **URBANAGE** focuses, as explained above, on three domains. In **the 2D and especially 3D domain**, CityGML [2] and probably IndoorGML [3] are of interest. The first one focuses on describing the public domain and outdoor world when indoorGML can be used to guide people when visiting large buildings like shopping malls, big train stations or airports**. IoT support** is a second domain where OGC standards like Sensor Model Language, Sensor Observation Service, Sensor Planning Service, Sensor Things, and Semantic Sensor Network (SSN) are valuable candidates. **Model interaction** standards like openMI enable the runtime exchange of data between process simulation models and between models and other modelling tools such as databases and analytical and visualisation applications. Its creation has been driven by the need to understand how processes interact and predict the likely outcomes of those interactions under given conditions.

3.3 ISO – International Organisation for Standardisation

The International Organisation for Standardisation (ISO) unites standards organisation of 165 member countries and is active in many fields. In the **geospatial domain**, closely related to Digital City Twins, the ISO 19115:2003 and 2014 standard on "Geographic Information **Metadata**" and ISO 19119:2005 and 2016 standard on "**Geographic Information Services**". Other useful ISO standards are ISO 19110:2005 and 2016 regarding "Geographic information for **feature cataloguing**" and ISO/TS 19139:2007 on "Geographic Information – Metadata – XML Schema Implementation". ISO 19150-1:2012 and ISO 19150-2:2015 "Rules for developing **ontologies** in the Web Ontology Language (OWL) makes the (modelling) link between UML and OWL.

3.4 ISA² - Interoperability solutions for Public Administrations, Businesses and Citizens

The ISA² programme by the European Union stands for "Interoperability solutions for public administrations, businesses and citizens". The programme supports the development of digital solutions that enable public administrations, businesses and citizens in Europe to benefit from interoperable cross-border and cross-sector public services. The programme is managed by the EC Interoperability unit of DG informatics DIGIT.D2. Since a Digital Urban Twin strongly relies on data interoperability, many ISA² actions are of particular interest.

The New European Interoperability Framework (EIF) gives guidance on setting up interoperable digital public services. The EIF recommendations and goals are closely related to the DCT goals, improving governance of their interoperability activities, establishing cross-organisational relationships, and streamlining processes supporting end-to-end digital services. The EIF categorises the data interoperability challenges into four main categories: legal, organisational, semantic, and technical interoperability. It also provides a set of principles that can be used to drive towards a more interoperable digital environment. Member states and regions are encouraged to realise their own EIF instantiations. In Flanders, the Open Standards of Linked Organisations (OSLO) [4] is an example to ensure greater coherence, better comprehensibility, and findability of data and services.

Another relevant ISA initiative is the DCAT Application Profile (DCAT-AP) for data portals in Europe. Its purpose is to give owners of geospatial metadata the possibility to achieve more visibility by providing an additional RDF syntax binding.

3.5 CEN – European Committee for Standardisation

The European Committee for Standardisation has three main goals: the promotion of the growth of the European Single Market economy and the economy of Europe as a whole in global trading, the promotion of the welfare of its citizens and its environment. This public standards organisation provides efficient infrastructure for developing, maintaining, and distributing coherent sets of standards and specifications for those interested in its use. It is also known as CEN, the abbreviation from the French 'le Comité Européen de Normalisation'. CEN is the officially recognised standardisation representative for sectors other than electrotechnical (CENELEC) and telecommunications (ETSI). Especially ETSI is relevant in the field of Smart Cities and Digital Twins (see FiWare/ETSI).

3.6 FiWare / ETSI – European Telecommunications Standards Institute

FiWare is a curated framework of open source platform components to accelerate the development of smart solutions in the domains of Smart Cities, Smart Energy, Smart Industry and Smart AgriFood. Fiware has an active ecosystem driven by the Fiware foundation. Fiware is a component-based solution around a central core context management system (context broker). Around the FIWARE Context Broker, a rich suite of complementary open source FIWARE Generic Enablers is available. These “enablers” are dealing with the following:

- Interfacing with the Internet of Things (IoT), Robots and third-party systems for capturing updates on context information and translating required actuations;
- Context Data/API management, publication, and monetization bringing support to usage control and the opportunity to publish and monetize part of managed context data;
- Processing, analysis, and visualization of context information implementing the expected smart behaviour of applications and/or assisting end-users in making smart decisions.

Fiware is closely related to the NGSI standard. FIWARE NGSI is the API exported by a FIWARE Context Broker, used to integrate platform components within a “Powered by FIWARE” platform and by applications to update or consume context information. FIWARE NGSI API specifications have evolved over time, initially matching NGSI-v2 specifications, now aligning with the ETSI NGSI-LD standard. The FIWARE Community plays an active role in the evolution of ETSI NGSI-LD specifications which were based on NGSIv2 and commits to deliver compatible open-source implementations of the specs.

3.7 OASC – Open Agile Smart Cities

Open Agile Smart Cities (OASC) is an international city network of local public administrations to support their digital transformation journey by building global market solutions, services and data based on the needs of cities and communities. To achieve the mission, OASC uses Minimal Interoperability Mechanisms (MIMs). MIMs are a set of practical capabilities based on open technical specifications that allow cities and communities to replicate and scale solutions globally.

The three current OASC MIMs are all relevant in a Digital City Twin context:

- **MIM1: Context Information Management** including a Context Information Management API allowing to access real-time context information from different cities;
- **MIM2: Common data models** containing a set of guidelines and catalogue of common data models in different verticals (domains) to enable interoperability for applications and systems among different cities;
- **MIM3: Ecosystem Transactions management,** including a Marketplace API. The Marketplace API exposes functionalities such as catalogue management, ordering management, revenue management, Service Level Agreements (SLA), license management, etc. Complemented by marketplaces for hardware and services.

Other MIMs in preparation are MIM4 on Personal Data Management and MIM5 about Fair Artificial Intelligence.

3.8 DTC – Digital Twin Consortium

The Digital Twin Consortium (DTC) is a member organisation active in the field of Digital Twin technology, recently including Digital City Twins. The Digital Twin Consortium is a global ecosystem of users who are accelerating the digital twin market and demonstrating the value of digital twin technology.
Digital Twin Consortium working groups. There are working groups dealing with infrastructure, natural resources, manufacturing, healthcare, aerospace and defence. The Working Groups are teams of members that collaborate to address the technology and business needs of specific vertical markets.

3.9 BDVA/DAIRO – Big Data Value Association - Data, AI and Robotics

The BDVA is an industry-driven international not-for-profit organisation active in Europe to position Europe as the world leader in creating Big Data Value. The mission of the BDVA is to develop the Innovation Ecosystem that will enable the data and AI-driven digital transformation in Europe, delivering maximum economic and societal benefit and achieving and sustaining Europe's leadership on Big Data Value creation and Artificial Intelligence. To strengthen the BDVA goals, a name changes to DAIRO will take place in 2021. The BDVA has different programmes around:

- **Education**: The goal is to increase skill in Data Science and bringing new talent to the data economy;
- **Mobility programme**: Goals is providing an overview of funded and open access initiatives around the exchange of professionals, students and researchers;
- **Skills recognition program**: The goal is to provide the possibility for data scientists to display their skills online;
- **Big data centres of excellence**: The goal is to facilitate collaboration, exchange best practices, and promote the establishment of new centres.

The BDVA is, in the first place, an ecosystem and network organisation. The impact on standardisation is rather indirect via, e.g. Its education and skills programmes and its centres of excellence.

3.10 Gaia-X and IDSA

GAIA-X [5]
GAIA-X is a proposal for a next-generation European data infrastructure. It is a secure, federated system meeting the highest standards of digital sovereignty and constructed to promote innovation. Created by representatives from business, science and politics all over Europe, the project has the ambition to be the cradle of an open and transparent digital ecosystem. This digital ecosystem will be a safe haven where data and services can be made available, collated and shared in an environment of trust.
The GAIA-X Hubs act as the voices of user ecosystems on a national level. Their main objectives are to develop ecosystems, bundle national initiatives, and provide a central point of contact to interested parties in their respective countries. All GAIA-X Hubs are in close exchange with each other to ensure international

alignment regarding their activities, the definition of requirements, and the identification of regulatory hurdles. A future network of GAIA-X Hubs will support the growth of a bottom-up dynamic from the National toward the European level.

International Data Spaces [6]

The international data spaces (IDS) standard, which enables open, transparent and self-determined data exchange, is a central element of the GAIA-X architecture. IDSA contributes its knowledge to GAIA-X. They have been active in GAIA-X from the start as a founding member of the GAIA-X AISBL. This is the not-for-profit association within the initiative. IDSA is a leader in the field of data spaces' concept and design principles. They have been working in this field since 2016 as a cross-industry, transnational coalition of over 120 leading companies and research organisations.

4 Standards related to the URBANAGE Digital Twin solutions

Based on the work done by the relevant standardisation organisations, a preliminary set of standards and their potential use for **URBANAGE** are described below.

4.1 Internet of Things (IoT) Standards

The IoT standard landscape has been split up into:

- Sensor Data Standards;
- Context Information Management Standards;
- Time Series Standards.

4.1.1 Sensor Data standards

4.1.1.1 Formats & schemas

SSN/SOSA

The Semantic Sensor Network (SSN) ontology is an ontology used to describe sensors and their observations, the studied features of interest, etc... SSN follows a horizontal and vertical modularization architecture by including a lightweight but self-contained core ontology called SOSA (Sensor, Observation, Sample, and Actuator) as its core. Vertical modules build upon each other and import lower-level modules. These lower-level modules are independent of the higher-level modules and logically are self-contained. On the other hand, horizontally layered modules may depend on each other, e.g. they may rely on the directional import of another horizontal module.

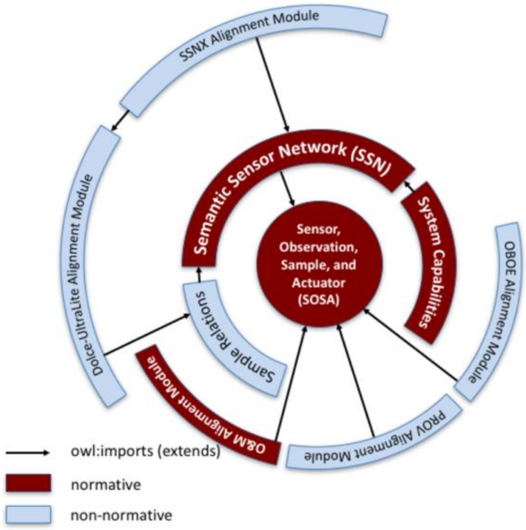


Figure 1: SOSA (Sensor Observation Sample and Actuator) components scheme

Observations & Measurements

Observations & Measurements (O&M) is an international standard for modelling observation events and describing their relations to the measured properties, measurement procedure and the captured data resulting from the observation events. The standard originates from the Open Geospatial Consortium's Sensor Web Enablement (SWE) activity. It was needed as the common standardized data model for handling the measurement events occurring in different kinds of sensors, enabling a "Sensor Web", which will allow applications to access all types of sensors, their observations, all of it over the Web. Together with other SWE framework open standards as SensorML and Sensor Observation Service (SOS), O&M provides a system-independent, Web-enabled way of data exchange between different parts of sensor network and other systems using the captured sensor information.

SAREF

The Smart Applications Reference (SAREF) is an ontology that facilitates matching existing assets in the smart applications domain. Their main goal is specifying recurring core concepts in the smart applications domain, the main relationships between these concepts, and hypotheses to constrain the usage of these concepts and relationships. The SAREF suite of ontologies provides definitions of generic classes that can be mapped to the NGSI-LD. Similar to NGSI-LD, SAREF also publishes domain-specific vocabularies in various domains, such as SAREF4CITY, a vocabulary for the Smart Cities.

