Publications Office

Directorate A – Information Management

Data Structures and Standards used at the Publications Office

Version: 4.0.1

DOCUMENT HISTORY			
Version	**Release Date**	**Author**	**Description**
0.0.1	2016-05-02	Najeh Hajlaoui (NH)	Creation
0.0.2	2016-05-04	NH	Comments from Holger Bagola (HB) and Claude Schmit (CS)
0.0.3	2016-05-04	NH	Comments from June Lowery-Kingston (JLK)
0.0.4	2016-06-01	NH	Comments from HB, CS and Corinne Frappart (CF)
0.0.5	2016-06-17	NH	Comments from HB, CS and CF
0.0.6	2016-07-06	NH	Comments from HB
0.0.7	2016-07-20	NH	Improvement
0.0.8	2016-07-26	NH	Comments from HB, CS and CF, adding INVEX (input from Vivian Boyer), and OP Core Metadata
0.0.9	2016-08-01	NH	Improvement, reviewed by CF
0.0.10	2016-08-29	NH	Improvement, adding XForms standard
0.0.11	2016-09-13	NH	Comments from CF
0.0.12	2016-10-10	NH	Comments from B1, B2, C1, C2, C3, C4
0.0.13	2016-10-15	NH	Improvements, comments from CS and CF
0.0.14	2016-12-06	NH	adding examples, improvements
1.0.0	2016-12-19	NH	Publication
1.0.1	2017-01-06	NH	Improvement, adding Premis, OAIS (input from Vasiliki Bountouri), TOGAF eProcurement ontology.
2.0.0	2018-08-01	NH	Publication
2.0.1	2019-10-22	NH	Updating contacts, versions, links and annex,
2.0.2	2019-10-28	NH	Updates following the grouping of Metadata Registry and EuroVoc in EU Vocabularies website, improvements.
3.0.0	2019-10-29	NH	Publication
3.0.1	2019-11-06	NH	Improvements, updates
3.0.2	2019-12-09	NH	Improvements
4.0.0	2019-12-09	NH	Publication
4.0.1	2020-01-08	NH	Updating contact

Remarks

- We will keep up to date this document. An updated version is (will be) available online.

- If links are not publically accessible, please contact the Publications Office.

- Please do not hesitate to send your comments/remarks to OP-INFO-HELPDESK@publications.europa.eu.

Contents

1.   Introduction

1.1   Context

The Publications Office of the European Union (OP) is the publishing house of the European institutions and bodies. Its main mission is to publish and disseminate all EU publications. The OP is working to offer both open data and linked data, to improve machine-readable access to its vast collections hosted in Cellar (Official Journal, Case law, General Publications of EU institutions, etc.). Cellar[1] is the storage for all oeuvres published by the Publications Office and institutions and bodies.

OP Portal provides an easy-to-use main point of access to all the content managed by the Office (law, publications, procurement notices, contacts, vocabularies) as well as corporate information about the Publications Office.   It also provides OP with a modern single and common dissemination platform. OP Portal consolidates also a single point of entry to all specialized services like EUR-Lex (European Union Law in the 23 official EU languages (24 when Irish is required)), EU Publications (an online library and bookshop of publications from the institutions and other bodies of the European Union), EU Open Data Portal, TED (online information European public procurement), EU Whoiswho (the official directory of the European Union), CORDIS (the European Commission's primary public repository and portal used to disseminate information on all EU-funded research projects and their results in the broadest sense).

The EU Open Data Portal is a single point of access to EU data published by EU institutions, agencies and other bodies in line with the EU open data strategy and reuse principles. It allows to search, explore, link, download and reuse data for free. Ways of querying the EU Open Data Portal via application interfaces can be consulted at http://data.europa.eu/euodp/en/developerscorner.

The contents of the EU Open Data Portal can also be accessed via the larger European Data Portal http://www.europeandataportal.eu/. This project harvests the metadata of public data made available by public services (including the EU itself) across the 28 EU Member States, the European Economic Area, countries involved in the EU's neighbourhood policy and Switzerland. It also has a SPARQL endpoint.

The EU Vocabularies website is the reference data repository of the Publications Office of the European Union. It provides access to data models, controlled vocabularies and other interoperability solutions mentioned in this document.