Typically, three groups of standards can be identified when summarizing various data interoperability standards. The first group are FIWARE, NGSI-LD and SensorThings and are closest to the physical object and define a protocol on how data can be retrieved from those objects. The second group of standards are considered generic meta models because they define top-level terminology such as entities, properties, observations, measurements, etc... Examples of standards in this group are Observations & Measurements, SSN/SOSA and SAREF. The last group consists of standards that define domain-specific information models.

They extend the generic meta models and define terminology in the context of a specific application domain. SAREF-extensions and the SmartCity models are standards that belong to this group.

FIWARE Smart Data Models

The FIWARE Smart Data Models is a program to support adopting common compatible data models in smart solutions. A smart data model includes three elements: the schema or technical representation of the model defining the technical data types and structure, the specification of a written document for human readers, and the examples of the payloads for NGSIv2 and NGSI-LD versions. All data models are public and of royalty-free nature.

LDES

Linked Data Event Streams (LDES) is a lightweight publishing format. Each object is described in RDF, and each object in the stream is immutable. There are no underlying assumptions on the kind of data, so it is quite universal and supports both contextual (e.g. an address registry) and sensor data. Although these two types of data are very different, because contextual data is slow-changing, while sensor data changes quickly, both can be published as LDES. One of the design goals is efficient access, supported by the ability to publish multiple fragmentations enabling consumers to choose the most efficient ordering for their use case.

4.1.1.2 Protocols

NGSI-v2 & NGSI-LD

NGSI provides a simple data model for handling context information and is based on Entities, Attributes and Metadata for v2, and Entities, Properties, Relationships and Values for LD. Version 2 is based on JSON messages with a specific structure, whereas the LD version supersedes v2 and uses JSON-LD messages.
The FIWARE Context Broker provides an API supporting both NGSI-v2 and NGSI-LD to ingest context information.

LDES

LDES differs with the rather classic way of data publishing and data protocols, where data controllers have to publish multiple APIs on top of the dataset or even a data dump of the dataset, trying to meet the needs of their consumers. In the LDES ecosystem the only requirement for the data controller is to make its data available as a LDES. A data controller is free to provide additional functionality on top of the LDES in the form of a fragmentation and/or query API, but he can also leave that up for others, who may have other use cases.
The LDES specification is hypermedia-driven, aiming for evolvability rather than versioning of an API. The objects in the LDES should be described as much as possible with existing data standards (e.g. SSN/SOSA for sensor data), which will raise the semantic interoperability. Each LDES also has a technical contract to which the objects comply, called a Shapes Constraint Language (SHACL) shape. The shape of the LDES tells a consumer that all objects have been and will be validated by that shape, and informs the consumer on what information these objects contain.

MQTT

MQTT (Message Queuing Telemetry Transport) is a lightweight publish-subscribe network protocol that transports messages (as opposed to "bytes") between devices. MQTT is a standard protocol for publishing IoT sensor data since it allows messages to be buffered at the device side when connection issues arise.
MQTT is undoubtedly a sound choice for the southbound interface with the sensors, but there are better northbound interfaces.

OGC SensorThings API

The SensorThings API provides an open-source and uniform API to connect IoT devices, data and applications on the Web. SensorThings provides two main functionalities, each of which has its API: sensing and tasking. The Sensing part provides a standard way to manage and retrieve observations and metadata from IoT sensors and offers similar OGC Sensor Observation Service (SOS) functionality. The Tasking part provides a standard way for parameterizing - tasking - of taskable IoT devices, such as individual sensors and actuators. In addition, the Tasking part provides functions similar to the OGC Sensor Planning Service (SPS). The main difference between the SensorThings API and the OGC SOS and SPS is that the SensorThings API is designed specifically for resource-constrained IoT devices and the Web developer community.
The Sensing Part of the SensorThings API was designed based on the model of Observations and Measurements.

4.1.1.3 Potential use in URBANAGE

Smart city solutions, including digital twins, rely on actual data for historical what-if and predictive analysis. IoT devices are an important source of such actual data, which is collected through event sourcing. Relevant examples of such data sources are traffic intensity, weather and air quality conditions, lighting conditions, etc.
Besides IoT events, other events may be of interest as well. This is why we specifically mention LDES in this category. LDES provides a more generic way to perform event sourcing. It supports fast-moving IoT data sources and slower moving events that track the evolution of the context. Relevant examples in this context are points of interest, weather, parks and gardens, addresses, demography data.
Data analysis and programmed rules can also be used to control aspects of the city such as smart lighting, traffic control, prevention, etc. Both the OGC and NGSI standards mentioned support actuation by automated systems to implement such patterns.

4.1.2 Context Information Management Standards

4.1.2.1 NGSI-v2 & NGSI

NGSI and NGSI-LD are information models and API's for publishing, querying and subscribing to context information. Its added value lies in the open exchange and sharing of structured information between different stakeholders. Important domains where NGSI is used are Smart Cities, Industry and Agriculture.
NGSI-v2 and NGSI-LD are de facto standards by the FIWARE association. Both are designed to ingest and publish context data.
NGSI-LD adds linked data elements and is based on RDF. It improves interoperability by adding semantic data following the W3C Linked Data principles. The NGSI-LD Context Information Management API allows users to provide, consume and subscribe to context information. It enables close to real-time access to information coming from many different sources (named context sources), as well as publishing information through interoperable data publication platforms.

4.1.2.2 LDES & Subject Pages

Linked Data Event Streams can be used to publish a historical dataset. LDES supports only immutable data, which means that entities, properties or values can only be changed by publishing a new version of the affected item. This implies that once the data has been published, we can always refer to the same time period without the risk of the late arrival of events changing our reference data. This is a very important feature in the context of evidence-based policies. When models make predictions or inferences, we need to go back to the input data, and we need to record data lineage for all generated data in the platform.
Since LDES is based on RDF, each entity has its own URI, which can be followed using a browser (dereferenced) to find out more. Similarly, all relations also have their own URI. When dereferencing the URI of an entity, it leads to a subject page, containing more information about the entity. Typically, that subject page will contain URIs of other entities which in their turn can also be dereferenced. By describing data as RDF and including links to other entities, clients can automatically discover new data.
LDES provides one or more tree structures (fragments), that enable retrieving multiple nearby elements, depending on which criterion the tree uses. This, in turn, reduces the number of requests that need to be handled by the server. For example, if the fragmentation is by id, the next and previous history of the requested element is either on the same page or in the linked pages.

4.1.2.3 Potential use in URBANAGE

NGSI-LD offer a valuable approached of the management of context information, leveraging also the capabilities related to the linked data (since NGSI-LD is based on JSON-LD), in particular in combination with tolls suited for this purposed, such as the Context Broker [7] that offer publish-subscribe capabilities. The Context Broker is one of the building blocks of CEF Digital. LDES can capture fast-moving and slow-moving events and turn them into linked data using generic components. Additionally, it can be used to publish data efficiently. Finally, it can help publishers of data (public and private actors) make their data more accessible.

4.1.3 Time series standards

4.1.3.1 NGSI-v2 TSDB

FIWARE QuantumLeap is a generic enabler that is used to persist context data into a time series database like CrateDB or TimescaleDB. This is done using the TSDB specification [8], which provides an API to ingest and query the time-series data.

The information model for time series data is divergent from the information model for ingesting it. Ingesting time series into the time-series data store is done by having the TSDB API register a callback in the NGSI-v2 API that inserts one or more measurements into the TSDB.

4.1.3.2 NGSI-LD TEMPORAL API

Unlike the NGSI-v2 specification, the NGSI-LD specification includes a temporal API and information model as a part of the standard. Although quite similar to NGSI-v2, NGSI-LD puts forward a more powerful time-series query language.

4.1.3.3 LDES

Linked Data Event Streams is a rather new data publishing strategy that answers a lot of questions around costs. Smart APIs scale with adding more processing power, which can quickly become a bottleneck. LDES shifts processing power to the client, at the cost of (some) more bandwidth. It also provides a uniform API that is easy to use by machines instead of custom APIs that need to be implemented by each client again. LDES also scales in the usage pattern direction. When a client needs access that is inefficient with the provided fragmentations (similar to indexes in databases), anyone can host a new fragmentation to support those new use cases.

4.1.3.4 Potential Use in URBANAGE

Capturing data streams and turning them into timeseries is instrumental to understanding dynamics in the city and making predictions or simulations. When this data comes from sensors, the data volumes can take on considerable proportions. It is therefore essential to adhere to scalable technologies, strategies and standards as much as possible.

4.2 Metadata

Metadata is often described as “**data about data**”. ISO speaks about information about a resource. “Resource” is a purposely general term aimed at emphasizing the generality of the ISO metadata standards (and models). A resource can be a service, a collection site, software, a repository, or many other things. The NISO [9] National Information Standards Organisation) based Wikipedia definition “**Metadata is data [information] that provides information about other data**” describes distinct metadata types: descriptive metadata, structural metadata and administrative metadata. The difference between these metadata types is important as they can make discussing metadata more complicated than needed when people have different definitions of metadata in mind.

NISO distinct a number of basic types of metadata: descriptive metadata, structural metadata, and administrative metadata are defined as follow:

- **Descriptive metadata** describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords;
- **Structural metadata** is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials;
- **Administrative metadata** provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

In practice, metadata are not only relevant for describing data, but also for other resources that allow access to those data (e.g. web services) or resources that allow data processing (e.g. software tools).

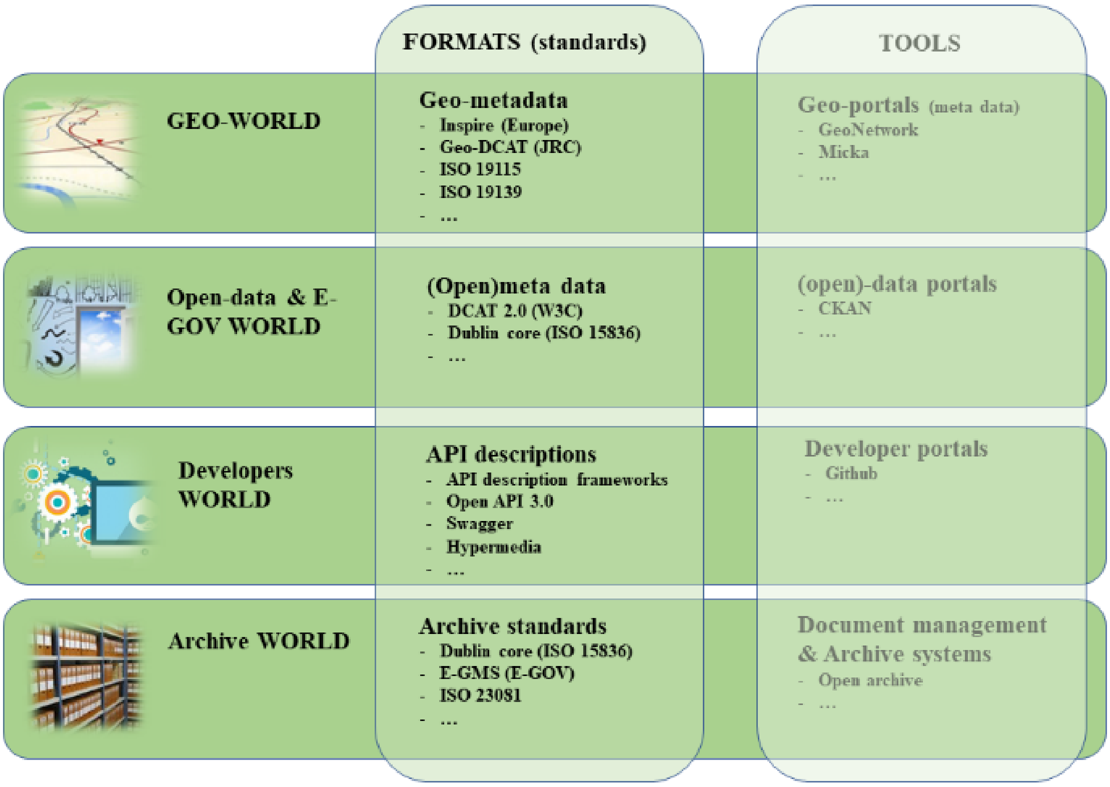


Figure 2: Metadata elements overview

In the **URBANAGE** Digital Twin, the focus will be on descriptive metadata and, to a lesser extent to administrative metadata. A Digital Twin can use and integrate metadata standards from different domains, including standards from the developer and archive domain. It is expected that a Digital Urban Twin will integrate mainly metadata information from the E-GOV & open data domain combined with data from the Geospatial domain.