According to the Digital Agenda for Europe, interoperability between public administrations is a challenge, hindering the provision of digital public services across borders and across sectors. Still today public administrations and EU institutions continue to use standards which are often based on non-interoperable solutions.

To avoid this redundancy and this effort duplication, we would like to facilitate the interoperability promoting the reuse of solutions approved at the Publications Office.

---

[1] Cellar: Information System storing all metadata and digital content managed by the Publications Office of the European Union in a harmonised and standardised format, accessible by SPARQL endpoint.

1.2   Aim of the document

In order to facilitate interoperability between public administrations, this document proposes an overview of standards used at the OP such as knowledge organisation systems, semantic technologies, data models and information technologies, etc.

The aim of the document is to optimise efforts collaboration between the OP and its contractors promoting the reuse and the sharing of interoperable standards and data structures used at the OP.

These standards are already approved and strongly recommended to be reused. Consequently, all contractors are expected to apply them avoiding to reintroduce other standards where not appropriate, however, OP remains open with authors and contractors in the frame of technology evolution (e.g. multimedia production).

1.3   Scope

This document summarises useful information to facilitate the market acceptance and the coherence criteria providing useful links to the tenderers to keep interoperability between services, interfaces, applications and products. It provides documentary information of standards used at the OP to be reused for a better normalisation of OP requirements avoiding reinvestigation.

The Publications Office of the European Union has already published an existing document which provides an overview of its technical environment with some general rules linked to the technical organisation of the Publications Office and applicable to all applications hosted at the Publications Office.

For better reusability and coherence of Information Technology applications and software used at the Publications Office, more details can be found in the annexed document (OP_IT_Technical_Annex.pdf).

Consequently, to avoid redundancy, the OP would like to exclude from the scope of this document information about the recommended environment, architecture required at the OP since these are already described in the document mentioned above.

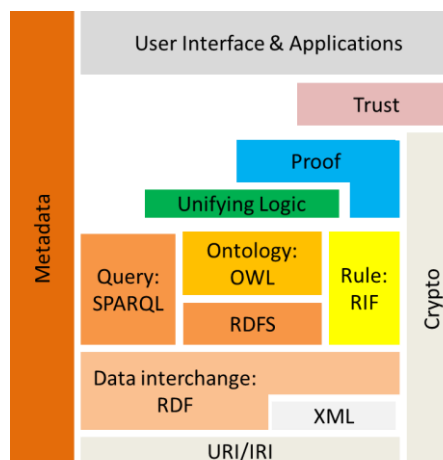The Publications Office applies semantic web technologies.

Figure 1: Semantic Web

According to the semantic web standards described in Figure 1, in section 2 we describe the standard identifiers used at the Publications Office. In section 3 we present XML technologies used at the Publications Office. In section 4 we describe the metadata. In section 5 we present technologies and data models. In section 6 we give an overview of the standards in output formats used at the Publications Office. Section 0 describes some other frameworks. In section 8 we provide useful links and in section 9 we present some ongoing standards whose development/implementation is in progress.

2.   Standard Identifiers

The Publications Office assigns the adequate identifiers listed below to each publication request issued by European Union institutions, bodies and organisations to classify publications all over the world in a unique and exclusive way.

The Publications Office is in charge of assigning international identifiers[2] such as ISBN, ISSN, DOI and ELI, and also internal identifiers such as the unique number for acts published in the Official Journal, catalogue number, Celex number and URI/IRI to EU publications according to the type of publication. The Publications Office is also responsible of creating and managing the bibliographical data associated to these identifiers.

The Publications Office assures the secretariat of the interinstitutional URI Committee and runs a redirection service under the data.europa.eu subdomain for resources from the EU institutions with persistent URIs. Examples are the European Legislation Identifier (ELI), the DOI, different ISA related resources, etc.[3]

2.1   URI/IRI

A Uniform Resource Identifier (URI) is a string of characters used to identify a resource. It enables interaction with representations of the resource over a network such as the World Wide Web, using specific protocols. The most common form of URI is the Uniform Resource Locator (URL), frequently used as a web address. A complementary form is the Uniform Resource Name (URN) used to provide a mechanism for the identification of resources in particular namespaces.