4.2.1 Open metadata standards

4.2.1.1 Dublin Core (DMCI)

The Dublin Core Schema or Dublin Core Metadata set can be used to describe documents, web resources (video, images, web pages, etc.), physical resources such as books or CDs, and objects like artworks. The small set of terms can be used for various purposes, going from simple resource descriptions to combining metadata vocabularies of different metadata standards or providing interoperability for metadata vocabularies in Linked Data Cloud and Semantic Web implementations.

There are minimal, strict regulations in the Dublin Core Standard. Each metadata element is optional, can be repeated, and there is also no prescribed order for presentation or use [10]. The fifteen element descriptions, available in The Dublin Core Metadata Initiative Metadata Terms in an abbreviated reference version, are formally authenticated in the ISO Standard 15836:2009 [11]. There are additionally 18 metadata terms that are designated as qualified Dublin Core and 55 metadata terms as full Dublin Core.

4.2.1.2 DCAT (W3C)

DCAT is an RDF[1] vocabulary designed to facilitate interoperability between data catalogues published on the Web. By using DCAT to describe datasets in data catalogues, publishers increase discoverability and enable applications to consume metadata from multiple catalogues. It further enables decentralised publishing of catalogues and facilitates federated dataset search across sites. DCAT makes extensive use of terms from other vocabularies, in particular, Dublin Core.

DCAT was originally conceptually based on FRBR (Functional Requirements for Bibliographic Records)[2] in a simplified way. But today, it has grown into a specification in its own right. It is about cataloguing resources. For that, it defined these core concepts:

- Catalogue: a curated collection of metadata about resources;
- Catalogued Resource: Resource published or curated by a single agent;
- Catalogue Record: a record in a Catalogue, describing the registration of a single Catalogued Resource.
- Dataset: A collection of data, published or curated by a single agent and available for access or download in one or more representations;
- Data Service: A collection of operations that provides access to one or more datasets or data processing functions;
- Distribution: Considered as a specific representation of a dataset. A dataset might be available in multiple serializations that may differ in various ways, including natural language, media type or format, schematic organization, temporal and spatial resolution, level of detail or profiles (which might specify any or all of the above).

A Catalogued Resource can be a Dataset, Data Service or even a Catalogue. Distributions are dependent entities of a Dataset; they cannot exist without. The introduction of Data Services has been an important change. It addressed the ever-ongoing discussion of whether an API is a distribution or not. Data Services are an entity in their own right distinctive from distributions.

[1] RDF Schema (Resource Description Framework Schema, variously abbreviated as RDFS, RDF(S), RDF-S, or RDF/S) is a set of classes with certain properties using the RDF extensible knowledge representation data model, providing basic elements for the description of ontologies, otherwise called RDF vocabularies, intended to structure RDF resources.

[2] Functional Requirements for Bibliographic Records.

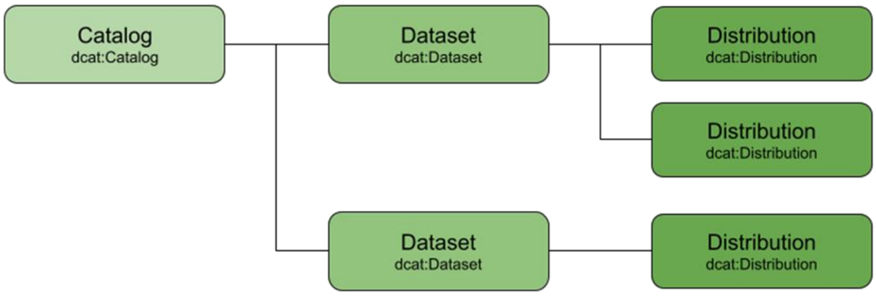


Figure 3: DCAT 2.0 Structure

DCAT (1.0) was released in January 2014, a revised version, DCAT (2.0), prepared by the W3C Data Exchange Working Group (DXWG)7 has been released in February 2020. The revision is a response to a new set of Use Cases and Requirements submitted based on experience gained with the DCAT vocabulary from the time of the original version and new applications initially not considered. Today a new release is being prepared to include, for instance, Data Series as a novel notion.

The new version deprecates and will not delete existing terms. In addition, the work of the DXWG investigated the development of guidance for the publication of application profiles of vocabularies as well as in the principle of content negotiation by application profile.

4.2.1.3 DCAT-AP (Application Profile)

The DCAT Application Profile (DCAT-AP) [12] for data portals in Europe is a DCAT profile describing public sector datasets in Europe. It enables cross-data portal search for data sets and improves public sector data search across borders and sectors in Europe.

DCAT-AP extends and precise the use of each term in DCAT. For instance, it determines mandatory, recommended and optional information and the code lists to be used. In addition, it provides guidelines or best practices for aspects that cannot be formally expressed.

DCAT-AP is the basis for the EU Data portal [13]

This portal harvests EU member states national data catalogues to provide a complete overview of Open Data available under the directive on Open Data and reuse of PSI [14] (Formally known as the PSI directive).

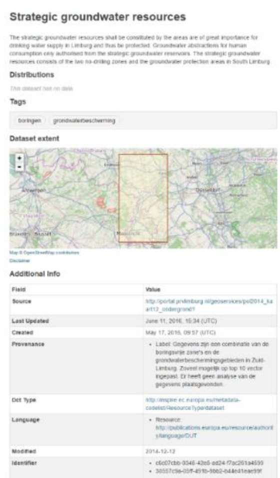


Figure 4: DCAT-AP implementation example – EU Data portal

DCAT-AP is further profiled by EU member states such as Germany [15], Denmark [16], Italy, the Netherlands, Sweden, Norway, and regional profile such as in the Flemish region (Belgium).

In the past year, the Flemish Region has profiled DCAT-AP for the Open Data [17], and created a generic DCAT profile, metadata DCAT [18], covering all datasets and data services, including closed or for internal use only. Furthermore, adopting a coherent, integrated version of the GeoDCAT-AP [19], an EU extension for geographical data covering other things the needs of INSPIRE, is in active development.

GeoDCAT-AP is particularly useful as it has been designed in the first place as a DCAT-AP profile allowing a mapping from the geospatial standards, discussed in the next section, into the broader data community. Although the latest version is expressed more as a normal DCAT-AP profile, the mappings from ISO are maintained as a reference implementation. The investigation for a Flemish GeoDCAT-AP profile shows that most of the collected data in the Flemish INSPIRE catalogue are captured, making the mapping almost without data loss. Via this mapping/profile, geospatial metadata descriptions can be explored in a DCAT setting.

4.2.1.4 ISO Standards

4.2.1.4.1 ISO 19115:2003 and ISO 19115:2014 – Geographic information metadata

ISO 19115 is a standard of the International Organization for Standardization (ISO). The standard is part of the ISO geographic information Suite of Standards (19100 series)[3]. ISO 19115 and its parts define how to describe geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use.
The objective of this International Standard is to provide a clear procedure for the description of digital geographic datasets and services so that users will be able to determine whether the data in a holding will be of use to them and how to access the data. By establishing a common set of metadata terminology, definitions and extension procedures, this standard promotes the proper use and effective retrieval of geographic data.
ISO 19115 was revised in 2013 to accommodate growing use of the internet for metadata management, as well as to add many new categories of metadata elements (referred to as code lists) and the ability to limit the extent of metadata use temporally or by the user.

In 2016 ISO/TS 19115-3:2016 was adopted. It defines an integrated XML implementation of ISO 19115-1, ISO 19115-2, and concepts from ISO/TS 19139.

4.2.1.4.2 ISO 19119:2005 and ISO 19119:2016 – Geographic information services

ISO 19119:2005 is the international standard for geographic information services. The scope of this standard ranges from identifying and defining the architecture patterns for service interfaces used for geographic information to defining its relationship to the Open Systems Environment model. It presents a geographic services taxonomy and a list of example geographic services placed in the services taxonomy. Additionally, this international standard prescribes how to create a platform-neutral service specification and how to derive conformant platform-specific service specifications. Finally, the standard provides guidelines for selecting and specifying geographic services from both platform-neutral and platform-specific perspectives. [20]

ISO 19119:2005 was revised and resulted in ISO 19119:2016. The 2016 version defines requirements for how platform neutral and platform-specific specification of services shall be created, in order to allow for one service to be specified independently of one or more underlying distributed computing platforms. It also defines requirements for a further mapping from platform neutral to service specifications, in order to enable conformant and interoperable service implementations. The service metadata model and

[3] Most of the standards of the ISO 19100 series were also adopted as CEN standards, they are usually indicated with EN and then the ISO name.

requirements have been moved to ISO 19115-1: 2014 (clause 6.5.14). So, this new version only speaks in very general terms about the need for service metadata.

The INSPIRE Metadata specifications or Implementing Rules and the 'INSPIRE Metadata Technical Guidelines' are based on EN ISO 19115 and EN ISO 19119:2005 for describing metadata for data sets, data sets series and services. [21]

4.2.1.4.3 ISO 19110:2005 and 2016 Geographic methodology for feature cataloguing

ISO 19110:2016 the successor of the 2005 version defines the methodology for cataloguing feature types. This document pertains to the cataloguing of feature types that are presented in digital form. It describes how to organise feature types into a feature catalogue. The document also explains how the feature catalogue can be presented to the users of a set of geographic data. This document can be used to create catalogues of feature types in previously uncatalogued domains and to revise existing feature catalogues to comply with standard practice. To catalogue other forms of geographic data, the principles in this document can also be applied. [22]

4.2.1.4.4 ISO/TS 19139:2007 Geographic information Metadata XML Schema implementation

Defines Geographic MetaData XML(GMD) encoding, an XML Schema implementation derived from ISO 19115 [23]. Geographic metadata is represented in ISO 19115 as a set of UML packages containing one or more UML classes. ISO 19115 provides a universal, encoding-independent view of geographic information metadata. ISO/TS 19139 provides a universal implementation of ISO 19115 through an XML schema encoding that conforms to the rules described in ISO 19118 (Encoding).

ISO/TS 19139 is currently under review and will be replaced by ISO/NP TS 19139-1. Also here, the specific metadata XML encoding will be removed and is now part of ISO/TS 19115-3: 2016 which defines an integrated XML implementation of ISO 19115-1 and ISO 19115-2 (for gridded data).

4.2.1.5 INSPIRE specification

In May 2007, the INSPIRE Directive came into force. This European Directive is put into force in various stages but has to be fully implemented by 2021 (see Inspire roadmap [24]). The Directive evolved from spatial information infrastructures established and operated by the European Union Member States. INSPIRE addresses 34 spatial data themes. [25]

The INSPIRE Directive wants to create a European Union spatial data infrastructure to the benefit of EU environmental policies and policies or activities that may impact the environment. Once in place, this spatial data infrastructure will facilitate the sharing of environmental spatial information among public sector

organisations, enhance public access to spatial information across Europe and assist in policy-making all over Europe.
The INSPIRE Directive emphasises the priorities for the development of the European spatial data infrastructure being, and we quote: "*the infrastructures for spatial information in the Member States should be designed to ensure that […] it is easy to discover available spatial data, to evaluate their suitability for the purpose and to know the conditions applicable to their use*."[4]

The 'INSPIRE Metadata Implementing Rules' [26], an EU Regulation[5] approved in 2008, have to ensure that the thematic information sources coming from the 28 EU member states are interoperable on the metadata level. It establishes the obligations under which public sector bodies in the EU should publish descriptive metadata on geographic data sets (series) and services. Additionally, it describes a set of twenty-one metadata elements to be used to describe geographic datasets, dataset series and services. This set was extended by five in 2010 and with one more in 2013 [27].