Uniform Resource Identifier (URIs) can be dereferenced using the HTTP protocol according to the Linked Data principles, such a dereferenced URI should result in a document that offers further data about the given URI.

The Internationalised Resource Identifier (IRI) was defined to generalise the existing Uniform Resource Identifier (URI) scheme. It was created by the Internet Engineering Task Force (IETF) in 2005 as a new internet standard published in RFC 3987. IRIs are not limited only to a subset of the ASCII character set as URIs, they may contain characters from the Universal Character Set (Unicode/ISO 10646), including Chinese or Japanese kanji, Korean, Cyrillic characters and so forth.

---

[2] ECLI is also an international identifier managed by the Publications Office, but it is assigned by the Court of Justice for the European case-law.

[3] http://data.europa.eu.

At the Publications Office, each resource in the Cellar[4] is globally identified by a URI composed as follows: http://publications.europa.eu/resource/{ps-name}/{ps-id} where ps-name identifies the name of the production system and ps-id is the resource's unique identifier and it has a structure that depends on the value of {ps-name}.

Some examples of resource URIs:

1. The following resource URI identifies a work with ps-name of type *cellar* and the given ps-id:

    http://publications.europa.eu/resource/cellar/b84f49cd-750f-11e3-8e20-01aa75ed71a1

2. The following resource URI identifies a work with ps-name of type *oj* and the given ps-id:

    http://publications.europa.eu/resource/oj/JOL_2014_001_R_0001_01

3. The following resource URI identifies a work with ps-name of type *celex* and the given ps-id:

    http://publications.europa.eu/resource/celex/32014R0001

Name	Status	Version(s)
URI/IRI	In production	Not specified

2.2 Celex

The Celex number is a unique identifier assigned by the Publications Office to documents on EUR-Lex, including the documents published in the Official Journal of the European Union and by the Court of Justice, and also treaties and preparatory acts. The Celex identifier is independent of the language(s) of the document. More details can be found in the following link: http://eur-lex.europa.eu/content/help/faq/intro.html#top.

Name	Status	Version(s)	Example
Celex	In production		32016R2104

2.3 Catalogue number

The Publications Office uses an internal identifier called catalogue number in addition to the international identifiers. This catalogue number is managed by an internal management tool used for physical distribution purposes and as identification key in various IT applications, as is the case for EU Publications, the European institutions' online library and bookshop.

For each product (book, leaflet, poster, etc.), the Publications Office assigns a catalogue number before the physical distribution of the publication and through EU Publications. The catalogue number is generally printed on the back cover, usually in the upper right-hand corner unless for technical reasons it has to be printed elsewhere.

---

[4]The Cellar currently uses the following production system names: cellar, celex, oj, com, genpub, ep, jurisprudence, dd, mtf, consolidation, eurostat, eesc, cor, nim, pegase, transjai, agent, uriserv, join, swd, comnat, mdr, legissum, ecli, procedure, procedure-event, eli, immc and planjo.

More details on the use of catalogue number at the Publications Office can be found in the following link: http://publications.europa.eu/code/en/en-240400.htm.

Name	Status	Version(s)	Example
Catalogue number	In production	Not applicable	OA-AS-16-001-EN-N

2.4   ISBN

The International Standard Book Number (ISBN) is a unique numeric identifier used in the book supply chain. It is assigned to each edition and variation of a book. Since 2007 the ISBN configuration consists of 13-digits.

The Publications Office uses ISBN number for any monographic. An ISBN cannot be modified, replaced or reused; a new ISBN is required for each separate format of a product (pdf, html, etc.) to each language version of that publication, and a new one should be issued for new editions and for any significant amendment to the form of a product, etc.

An ISBN is assigned to each volume in case of multiple volumes and a collective ISBN is assigned to the set of volumes. The collective ISBN and the ISBN of the respective volumes must be printed on the reverse of the title page of each volume.

More details on the use of ISBN at the Publications Office can be found in the following link: http://publications.europa.eu/code/en/en-240400.htm.

Name	Status	Version(s)	Example
ISBN	In production	Not applicable	ISBN: 978-92-78-41349-1

2.5   ISSN

The International Standard Serial Number (ISSN) is an eight-digit serial number used to identify serial publications such as newspapers, journals, magazines, all media-print and electronic periodical publications, etc. The ISSN is especially helpful in ordering, cataloging, interlibrary loans and other practices in connection with serial literature.

The ISSN system was first drafted as an International Organisation for Standardisation (ISO) international standard in 1971 and published as ISO 3297 in 1975.

The Publications Office is using ISSN for continuing resources (serial publications and ongoing integrating resources). The allocation of an ISSN has no legal meaning or value with regard to the copyright of the work concerned or its content.

An ISSN is assigned exclusively for the entire lifespan of a title, each language version, each edition (monthly, annual, etc.) and each separate format.

An ISSN can also be assigned to a collection of monographs, with an ISBN being assigned to each volume in the collection.

More details can be found in the following link: http://publications.europa.eu/code/en/en-240400.htm.

Name	Status	Version(s)	Example
ISSN	In production	Not applicable	ISSN 1977-0693

2.6 DOI

In 2000 [the Digital Object Identifier](#) (DOI) was created as a digital identifier used to uniquely identify objects such as journal articles, entire publication, photo, table, chapter, etc. The DOI system is implemented through a federation of registration agencies coordinated by the International DOI Foundation which developed and oversees the application of the ISO standard.

The DOI name is associated to the metadata and includes a location, such as a URL, where the object can be found or where information related to the object is displayed. It can be expressed as a URI, e.g. http://data.europa.eu/doi/10.2777/741268.

The DOI for a document remains fixed for its entire lifetime, while the metadata may change over time. For instance, the URL may change to a new one. Each DOI is unique and permanent, if ever the document is deleted; its DOI will not be reused but should continue to point to information on this object giving a reason as to why the object is no longer available.

The DOI number is composed of a prefix and a suffix, separated by a forward slash.

The Publications Office is using DOI system to identify products such as a complete publication, a photo, a table, a chapter, etc. More details can be found in the following link: http://publications.europa.eu/code/en/en-240400.htm.

Name	Status	Version(s)	Example
DOI	In production	Not applicable	DOI: 10.2830/339728

2.7 ECLI

The European Case-Law Identifier (ECLI) has been developed to facilitate the correct and unequivocal citation of decisions from European and national courts. It was introduced progressively by the Court of Justice of the European Union in 2014. It is assigned by the Court of Justice of the European Union and by each Member State to its own Case-Law.

The Court of Justice assigned an ECLI number to all decisions delivered by the European Union courts since 1954 (judgments, orders, opinions of the Court, decisions (review procedure)), to the opinions and views of the advocates general and to the information about these decisions (summaries and reviews, information on unpublished decisions).

More details can be found in the following links http://publications.europa.eu/code/en/en-240500.htm and http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV%3Ajl0056.

Name	Status	Version(s)	Example
ECLI	In production		ECLI:EU:C:2016:912

2.8   ELI

The European Legislative Identifier (ELI) is a system to make legislation available online in a standardised format, so that it can be accessed, exchanged and reused across borders. ELI facilitates a recognisable, readable and understandable identification by both humans and computers.

The Publications Office uses ELI, a system based on semantic technology to facilitate interoperability in the context of legal information and to interconnect legislative documents.

ELI was implemented at the Publications Office in three steps:

- The first step consisted in assigning in 2015 URIs to all acts published in the Official Journal, series L, including directives, regulations, decisions, international agreements, as well as treaties and consolidated texts of legal acts.

- The second step consisted in applying ELI metadata to these legal acts in a standard structure to make this information machine-reusable.

- The third step consisted in publishing the http ELI URIs and machine-readable ELI metadata through the EUR-Lex portal in March 2016.

The Publications Office has reached an important milestone by building a tailored information structure based on ELI. It makes easier to find legal acts published on EUR-Lex in more harmonised way facilitating the exchange, the access and the reuse of legal information across borders.

ELI is sufficiently standardised to provide interoperability of legal data, while being flexible enough to adapt to other systems and legal traditions.

More information about this identifier and the countries who have adopted it can be found on the following link: http://eur-lex.europa.eu/eli-register/about.html.