In order to further harmonise the implementation of these obligations, the 'INSPIRE Metadata Technical Guidelines' [28] were defined. Although these guidelines are non-binding, it is advised to use them. Otherwise, it is challenging to demonstrate that the metadata provided in a different model or format is in accordance with the binding INSPIRE Metadata Implementing Rules. The INSPIRE technical guidelines demonstrate how the metadata elements in the Regulation match metadata elements in ISO 19115: 2003 and ISO 19119: 2007. The specification also contains examples of how metadata records can be transformed into XML[6].

Although the INSPIRE technical Guidelines prescribe the use of ISO19115:2003, a metadata record that complies with the ISO19115: 2003 core elements does not fully comply with the INSPIRE Metadata Implementing Rules and Technical Guidelines. There are some additional metadata elements that INSPIRE requires on top of the ISO19115: 2003 core profile, for example, the (degree of) 'conformity' of spatial datasets with the INSPIRE Data specifications.

4.2.2 Potential use in URBANAGE

The use of open data and geospatial metadata related standards in URBANAGE will be mainly linked to data catalogue related activities.

---

[4] DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) https://inspire.ec.europa.eu/legislation-details/directive-20072ec-european-parliament-and-council

[5] COMMISSION REGULATION (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata

[6] Part 2.1.5. of the Information Flanders “metadata study”, Information Flanders, May 2017.

Dublin Core is one of the easiest ways, and most common ways to describe metadata and is supported by many metadata tools. Candidate metadata tools like the open-source tool CKAN [29] supports DUBLIN core mappings. Also, other open source widely used geospatial metadata tools like GeoNetwork (An application to manage spatially referenced resources [30]) support Dublin Core. [31] DCAT [32] builds further on Dublin Core and is one of the most frequently used standards to describe datasets and dataset catalogues and is also supported by CKAN.
Those tools uses a DCAT-AP mapping/plugin as the way to integrate DCAT. But as above is shown, it is possible or even recommended to consider creating a dedicated profile for the application context. It should be investigated in **URBANAGE** what information is required to drive the pilots. If the catalogued resources are all part of the PSI directive, then a full profile of DCAT-AP can be considered.

The advantages of DCAT are numerous: a lively community, a standard suited for profiling, not limited to a single kind of catalogued resources and public catalogues in every member state ready to be queried. This allows for the following approach to be investigated: instead of designing a tool to manage the metadata of the datasets for a pilot locally, one can consider reinforcing the community by the following actions:

- Defining the metadata needs for **URBANAGE**;
- Express the metadata needs as a profile;
- Explore if the data can be extracted from the EU data portal (SPARQL is a machine queriable interface to the EU Data Portal [33]) (or the respective national portals) These data portals might already cover the discoverability aspect **URBANAGE** need.

According to this approach, the used services and datasets can then be included in the pilot configuration file. Then, using their URI, interested people can find the metadata.

In the geospatial field additional metadata is required. The ISO geospatial standards family was designed to extend the more generic Dublin Core standard to Geospatial use by adding e.g., coordinate reference systems and geospatial extends. Open and proprietary geospatial information systems adopt the ISO standards (e.g. ESRI ArcGIS) and metadata systems like GeoNetworks or Micka [34]. The use of ISO19115 and ISO19139 is an obvious and widely supported choice for using geospatial metadata in a Digital Twin solution like **URBANAGE** without many needs to change existing geospatial metadata catalogues.

The EU Inspire regulation adds extra metadata implementing rules on top of the ISO19115:2003. Since many governments owned datasets cover themes that are part of the Inspire regulations and at least some of these datasets will be used in a Digital Twin, it makes sense to use an Inspire compatible metadata tool. This must guarantee that all metadata elements will be available in a Digital Twin. GeoNetwork, Micka, and the ESRI ArcGIS are examples of tools supporting the INSPIRE Metadata Implementing Rules.

4.3 AI related standards

4.3.1 Model Storage Formats

Different standardization initiatives and groups are active on the definition of standards and reference languages to represent and exchange data mining/analytics and artificial intelligence models. Here, the main interest is reducing deployment fragmentation and reducing the gap among training and inference engines, extending the range of devices and platforms supported.

The Data Management group [35], hosts the working groups in charge of the definition of Predictive Model Markup Language (PMML [36]) and the Portable Format for Analytics (PFA [37]), two complementary standards for analytic models. PMML is widely adopted by the industry, simplifies the exchange of information among systems. PFA is an emerging standard for statistical models and data transformations, supporting flexible chains of tools, integrating models, pre-processing, and post-processing to define complex workflows. The Open Neural Network Exchange (ONNX [38]) is an open-source artificial intelligence partner ecosystem that establishes open standards for representing machine learning algorithms (artificial neural networks) and software tools to promote collaboration.

Khronos Group, an American non-profit member-funded industry consortium pushing open standards as OpenCL, is currently proposing Neural Network Exchange Format (NNEF [39]), a dedicated data exchange format. Also, the Apache Community defines SystemML as a workplace for machine learning, supporting a declarative large-scale machine learning algorithms specification with a generation of hybrid runtime plans over Apache technologies. Spark MLlib [40] also standardizes APIs for machine learning algorithms to define processing pipelines. Artificial Intelligence Markup Language (AIML [41]), defines an XML format for creating natural language software agents.

These standards focus on interoperability, and algorithm design and implementation aspects, as model debuggability and interpretability are missing.

4.3.2 Architecture and process standards

Some of the most relevant standardization bodies are addressing intelligent systems and their underlying technologies as artificial intelligence, data analytics, etc.

- IEEE portfolio of AIS technology and impact standards and standards projects [42]. After a review of the existing and ongoing standards, potentially, the following are the most relevant at this stage:
    - IEEE P2807. Framework of Knowledge Graphs (Representation), including input requirement, construction process of KG, i.e., extraction, storage, fusion and understanding, performance metrics, applications, related artificial intelligence (AI) technologies and other required digital infrastructure.

- IEEE P283 - Standard for Technical Framework and Requirements of Shared Machine Learning, defines a framework and architectures, defining functional components, workflows, security requirements, technical requirements, and protocols, for a federated training of models.
- IEEE P2840 - Standard for Responsible AI Licensing. The standard describes specifications for the factors for a Responsible Artificial Intelligence (AI) license. Includes: Standardized definitions for referring to components, features and other such elements of AI software, source code and services, standardized reference to geography specific AI/Technology specific legislation and laws.
- IEEE P2863 - Recommended Practice for Organizational Governance of Artificial Intelligence specifies governance criteria and process steps for effective implementation, performance auditing, training, and compliance in developing or using artificial intelligence within organizations. IEEE P2894 (Guide for an Architectural Framework for Explainable Artificial Intelligence) is also very aligned.
- IEEE P7003 - Standard for Algorithmic Bias Considerations, methodologies to reduce negative bias in the creation of their algorithms, avoiding overly subjective or uniformed data sets or information known to be inconsistent with legislation concerning certain protected characteristics, or with instances of bias against groups.

- National Institute of Standards and Technology (US) (NIST [43]) has released a plan for prioritizing federal agency engagement in the development of standards for artificial intelligence (AI), described in the Executive Order on Maintaining American Leadership on Artificial Intelligence (EO 13859 [44]). The plan recommends to "commit to deeper, consistent, long-term engagement in AI standards development activities to help the United States to speed the pace of reliable, robust, and trustworthy AI technology development."
- Standard and/or project under the direct responsibility of ISO/IEC JTC 1/SC 42 Secretariat that includes some potentially relevant standards as:
    - ISO/IEC DTS 4213. Information technology — Artificial Intelligence — Assessment of machine learning classification performance.
    - ISO/IEC AWI 5259. Data quality for analytics and ML — Part 1: Overview, terminology, and examples, Part 2: Data quality measures, Part 3: Data quality management requirements and guidelines and Part 4: Data quality process framework
    - ISO/IEC AWI 5338. Artificial intelligence — AI system life cycle processes.
    - ISO/IEC AWI 5339. Artificial Intelligence — Guidelines for AI applications
    - ISO/IEC AWI 5392. Artificial intelligence — Reference architecture
    - ISO/IEC AWI TS 5471. Artificial intelligence — Quality evaluation guidelines
    - ISO/IEC AWI TS 6254. Artificial intelligence — Objectives and approaches for explainability of ML models and AI systems
    - ISO/IEC 20546:2019. Information technology — Big data — Overview and vocabulary
    - ISO/IEC TR 20547-1:2020. Information technology — Big data reference architecture — Part 1: Framework and application process, Part 2: Use cases and derived requirements, Part 3: Reference architecture and Part 5: Standards roadmap

- ISO/IEC DIS 22989. Information technology — Artificial intelligence — Artificial intelligence concepts and terminology
- ISO/IEC DIS 23053. Framework for Artificial Intelligence (AI) Systems Using ML
- ISO/IEC CD 23894.2. Information Technology — Artificial Intelligence — Risk Management
- ISO/IEC DTR 24027. Information technology — Artificial Intelligence (AI) — Bias in AI systems and AI aided decision making.
- ISO/IEC TR 24028:2020. Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence

4.3.3 Model life cycle

All Artificial Intelligence (AI) projects follow a model life cycle consisting of well-defined steps. The typical steps involve various interdependent tasks using a variety of tools, techniques or programming languages (e.g. ISO/IEC AWI 5338).

1. Business understanding. Before starting an AI project, it is important to understand the project, its specifications, requirements, priorities, among others. In this step, it is also relevant to identify the project's main objective, the variables needed and formulate initial hypotheses to test.
2. Data collection. This step involves data acquisition that can come from multiple sources and formats.
3. Data cleaning. This step is a time-consuming task that involves preparing the data, removing discrepancies, missing values, outliers, structuring data from raw files, etc. Note that data collection, data understanding, and data preparation may take up to 70-80% of the overall project time.
4. Data exploration. The exploratory analysis involves understanding the variables, the patterns, correlation between variables, data distributions (measures of frequency, dispersion, central tendency), univariate and multivariate analysis to identify trends in the data.
5. Model development. This step involves the modelling of the AI project. The AI modelling is employed to find patterns or behaviours in data, feature engineering and can be done by descriptive analytics or predictive modelling. The type of problem (supervised/unsupervised) leads to some Machine Learning algorithms or others.
6. Data visualization. The graphical representation of the data can help understand trends, patterns in data and give insights into the problem.

4.3.4 Potential use in URBANAGE

In the **URBANAGE** project, the lifecycle as mentioned above will be used together with the related tools and techniques. A deep analysis of existing standards, mainly those related to critical aspects, risks and compliance with new European regulations, will be carried out once the technical approach and methods or techniques to be applied are clearer. In the Flanders case, AI can be used to compare calculated scores (for example, a calculated heat stress zone) with the scores given by human users. AI can help to close the gap between the perceived and calculated scores.

4.4 Big Data related standards

Big Data has been addressed in several standardisation bodies during the last years, with several standards (including ISO/IEC) already published related to architecture and specific aspects of the solutions. With the Big Data Interoperability Framework (NBDIF [45]) a broadly supported ecosystem and the IEEE Big Data Community [46] both steer big data related standards and architecture related developments.

The Big Data Interoperability Framework (NBDIF) is a consensus-based, vendor-neutral, technology- and infrastructure-independent ecosystem established by the NIST Big Data Public Working Group (NBD-PWG) integrating partners from the industry, academia and government. It enables Big Data stakeholders to utilize the best available analytics tools to process and derive knowledge through standard interfaces between swappable architectural components.

The NBDIF is being developed in three stages to achieve the following with respect to the NIST Big Data Reference Architecture (NBD-RA):

- Identify the high-level Big Data reference architecture key components, which are technology, infrastructure, and vendor-agnostic;
- Define general interfaces between the NBD-RA components;
- Validate the NBD-RA by building Big Data general applications through the general interfaces.