Name	Status	Version(s)
ELI	In production	1.0

2.9   Unique number of acts published in the Official Journal

Acts published in the Official Journal receive a unique number by the OP. The number of an act is made up of three parts, presented in the following order:

— the abbreviation for the domain placed within brackets ('EU' for the European Union, 'Euratom' for the European Atomic Energy Community, 'EU, Euratom' for the European Union and the European Atomic Energy Community, 'CFSP' for the common foreign and security policy),

— a reference to the year of publication, comprising four digits,

— the sequential number, assigned from one single series, irrespective of the type and the domain of the act, and comprised of as many digits as necessary.

(domain) YYYY/N

More details can be found in the following link: http://publications.europa.eu/code/en/en-110202.htm

Name	Status	Version(s)	Example
OJ Act Unique number	In production	Not applicable	(EU) 2015/299

3.   XML

The Publications Office provides various documents harmonised in XML format such as the Official Journal, Case Law, General Publications, public procurement notices, etc. of the European Union. All of these documents are converted in XML formats. Some XML standards are already defined to be reused at the OP such as FORMEX, BITS, etc.

3.1   XML standard

Extensible Markup Language (XML) is a markup language that defines formats to encode documents. XML was designed to be both human- and machine-readable. It is defined by the W3C's XML 1.0 Specification and by several other related standard specifications.

In 2001, the Publications Office decided the migration from SGML to XML as a new and compatible standard accompanied by additional standards which supported the navigation and the transformation of documents.

Name	Status	Version(s)
XML	In production	1.0/1.1

3.2   XML Schema

XML Schema is a recommendation of the World Wide Web Consortium (W3C) published as a W3C standard in May 2001. It is one of several XML schema languages. XML Schema describes the structure of an XML document specifying its elements. It is a set of rules to which an XML document must conform to be considered valid.

The first version XSD 1.0 was originally specified and published in 2001, with a second edition following in 2004 to correct large number of errors. XSD 1.1 became a W3C Recommendation in April 2012.

Since the new standard for the specification of XML grammars was adopted in 2001, the Publications Office adopted the replacement of the DTD by XML Schema. In the sections below, we will give more details about the XML schema used at the Publications Office such as FORMEX, BITS, etc.

Name	Status	Version(s)
XSD	In production	1.0/1.1

3.3   Schematron

Schematron is a rule-based language used for the validation of XML document structure. It makes assertions about the presence or absence of patterns in XML trees.

The Publications Office is using Schematron for the validation of the XML automated production mainly in the Metadata Registry published in EU Vocabularies: Controlled vocabularies (Authority tables, Thesauri, ATTO tables, Alignments), Models (Schemas, Ontologies, Presentation style sheets, Application profiles), etc.

Name	Status	Version(s)
Schematron	In production	

3.4   XSLT

Extensible Stylesheet Language Transformations (XSLT) is a style sheet language for XML documents to transform them into other XML documents or other formats such as HTML, plain text, etc. XSLT keeps the original XML documents as it is and creates a new document based on the content of the original one. XSLT uses XPath for the identification of subsets of the source document tree (see next section).

The Publications Office is using version 2.0 and version 3.0.

Name	Status	Version(s)
XSLT	In production	2.0/3.0

3.5   XPath

The XML Path Language (XPath) is a query language defined by the W3C and used to select nodes from an XML document and to compute their values. It allows the navigation around the XML document based on a tree representation. It has been adopted by several XML processing libraries and tools.

Name	Status	Version(s)
XPath	In production	1.0/2.0/3.0

3.6   FORMEX

The current Official Journal production flow does not use XML from the drafting until the publication stage. Therefore a contractor is in charge of marking-up the legislative texts in the [Formalized Exchange of Electronic Publications](#) (FORMEX) XML. As the use of some elements requires legal content analysis, the contractor submits to the OP services the question on the correct mark-up of the text in case of doubt.

FORMEX is the XML schema developed by the Publication Office. It describes the XML format to exchange data between the Publication Office and its contractors. It defines the logical description for documents which are published in the Official Journal (series L and C) of the European Union. The current version 4 is used since May 2004.

The actual version 4 is presented as an XML Schema grammar where character set definition are represented in Unicode (UTF-8) to encoding errors of special characters and symbols of different EU languages.

While the previous version (version 3) contains about 1200 tags, FORMEX version 4 starts with only about 260 tags. The future version (v5.55) contains 488 elements; it is an evolutive standard that progresses to fit the publication needs of European legislation.

More details about the physical specifications which contain information on the exchange of data, the construction of filenames and, in particular, on the character sets and the grammar for the mark-up based on XML Schema can be found in FORMEX website https://op.europa.eu/en/web/eu-vocabularies/formex .

Name	Status	Version(s)
FORMEX	In production	4
FORMEX	Under development	5

3.7   BITS

[Book Interchange Tag Suite](#) (BITS) provides a common XML format in which publishers and archives can exchange book content (entire book or a part of it such as chapter). It provides a set of XML DTD and schemas that define elements and attributes to describe textual and graphical contents.

In order to facilitate the reusability, BITS makes a tag set adequate for supporting interchange, archiving, format-conversion and publishing for scientific, technical and medical books.

The BITS Book Interchange DTD is a superset customisation of the ANSI/NISO JATS Z39.96-2012 Journal Archiving and Interchange Tag Set, with added material to describe STM books, book components such as chapters and information concerning the inclusion of books and book components in book series.

The BITS book model is intended to support scholarly, reference, higher education, medical and technical books. Just as the NISO JATS journal article models do not attempt to support magazines or any of a wide variety of other serial publications, the book models are not intended to describe trade

books, cook books, grade school text books, or any of the wide variety of books outside the scientific, technical and medical realm.

More details can be found here http://jats.nlm.nih.gov/extensions/bits/.

The Publications Office has prepared a customised version of BITS schema for General Publications. More details can be found in the following link:

https://op.europa.eu/en/web/eu-vocabularies/bits

Name	Status	Version(s)
BITS-book-01.07-20160204.xsd	Under evaluation	01.07

3.8   LegalDocML

Based on the Akoma Ntoso-UN project, the Advancing Open Standards for the Information Society (OASIS) works on the standardisation of LegalDocML for the use of XML standard for the specification of parliamentary, legislative and judicial documents, for their interchange between institutions.

Akoma Ntoso ("linked hearts" in the Akan language of West Africa) defines representations in XML format of parliamentary, legislative and judiciary documents. The XML schemas of Akoma Ntoso make explicit the structure and semantic components of the digital documents in parliamentary, legislative and judiciary contexts. More details can be found in the following links:

https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legaldocml
http://www.akomantoso.org/
https://op.europa.eu/en/web/eu-vocabularies/akn4eu

Name	Status	Version(s)
LegalDocML	Under evaluation	1.0

3.9   DITA

Darwin Information Typing Architecture (DITA) is an XML data model used for authoring and publishing. It is an open standard that is defined and maintained by the OASIS DITA Technical Committee.

At the Publications Office, DITA was proposed for adoption throughout 2018 and currently in production since 2019. It is used for two types of documentation: Editorial documentation projects (Style Guide Editorial Platform (SGEP)) and Data-driven documentation projects (Release notes (IMMC and NAL tables), Asset Description Documents (ADDs), SKOS-to-DITA (EuroVoc and NALs), XSD-to-DITA (Common Vocabulary, IMMC, and OJEEP XSDs); XMI-to-DITA (ontologies, application profiles and similar assets); and Formex documentation).

Name	Status	Version(s)
DITA	In production	1.3

3.10 TED schema

Tenders Electronic Daily (TED) is the online version of the EU Supplement to the Official Journal (OJS). It publishes public procurement notices that are eligible for publication in OJ S. Rules and conditions applied to publication are set by EC directives The OJ S contains notices on public contracts for works, supplies, and services from all the EU Member States and European Institutions/Agencies or International Organisations.

TED public procurement notices are shared by the Publications Office of the European Union in HTML, PDF and the XML format.

A set of standard form is defined by Regulation (EU) 2015/1986. It was transformed into a data structure schema of XML format. There are three individual versions of the schema to support all the main stages of the publication process – Reception (TED eSenders XML schema, used by eSenders, eNotices and eTendering), Production (Internal, used by the OP) and Publication delivery (XML schema published on the TED Website, used also by eTendering).

XML schema 2.0.9 is compliant with:

- Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC;
- Directive 2014/25/EU of the European Parliament and of the Council of 26 February 2014 on procurement by entities operating in the water, energy, transport and postal services sectors and repealing Directive 2004/17/EC;
- Directive 2014/23/EU of the European Parliament and of the Council of 26 February 2014 on the award of concession contracts.