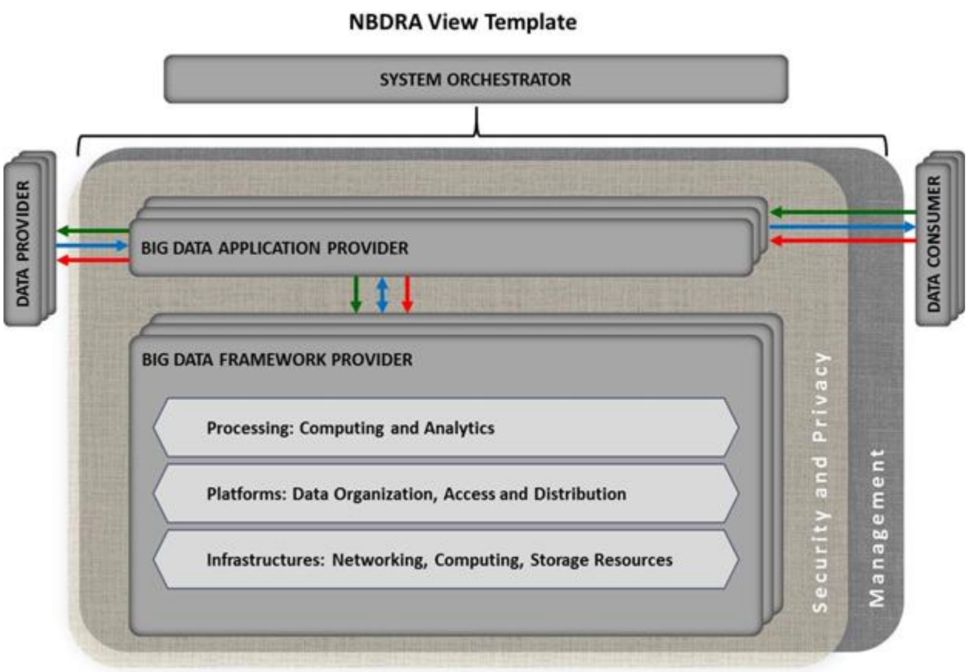


Figure 5: Top level roles in NIST Big Data Reference Architecture (NBD-RA [47])

The five main NBD-RA roles, shown in Figure 5, represent different technical roles in every Big Data system, and they refer to classes of functional components. These roles are System Orchestrator, Data Provider, Big Data Application Provider, Big Data Framework Provider and Data Consumer. Specifically, the System Orchestrator is related to policies and processes derived from community best practices or standards such as International Organization of Standardization (ISO) 20000 for IT Services Management or ISO 27000 for Information Technology Security. The functional components within the Big Data Application Provider include workflows, transformations, visualisations, access services and algorithms. Regarding the algorithms, the MapReduce and the Bulk Synchronous Parallel (BSP) models are quite popular.

Components within the Big Data Framework Provider fall within three categories: processing, platforms and infrastructures, and specific cross-cutting aspects. Specifically, the infrastructure frameworks provide the necessary resources (physical or virtual) for other components to run on top. The resources can be classified as networking, computing and storage. Virtualisation is frequently used to achieve elasticity and flexibility in allocating physical resources and is often referred to as infrastructure as a service (IaaS) within the cloud computing community. Virtualisation is implemented via hypervisors, either native or containerised/dockerised. The logical distribution of cluster/computing infrastructure may include a high-performance computing (HPC) cluster, a grid of physical commodity machines in a rack or a set of virtual machines running on a public cloud provider.

Moreover, computing resources may also include computation accelerators such as Graphics Processing Units (GPU) and Field Programmable Gate Arrays (FPGA). The AI algorithms developed in the context of **URBANAGE** can benefit from those accelerators. Finally, the storage infrastructure may include any resource from isolated local disks to network-attached storage (NAS).

Data platform frameworks support the logical data organisation and distribution and may range from simple delimited files to fully distributed relational or columnar data stores. Specifically, common Big Data Organization approaches include relational, key-value, columnar, document-based and/or graph storage platforms. The processing frameworks differentiate based on whether they support batch or streaming processing. Batch frameworks are mature enough and are typically distributed on a cluster of nodes. On the other hand, streaming frameworks are ideal when dealing with data that requires processing as fast as the velocity at which it arrives into the Big Data system.

The IEEE Big Data Community, an open group, supported by the IEEE Cloud Computing Initiative and multiple Societies, as the Reliability or Communications Societies, has already taken the lead on the technical aspects of big data. Some identified areas of need and opportunities for standardisation of data-related technologies have been included. The most relevant is the IEEE Big Data Governance and Metadata Management (BDGMM), developing an interoperable data infrastructure through extensible governance and metadata lifecycle framework.

The ISO/IEC JTC1 report [48] recaps a survey of the existing ICT landscape for key technologies and relevant results and scenarios for Big Data from the most relevant standardisation and relevant organisations. As a result, was the creation of WG9 for Big Data related standardisation activities, now discontinued, and its activities are continued in ISO/IEC JTC1 SC42 Artificial Intelligence. From the definition of the group, some topics of interest addressed were:

- Big Data Use Cases –application scenarios that require either mixed legacy system and/or horizontally scalable from the analytics;
- Big Data Analytics Processing – provide well-defined generic algorithms for distributed computing, metadata extraction, search and retrieval
- Big Data Architecture/Infrastructure – provide effective provision and configuration of horizontal applications, databases and resources;
- Big Data Security and Privacy – deal with infrastructure security, data privacy, data management, integrity and reactive security;
- Big Data Management – monitor scalable distributed high-performance systems to ensure Big Data applications are efficiently and securely executed;
- Big Data Challenges and Opportunities – explore approaches to address heterogeneity, scale, timeliness, complexity, security, and privacy problems in Big Data that can create value;
- Some standards published are:
    - ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary
    - ISO/IEC TR 20547-1:2020 Information technology — Big data reference architecture — Part 1: Framework and application process, Part 2: Use cases and derived requirements, Part 3: Reference architecture and Part 5: Standards roadmap

4.4.1 Potential use in URBANAGE

In general, all of the existing approaches for big data reference architectures have modules for collecting, storage, integration & preparation, analysis and visualization of data, with similar abstraction levels of these components. For example, the NIST Big Data Interoperability Framework or the ISO/IEC TR 20547 reference architectures can be seen as a meta-reference architecture used to derive specific instances, as for **URBANAGE**. Topics under interest will be considered in the context of T5.1 for the definition of the complete system architecture. Its functional and non-functional requirements and the programming interfaces among the individual components and in T3.3 in charge of the Big Data analytics components to process a large amount of data produced in the cities to extract knowledge and present the results.

4.5 Security related standards

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. ENISA works with the EU, its member states, the private sector and Europe's citizens to develop advice and recommendations on good practices in information security. It assists EU member states in implementing relevant EU legislation and works to improve the resilience of Europe's critical information infrastructure and networks. Since 2019, it has been drawing up cybersecurity certification schemes. In addition, ENISA has issued various studies introducing good practices for software security, with a particular focus on software development guidelines for security products and services throughout their lifetime. These guidelines conform to international standards of software security like ETSI TR 103 309 and OWASP Application Security Verification Standard.

ETSI (European Telecommunications Standards Institute) is an independent, not-for-profit, standardisation organisation in the field of information and communications. ETSI supports the development and testing of global technical standards for ICT-enabled systems, applications and services. It is one of the three bodies, next to CEN and CENELEC, officially recognised by the European Union as a European Standards Organization (ESO). The role of the European Standards Organisations is to support EU regulation and policies through the production of Harmonised European Standards and other deliverables. The standards developed by ESOs are the only ones that can be recognised as European Standards (ENs). Part of ETSI's operations is the creation of standards in the area of software security.

4.5.1 OAuth 2.0

The OAuth 2.0 authorization framework is a protocol that enables a third-party application to obtain limited access to an HTTP service. It is used in a wide variety of applications, including providing mechanisms for user authentication, and it's used to obtain limited access to services and resources by orchestrating an approval interaction between the resource owner and the HTTP service or by allowing the third-party application to obtain access to these.

The OAuth 2.0 is the industry-standard protocol for authorization that focuses on client developer simplicity while providing specific authorization flows for web/desktop applications, devices and services. This specification defines a delegation protocol useful for conveying authorization decisions across a network of web-enabled applications and APIs.

This specification and extensions are being developed within the **IETF OAuth Working Group** and more specifically follows the [49]

Figure 6 depicts the architectural diagram and a sequence diagram showing the authorization steps followed by this protocol. In particular, a client who calls a resource server needs to obtain a valid access token provided by OAuth 2.0 (steps 1,2,3). After successful authorization, the client can have access to the resource (steps 4,5,6,7,8).

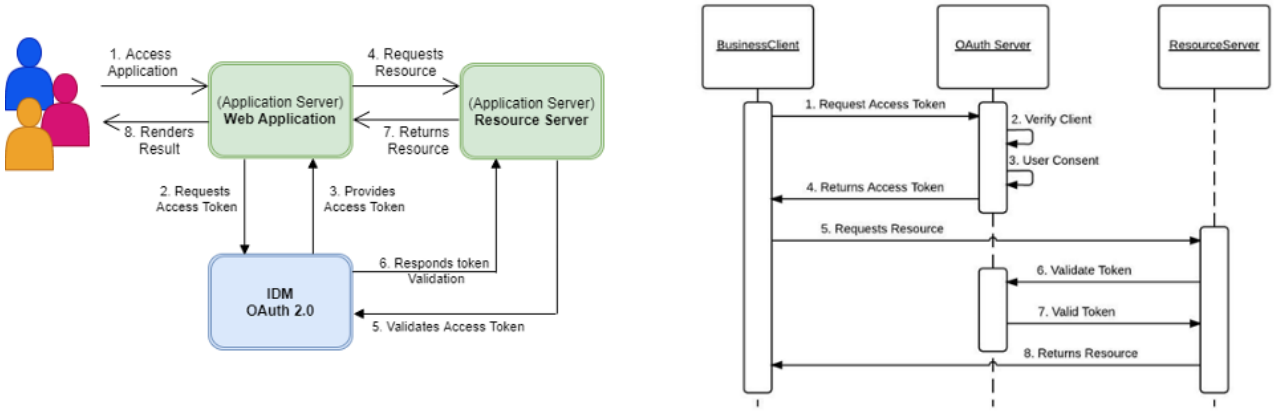


Figure 6: OAuth 2.0 Architecture and Sequence diagrams

4.5.2 SAML

The SAML 2.0 [50] stands for Security Assertion Markup Language 2.0. This is a specification for exchanging authentication and authorization identities between security domains. More specifically, the SAML version 2.0 is a protocol that uses security tokens containing assertions to pass information about an end-user /principal between a SAML authority (also called Identity Provider) and a SAML consumer (also called a Service Provider).

The SAML 2.0 specification was ratified as an **OASIS** Standard in March 2005, replacing a previous version of the security mark-up language, which was SAML 1.0. SAML 2.0 is based on the key concept of “assertion”. In SAML 2.0 specifications, there are three different kinds of assertion statements:

- **Authentication Assertion:** The assertion subject was authenticated by a particular means at a particular time;
- **Attribute Assertion:** The assertion subject is associated with the supplied attributes;
- **Authorization Decision Assertion:** A request to allow the assertion subject to access the specified resource has been granted or denied.

Furthermore, another important type of SAML assertion is the so-called “bearer” assertion used to facilitate web browser SSO.

4.5.3 OpenID Connect

OpenID Connect (OIDC) is an open standard published in 2014 (OpenID Connect [51] that defines an interoperable way to use OAuth 2.0 to perform user authentication. Instead of building a different protocol for each potential identity provider, an application can adopt one protocol for as many providers as they want to work with. Since it's an open standard, OpenID Connect can be implemented by anyone without restriction or intellectual property concerns.

As an extension of OAuth 2.0, the OpenID Connect is one of the most frequently used authentication protocols. OIDC is a full-fledged authentication and authorization protocol, while OAuth 2.0 works as a framework for building authorization protocols. The OIDC also makes heavy use of the JSON Web Token (JWT Web Tokens [52]) of standards following **RFC 7519-2015** [53]. These standards define an identity token JSON format and ways to digitally sign and encrypt that data in a compact and web-friendly way.

4.5.4 Potential use in URBANAGE

In the context of the **URBANAGE** project, the most suitable security guidelines are provided by ENISA. In compliance with the relevant standards provided by ETSI and other ones available such as OAuth 2.0, SAML and OpenID Connect, briefly described in the following, will be used to design and implement a secure system. This is needed to cope with aspects related to authentication and authorization, acting as an identity and access manager. There are diverse open-source identity and access management solutions available (such as Keycloak [54] and Keyrock [55]). These solutions are mainly based on standard protocols such as OpenID Connect, OAuth 2.0 and SAML.

4.6 Data visualisation and presentation

4.6.1 City data visualization

An important, eye catching, element of a Digital Twin in the digital representation of the city itself in 2D and 3D. Digital Twins are usually (re)using existing GIS and CAD generated data using the common domain standards. Two particular standards related to BIM (Building Information Modelling) are of particular interest for **URBANAGE**, CityGML [56] and IndoorGML [57].

CityGML or City Geographic Markup Language is a concept for the modelling and exchange of 3D city and landscape models. CityGML is a common information model for the representation of 3D urban objects. It defines the classes and relations for the most relevant topographic objects in cities and regional models with respect to their geometrical, topological, semantical and appearance properties. Targeted application areas include urban and landscape planning, architectural design, 3D cadastres, environmental simulations, disaster management and vehicle and pedestrian navigation. CityGML is realised as an open data model and XML-based format for the storage and exchange of virtual 3D city models. It is implemented as an application schema for the Geography Markup Language version 3.1.1 (GML3), the extendible international standard for spatial data exchange issued by the Open Geospatial Consortium (OGC) and the ISO TC211. CityGML is an official OGC Standard and can be used free of charge.

IndoorGML is an OGC standard for an open data model and XML schema for indoor spatial information. It aims to provide a common framework of representation and exchange of indoor spatial information. The main application domains are management of building components and indoor facilities and the usage of indoor space. IndoorGML is defined as an application schema of OGC Geographic Markup Language (GML) 3.2.1.

4.6.2 ISO 9241-210

The international standard on ergonomics of human-system interaction, ISO 9241 - 210:2019 [58] defines the user experience as a person's perceptions and responses that result from the use or anticipated use of a product, system, or service. User experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors, and accomplishments before, during and after use. The ISO lists three factors that influence user experience: system, user, and the context of use.

Figure 7: Human system interaction elements overview

During the development of a product and, more specific a software solution, it is essential to take the users' opinions and reactions into account. Furthermore, it is crucial to understand the user needs and use the knowledge to create a solution addressing these needs. But also, to create a solution that does so in a friendly and ergonomic way.

The ISO 9241-210:2019 provides requirements and recommendations for human-centred design principles and activities throughout the life cycle of computer-based interactive systems. It is intended to be used by those managing design processes and is concerned with how interactive systems' hardware and software components can enhance human-system interaction. It also provides an overview of human-centred design activities.

4.6.3 Potential use in URBANAGE

Based on ISO 9241-210:2019, several techniques and methodologies have been proposed to aid the development of a system and gather information about what the users say about the system and what the users do with the system. Some of them are depicted in the following two images.

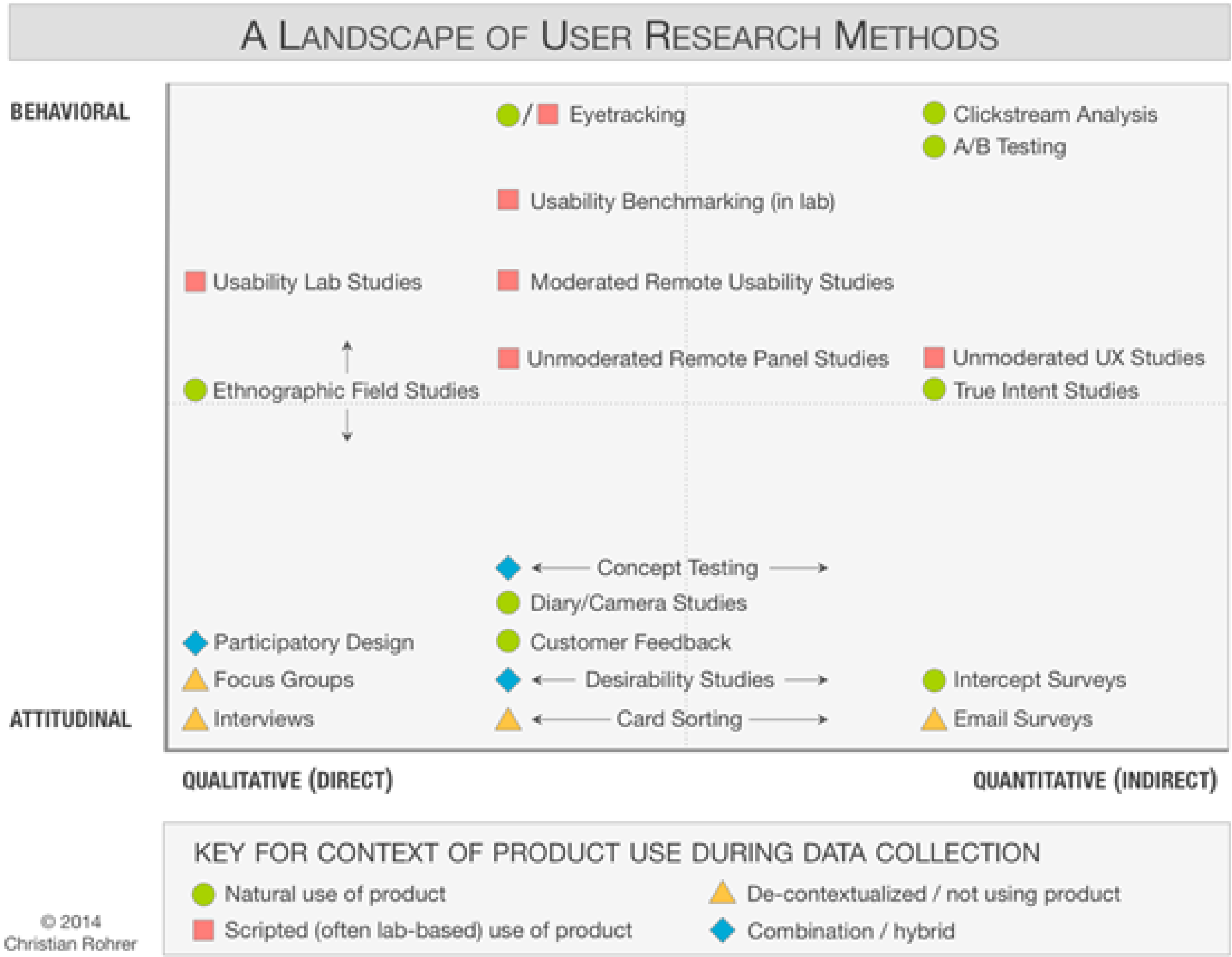
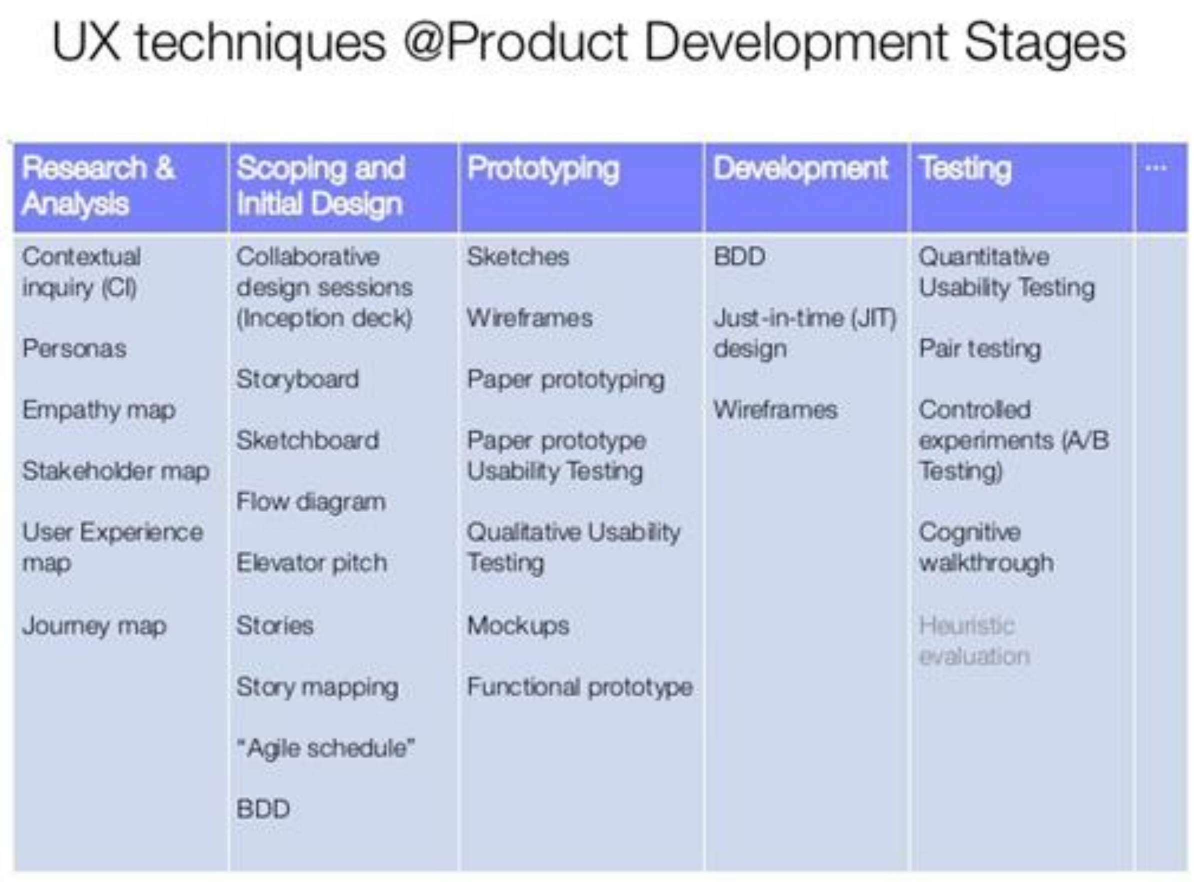


Figure 8: UX Research methods landscape


UX techniques @Product Development Stages

Research & Analysis	Scoping and Initial Design	Prototyping	Development	Testing	...
Contextual inquiry (CI) Personas Empathy map Stakeholder map User Experience map Journey map	Collaborative design sessions (Inception deck) Storyboard Sketchboard Flow diagram Elevator pitch Stories Story mapping "Agile schedule" BDD	Sketches Wireframes Paper prototyping Paper prototype Usability Testing Qualitative Usability Testing Mockups Functional prototype	BDD Just-in-time (JIT) design Wireframes	Quantitative Usability Testing Pair testing Controlled experiments (A/B Testing) Cognitive walkthrough Heuristic evaluation	

Figure 9: UX Techniques

In the context of the **URBANAGE** project, the most suitable techniques and recommendations of the ISO will be adopted and applied to the design and implementation of the system's User Interfaces. CityGML will be used to visualize the city topography and buildings, while IndoorGML will be considered when information is needed from inside the building. While the first can play a role in getting insight into the city's hilliness, the latter can be used to guide people with disabilities through the building itself.

5 Standardisation plan

The standardisation plan combines research on existing standards based on chapters 3 and 4 and the activities in the technical work packages WP2, WP3, WP4 and WP5. This standardisation plan provides a first overview of the use of standards and how the project could improve standards related to the **URBANAGE** objectives of using and contributing to new disruptive technologies. The standardisation plan and strategy cover the use of existing standards including the standardisation approach during the technical design and an overview of (data) standardisation evolutions where **URBANAGE** can play a role, based on the current knowledge, to contribute by e.g. extending or improving the (data) standard.

5.1 Use of existing standards and data standardisation approach during the technical design

The first step of the standardisation plan is to elaborate on the use of existing relevant standards. The first part of the approach is to elaborate on the standardisation body (SDO) and the standards themselves (community activity). A second step is to elaborate on the use itself. Chapter four, "potential use in **URBANAGE**", is the first step in this exercise and provide an overview of promising standards that can be re-used. The second step is to elaborate on applying the standards in the **URBANAGE** data and software architecture.

5.1.1 Overview of the used standards and data standardisation related activities (as is)

Based on the current insights on the potential use in **URBANAGE**, the standards below are listed as relevant and re-usable in **URBANAGE**.

Standard	Organisation	Type	Potential use in URBANAGE
LDES	SEMIC	IoT	Fast (IoT) and slow moving data event processing (generic)
Dublin core	W3C	Metadata	Generic metadata description as part of the data catalogue
DCAT	W3C	Metadata	Generic metadata description as part of the data catalogue
DCAT-AP	W3C	Metadata	Generic metadata description as part of the data catalogue
ISO19115 & ISO19139	ISO	Metadata	Geospatial metadata description as part of a geospatial data catalogue
Inspire	EC / JRC	Metadata	Geospatial metadata description as part of a geospatial data catalogue
NBDIF (components)	NIST	Big data	Architecture application (data collection, data storage, data processing and analysis
OAuth 2.0	IETF	Security	Digital City Twin Platform authorisation
SAML	OASIS	Security	Digital City Twin Platform Authentication and authorisation
OpenID	OpenID Foundation	Security	Digital City Twin Platform Single Sign on
ISO 9241-210:2019	ISO	Presentation	Digital City Twin Platform Standards on ergonomics and human-system interaction

Table 1: Overview of used standards and data standardisation tasks that might to be used "as-is"

5.1.2 Applying standards to the URBANAGE data and software architecture

However the definition of the **URBANAGE** Architecture is still under development, two architectures were identified as a starting point: a *high-level ICT Architecture of **URBANAGE**, based on reference approaches and best practices (e.g.* BDVA, FiWare) and following the layered architectural pattern with different levels of abstraction: the *Infrastructure layer*, the *Data Layer* and the *Data Management layer* includes components that serve four primary purposes: a) to access and aggregate data from the various external data sources like city databases, open data and sensors; b) to apply to condition, *the Business layer, the Presentation layer and the Security layer*. In parallel, the **URBANAGE** Digital Twin platform is an extensible platform that allows modelling the city and its processes through a virtual replica, supported by a set of components to facilitate data modelling and mapping, geospatial analysis and data retrieval through web services. A city information model (CIM) modelling the city elements, related to the four purposes, will connect the different elements that define the city and its processes. In short, the main challenge is the integration of both technical solutions, the agreement on a global style (e.g. federated vs centralized approaches) and a communication mechanism among the different components and subsystems to deal with the **URBANAGE** use cases.

Although a more detailed analysis will be conducted in the following months of the project, aligned with the whole architecture definition, the following paragraphs summarize some initial insights. In general, the initially proposed architecture is aligned with the reference BigData architectures identified in the previous section.

Infrastructure layer. The Big Data Interoperability Framework (NBDIF) infrastructure layer will be relevant here, identifying applicable technologies and resources.

Data Layer: Several standards, as the ISO19115 and ISO19139 and the Inspire regulation adding for geospatial metadata, will be adopted for the URBANITE Digital Twin solution. Additionally, DCAT, standard to describe datasets and dataset catalogues, will define the project metadata needed for **URBANAGE**.

Data Management layer: A deeper review of the BigData and AI-related standards will be conducted to ensure that the selected technologies are compliant with the regulations. The Data collection consists of acquiring data from different data sources. Sqoop allows to extract data from differently structured sources and store it in HDFS directly. It was created to handle large data sources that can be imported or exported in a Hadoop cluster. Regarding message queues, based on the publisher-subscribe pattern, MQTT is considered a lightweight protocol where the devices used have low power. Despite not being specifically a Big Data technology, it is worth mentioning for being ideal in the context of the Internet of Things. Apache Kafka is a distributed streaming platform that allows publishing and subscribing stream records, storing stream logs in a durable and fault-tolerant manner, and processing stream logs as they occur. It is mainly used to build real-time data flows, obtaining data between systems and applications quickly.

The Data Storage layer constitutes the persistent storage of the incoming data. Choice of the storage system is directly connected to chosen processing tools and must integrate with it seamlessly. Popular options include MongoDB, InfluxDB, Elasticsearch and HDFS. HDFS is a widely used open-source distributed file system (Hadoop Distributed File System), which provides high fault tolerance. It is designed to be deployed on low-cost hardware providing high network performance for accessing HDFS. It is an ideal data format for applications that have large datasets for exchanging data sets. NoSQL databases are more flexible than traditional databases by not being so structured and providing higher velocities for specific situations. In addition, they are often conceptualized to be distributed. There are different types of NoSQL, with different ways of saving information: Apache CouchDB or MongoDB. Time-series databases are designed to store events or measurements that arrive continuously marked by a moment in time. They are optimized for real-time read and write to apply filters and aggregates on the timestamp. Generally, the timestamp acts as a key and is associated with key-value pairs. Their main applications are real-time system performance monitoring, stream analysis and real-time sensor monitoring. Examples of this technology are InfluxDB or OpenTSDB. Search engines provide amplified capabilities to perform complex searches and aggregations among data. For this purpose, they index a more significant number of fields than traditional databases. The most used technologies in this area are Apache Solr and ElasticSearch. Data processing and analysis provides the functionality of large-scale data processing. In **URBANAGE** we study using Apache Spark. It is an opensource, general-purpose analytics engine with multiple connectors to other architecture fragments that allow for maximum elasticity of the proposed solution. Multiple supported high-level APIs (Java, Scala, Python, R) will enable developers to use the adequate programming stack to develop algorithms (including Machine Learning) necessary for the project. Relevant languages to specify AI, analytical models, and their life-cycle will be explored.

The Presentation layer will be based on the most suitable techniques and recommendations of the ISO. It will be adopted and applied to the design and implementation of the system's User Interfaces.

The Security layer as mentioned before, will be based on the most suitable security guidelines. Those are issued by ENISA and will be implemented in compliance with the ETSI regulation. They will be used to design and implement the securitization, authentication and authorization mechanisms. Some potential solutions as Keycloak and Keyrock are compliant.

5.2 URBANAGE and extending existing standards and data standardisation related tasks

This section describes standards and data standardisation related tasks that are interesting but presumably only fit partial. Based on needs and memberships, it will be elaborated if and how the consortium must or can play a role in the standardisation process. Starting from participation in practice also EU standardisation supporting mechanisms are discussed in this section.

5.2.1 Participating to standardisation activities and cooperation with SDO’s

5.2.1.1 Participation in practice

SDO's like W3C and OGC have a series of working groups open for members to participate. The working groups start from a charter describing the scope, goals, activities, milestones, deliverables, communication, decision processes, patents and licenses.
Depending on the type of activities (knowledge exchange, discussion or standard developments), working groups are open for all members or only open for active contributors amongst the members. The latter is mostly the case once the standard core writing process is in progress.
The period between setting up a new standardisation working group or starting a new standardisation initiative in an existing working group and the publication of a finalised and excepted standard by the SDO community is difficult to achieve during a three-year H2020 project.
Participation in standardisation activities can only be achieved by the commitment of a consortium partner who is a member and actively participating in an SDO like W3C, ISO or OGC.
Contributing to data standardisation activities like a new NGSI-LD (thematic) data model is less formal and less time consuming compared to the SDO processes. Fiware, for example, has two application forms, one

for new contributors and one for existing contributors. The Fiware technical steering committee manages the Fiware technologies. The core principle of the committee's work is to maintain the open nature of the undertaken activities. The TSC, therefore, strives to have inclusive policies around FIWARE, contributing to its growth and wider adoption. Generally speaking, participating in standardisation-related organisations' working group activities is less formal and probably less impactful.

5.2.1.2 EU support to standardisation activities

The StandICT.EU "Supporting European Experts Presence in International Standardisation Activities in ICT" is an initiative that aims to support European Experts contributing in standardisation initiatives. The initiative started in 2018 and is extended to 2023 as a CSA (Communication Support Action). StandICT uses open thematic calls. The upcoming calls are the 4th open call "Digital, Industry and Space" – Sept 2021 and the 5th open call "Theme 5 Climate, Energy and Mobility". [59]

5.2.2 Overview of standards and standard related activities to extend/improve

This section offers a preliminary and non-exhaustive overview of the standards for which the project could investigate possible extensions and/or changes. Table 2 summarises the main actions that could be investigated during the project.

Standard	Organisation (owner)	Need for extension or change
NGSI-LD	ETSI	Contributing to existing of new NGSI-LD Fiware supported smart data models (see: https://fiware-datamodels.readthedocs.io/en/latest/index.html)
SAREF	ETSI	Contributing to existing extensions of SAREF Ontology (such as SAREF4CITY and SAREF4BLDG) by revising them and proposing possible new elements.
CityGML	OGC	Investigate a possible extension of CityGML to cover the representation and interaction of all the elements in the four layers of the **URBANAGE** CIM (City Information Model).
GeoDCAT-AP	GeoDCAT-AP Working Group;	Investigate improvements of the mapping from ISO based standards (e.g. INSPIRE)
Open MI	OGC/OASC	OGC® Open Modelling Interface (OpenMI) Interface Standard | OGC (Follow-up of openMI in the context of emerging technologies)
OSM	OpenStreetMap Foundation	Analyse a possible extension of OSM format to manage accessibility details

Table 2: Overview of standards and data standardisation tasks that might to be extended of changed

Information reported in this section represents the initial aiming of the project for what concerns potential activities related to contribution to existing standards. This list of activities will be revised during the project, also according to the reached technical development and advancements.

5.2.3 Overview of standardisation organisation and related organisations memberships

The table below provide an overview of the membership of the **URBANAGE** partners in SDO's and the way they are currently active. The lowest activity listed is "membership", which means the organisation is not taking part to any standardisation related activities.

Partner	SDO membership	Role
AIV	W3C	Member of W3C
AIV	OGC	Coordinator Metadata and catalogue DWG
AIV	ISA²	Active member – Influencing ISA² via the OSLO programme
ENG	FIWARE	Platinum member FIWARE (ENG is one the founding members).
ENG	BDVA	Full member and member of the Board of Directors. ENG leads the Smart Cities group, the Smart Manufacturing Industry group and the Security group.
FVH	OASC	FVH via the city of Helsinki
SANT	OASC	Member of the OASC
ATC	BDVA	Member of the BDVA
ATC, TEC	NESSI http://www.nessi-europe.com/default.aspx?Page=home	Member of the Networked European Software and Services Initiative (http://www.nessi-europe.com/)
IMEC	DTC	Member of the Digital Twin Consortium
IMEC	IDSE	Member of International Dataspaces
TEC	ENISA	Co-chair of the CSPCERT (European Cloud Service Provider Certification Working Group)

Table 3: SDO Membership

6 Conclusion

This report emphasises the availability of standards in Digital Twin related domains like big data, metadata, AI, ICT security, e.g. There are many relevant organisations in the broader field of ICT standardisation.

The **URBANAGE** consortium embraces standards because it believes that innovation also lies in how disruptive technologies can only cooperate successfully when they cooperate in a well-designed way, implying the re-use of standards.

But, despite the abundance of standards in many fields, there is no set of standards or mature, standardised architecture dealing with Digital City Twins. This is no surprise when you take into account that a Digital City Twin can be considered as a kind of a Swiss knife (a collection of tools in its own right). For some of the tools, e.g. a (BIM) and visualisation standard (e.g. city GML) standard exist. For other parts of the Swiss Knife, like interacting simulation models, no actively used standards (e.g. OGC OpenMI hasn't made any progress since 2014) exists.

Core standards are only one part of the puzzle. But also, the interaction and interoperability between datasets heavily influence a Digital City Twin richness. W3C principles of the semantic web play an important role. For example, describing the envelope where messages are sent is often the subject of a core standard, but it is the envelope's content that provides the real richness of the Urban Twin. Initiatives like Fiware (NGSI-LD) contribute to setting up ontologies and data interoperability. Also, the EU with ISA² and Inspire play an important role. The added value of **URBANAGE** can be to test these data ontologies and interoperability in a complex smart city setting.

URBANAGE will also be successful when creating best practices based on existing ontologies and standards. As an example, **URBANAGE** implementation of the OASC MIM's to show interoperability in practice can be one of the desirable outcomes. By implementing the MIM's in practice, a network of more than 100 front-running cities can be reached and inspired. Participating in new or improving standards is another option. Still, it will probably overarch the project's duration and align with the long-term objectives of one or more **URBANAGE** partners.

Based on the first analysis of the **URBANAGE** pilots, the existing Digital City Twin solution will combine the need for 3D city models (BIM-based), data visualisation technologies and AI-related predictive algorithms for long-term predictions combining personal data and city-data. Also, metadata standards will play an important role as part of a data catalogue together with the underlying data ontologies.

Based on the experiences of **URBANAGE** partner's with the long and difficult to assess SDO standardisation processes and initiatives that often exceed the project duration of EU funded projects, a long-term commitment of the partners participating in a (new) standardisation initiative is necessary. In that context,

URBANAGE will focus on the core standards itself and more data and ontology related standards (e.g., NGSI-LD) and the implementation of concepts like the OGC MIM's (Minimal Interoperability Mechanisms).

This report outlines the different standardisation possibilities and changes while developing the disruptive and innovative Digital City Twin solution directed to older adults. In cooperation with standardisation organisations and other EU Digital Urban Twin Initiatives like destination earth, Living-in.EU, Gaia-X and sister projects like DUET [60], **URBANAGE** will design, test and contribute to the Digital City Twin ecosystem in Europe. The journey will be part of the D7.7 standardisation Plan and Report final version.

7 References

[1] „W3C - Web of Things," [Online]. Available: https://www.w3.org/WoT/. [Geopend July 2021].

[2] „OGC standards - CityGML," [Online]. Available: https://www.ogc.org/standards/citygml. [Geopend July 2021].

[3] „IndoorGML - OGC Standard for Indoor Spatial Information," [Online]. Available: http://www.indoorgml.net/. [Geopend July 2021].

[4] R. Buyle, „OSLO: Open Standards for Linked Organizations," *Proceedings of the International Conference on Electronic Governance and Open Society,* p. 126–134, 2016.

[5] „GAIA-X - Home," [Online]. Available: https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html. [Geopend July 2021].

[6] „International Data Spaces | The future of the data economy is here," [Online]. Available: https://internationaldataspaces.org/. [Geopend July 2021].

[7] „Context Broker (europa.eu)," [Online]. Available: https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Context+Broker. [Geopend July 2021].

[8] „OpenTSDB - A Distributed, Scalable Monitoring System," [Online]. Available: https://opentsdb.net/. [Geopend July 2021].

[9] „National Information Standards Organization | NISO website," [Online]. Available: http://www.niso.org/. [Geopend July 2021].

[10] „Dublin Core - Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Dublin_Core. [Geopend July 2021].

[11] „ISO - ISO 15836:2009 - Information and documentation - The Dublin Core metadata element set," [Online]. Available: https://www.iso.org/standard/52142.html. [Geopend July 2021].

[12] „DCAT-AP/releases/2.0.1 at master - SEMICeu/DCAT-AP - GitHub," [Online]. Available: https://github.com/SEMICeu/DCAT-AP/tree/master/releases/2.0.1. [Geopend July 2021].

[13] „The official portal for European data," [Online]. Available: https://data.europa.eu. [Geopend 2021 July].

[14] „EUR-Lex - 32019L1024 - EN - EUR-Lex (europa.eu)," [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG. [Geopend July 2021].

[15] „DCAT-AP.de - Start," [Online]. Available: https://www.dcat-ap.de/. [Geopend July 2021].

[16] „DCAT-AP-DK 2.0.2: Dansk anvendelsesprofil til beskrivelse af datasæt og datakataloger (digst.github.io)," [Online]. Available: https://digst.github.io/DCAT-AP-DK/releases/v.2.0/docs/. [Geopend July 2021].

[17] „DCAT-AP Vlaanderen (vlaanderen.be)," [Online]. Available: https://data.vlaanderen.be/doc/applicatieprofiel/DCAT-AP-VL/. [Geopend July 2021].

[18] „Metadata DCAT (vlaanderen.be)," [Online]. Available: https://data.vlaanderen.be/doc/applicatieprofiel/metadata-dcat. [Geopend July 2021].

[19] „GeoDCAT-AP - Version 2.0.0 (semiceu.github.io),” [Online]. Available: https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/. [Geopend July 2021].

[20] „ISO - ISO 19119:2005 - Geographic information - Services,” [Online]. Available: https://www.iso.org/standard/39890.html. [Geopend July 2021].

[21] „Inspire - download page,” [Online]. Available: http://inspire.ec.europa.eu/file/1705/download?token=iSTwpRWd. [Geopend July 2021].

[22] „ISO - ISO 19110:2016 - Geographic information - Methodology for feature cataloguing,” [Online]. Available: https://www.iso.org/standard/57303.html. [Geopend July 2021].

[23] „ISO - ISO/TS 19139:2007 - Geographic information - Metadata - XML schema implementation,” [Online]. Available: https://www.iso.org/standard/32557.html. [Geopend July 2021].

[24] „INSPIRE Roadmap | INSPIRE (europa.eu),” [Online]. Available: http://inspire.ec.europa.eu/inspire-roadmap/61. [Geopend July 2021].

[25] „About INSPIRE | INSPIRE (europa.eu),” [Online]. Available: http://inspire.ec.europa.eu/about-inspire/563. [Geopend July 2021].

[26] „INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119,” [Online]. Available: http://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf.

[27] „COMMISSION REGULATION (EU) No 1253/2013 of 21 October 2013,” [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ%3AL%3A2013%3A331%3A0001%3A0267%3AEN%3APDF . [Geopend July 2021].

[28] „INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119,” [Online]. Available: http://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf. [Geopend July 2021].

[29] „CKAN - The open source data management system,” [Online]. Available: https://ckan.org/. [Geopend July 2021].

[30] „Home - GeoNetwork opensource (geonetwork-opensource.org),” [Online]. Available: https://geonetwork-opensource.org/. [Geopend July 2021].

[31] „Home - GeoNetwork opensource (geonetwork-opensource.org),” [Online]. Available: https://geonetwork-opensource.org/. [Geopend July 2021].

[32] „Data Catalog Vocabulary (DCAT) - Version 2 (w3.org),” [Online]. Available: https://www.w3.org/TR/vocab-dcat-2/. [Geopend July 2021].

[33] „OpenLink Virtuoso SPARQL Query Editor (europa.eu),” [Online]. Available: https://data.europa.eu/sparql. [Geopend July 2021].

[34] „Open Micka - Open Micka is a web application for management and discovery geospatial metadata. (bnhelp.cz),” [Online]. Available: https://micka.bnhelp.cz/. [Geopend July 2021].

[35] „Data Mining Group (dmg.org),” [Online]. Available: http://dmg.org/ . [Geopend July 2021].

[36] „PMML 4.1 - General Structure (dmg.org),” [Online]. Available: http://dmg.org/pmml/v4-1/GeneralStructure.html . [Geopend July 2021].

[37] „PFA · Portable Format for Analytics (dmg.org)," [Online]. Available: http://dmg.org/pfa/. [Geopend July 2021].

[38] „ONNX | Home," [Online]. Available: https://onnx.ai/. [Geopend July 2021].

[39] „NNEF Overview - The Khronos Group Inc," [Online]. Available: https://www.khronos.org/nnef. [Geopend July 2021].

[40] „MLlib | Apache Spark," [Online]. Available: https://spark.apache.org/mllib/. [Geopend July 2021].

[41] „AIML Foundation," [Online]. Available: http://www.aiml.foundation/. [Geopend July 2021].

[42] „IEEE SA - AIS Standards," [Online]. Available: https://standards.ieee.org/initiatives/artificial-intelligence-systems/standards.html. [Geopend July 2021].

[43] „AI Standards | NIST," [Online]. Available: https://www.nist.gov/artificial-intelligence/ai-standards. [Geopend July 2021].

[44] „Federal Register - Maintaining American Leadership in Artificial Intelligence," [Online]. Available: https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence. [Geopend July 2021].

[45] „NIST Big Data Working Group (NBD-WG)," [Online]. Available: https://bigdatawg.nist.gov/V3_output_docs.php. [Geopend July 2021].

[46] „Home - IEEE Big Data," [Online]. Available: https://bigdata.ieee.org/. [Geopend July 2021].

[47] „NIST Big Data Interoperability Framework: Volume 6, Reference Architecture," [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-6r2.pdf. [Geopend July 2021].

[48] „ISO - ISO/IEC JTC 1 - Information Technology," [Online]. Available: https://www.iso.org/isoiec-jtc-1.html. [Geopend July 2021].

[49] „rfc6749 (ietf.org)," [Online]. Available: IETF RFC 6749-2012: https://datatracker.ietf.org/doc/html/rfc6749.txt. [Geopend July 2021].

[50] „SAML 2.0: Cantor et al. Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0 - Errata Composite. Working Draft 07, 8 September 2015. Document ID sstc-saml-core-errata-2.0-wd-07," [Online]. Available: http://www.oasis-open.org/committees/download.php/56776/sstc-saml-core-errata-2.0-wd-07.pdf. [Geopend July 2021].

[51] „The OpenID Foundation Launches the OpenID Connect Standard | OpenID," [Online]. Available: https://openid.net/2014/02/26/the-openid-foundation-launches-the-openid-connect-standard/ . [Geopend July 2021].

[52] „JWT - JSON Web Tokens," [Online]. Available: https://jwt.io/. [Geopend July 2021].

[53] „JWT RFC 7519-2015," [Online]. Available: https://datatracker.ietf.org/doc/html/rfc7519. [Geopend July 2021].

[54] „Keycloak," [Online]. Available: https://www.keycloak.org/. [Geopend July 2021].

[55] „GitHub - FIWARE/tutorials.Identity-Management: FIWARE 401: IDM - Managing Users and Organizations," [Online]. Available: https://github.com/FIWARE/tutorials.Identity-Management. [Geopend July 2021].

[56] „CityGML Wiki," [Online]. Available: https://www.citygmlwiki.org/index.php?title=Citygml_Wiki. [Geopend July 2021].

[57] „IndoorGML OGC," [Online]. Available: http://www.indoorgml.net/. [Geopend July 2021].

[58] „ISO - ISO 9241-210:2019 - Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems," [Online]. Available: https://www.iso.org/standard/77520.html. [Geopend July 2021].

[59] „StandICT calls - Relation with Strategic Plan for Horizon Europe | StandICT.eu 2023," [Online]. Available: https://www.standict.eu/standict-calls-relation-strategic-plan-horizon-europe. [Geopend July 2021].

[60] „Home | digital urban european twins (digitalurbantwins.com)," [Online]. Available: https://www.digitalurbantwins.com/. [Geopend July 2021].

[61] „AI Standards | NIST," [Online]. Available: https://www.nist.gov/artificial-intelligence/ai-standards. [Geopend July 2021].