In this context, a new release of the schema has been produced. It is identified as TED eSenders schema R2.0.9.S01.

Note: For notices covered by Directive 2009/81/EU (the defence directive — forms 16, 17, 18, 19) and Regulation (EC) No 1370/2007 (the public passengers transport services regulation — forms T01, T02), please refer to TED eSenders schema R2.0.8.S02.

Older version 2.0.8 is supported by TED because of its legal bindings with the "old" directives (see below). However, this version is planned for retirement and shall not constrain design of new solutions for eNotification domain.

XML schema 2.0.8 is compliant with:

- Directive 2009/81/EU (the defence directive — forms 16, 17, 18, 19);
- Regulation (EC) No 1370/2007 (the public passengers transport services regulation — forms T01, T02);
- Old directives :
  - Directive 2004/18/EC of the European Parliament and of the Council on public procurement;
  - Directive 2004/17/EC of the European Parliament and of the Council on procurement by entities operating in the water, energy, transport and postal services sectors.

Further details can be found at the following sites

- Tenders Electronic Daily – general information about eNotification
- TED eSenders Wiki (ECAS needed) – description of the standard
  - XML schema files
  - XML schema 2.0.9 description
  - XML schema 2.0.8 description
- FTP (guest/guest) – schemas R2.0.9.S01.E01 & R2.0.8.S02.E01 files / export versions, news, OJS packages, other resources supporting use of schemas
- TED Open Data Wiki (ECAS needed) – reuse cases for TED data (DG GROW)

Internal (production) versions are not available in extranet and OP shall be contacted to provide the respective schema files. An ongoing project shall make all the schemas available via EU Vocabularies website of the Publications Office.

Name	Status	Version(s)
TED	In production	2.0.9
TED	In production	2.0.8
UBL	Under evaluation	2.1

3.11 UBL

Universal Business Language (UBL) is a library of standard electronic XML business documents developed by an OASIS Technical Committee to be used for e-commerce. UBL is used by nations around the world for implementing cross-border transactions related to sourcing (e.g. tendering and catalogues), procurement (e.g. ordering and electronic invoicing), replenishment (e.g. managed inventory) and transportation (e.g. waybills, forwarding and status).

Following the recommend of the Multi-Stakeholder Platform on ICT Standardisation the UBL Standard has been designated by the European Commission as one of the first consortium standards officially eligible for referencing in tenders from public administrations. It is an electronic format to define and exchange business transaction documents, particularly in international supply chain processes. UBL provides the foundation for successful European public procurement frameworks such as ePrior (European Commission DIGIT). ePrior offers e-Procurement business services enabling the European Commission and Member States to exchange procurement documents in digital format with their suppliers.

Name	Status	Version(s)
UBL	In production	2.0/2.1

3.12 INVEX

INVEX is an XML format used at the Publications Office for the invoicing of OJ L&C, DD and Case Law. It is a multilevel invoice allowing detailed invoicing for the whole publication as well as for the underlying documents. It also includes all the production information (per document) making

additional supporting documents obsolete. It is used by the INVEX application for financial validation, Sicof for operational validation and Factelec2 for cost accounting.

Name	Status	Version(s)
INVEX	In production	1.0.4

4. Metadata

4.1 Standards

SKOS

Simple Knowledge Organisation System (SKOS) is a standard data model for sharing and linking knowledge organisation systems via the Semantic Web. It is a W3C standard which aims to enable an easy publication and use of such vocabularies as linked data. It can be used to represent thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary.

At the Publications Office, SKOS is used to represent and organise EuroVoc and Common Authority Tables (see section 4.3) to be machine-readable. The Publications Office provides also semantic alignments between EuroVoc and other international Thesauri (e.g Unbis, GEMET, INSPIRE etc.) in SKOS format.

On request, the Publications Office can also convert vocabularies of EU institutions and bodies existing in various non machine readable formats into SKOS. A SKOS version of EuroVoc is available for download using the following link https://op.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc .

Name	Status	Version(s)
SKOS	In production	2004/02
SKOSXL	In production	2008/05

METS

METS schema is a standard for encoding descriptive, administrative and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium.

The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress and is being developed as an initiative of the Digital Library Federation.

METS is used by the Publications Office in Cellar and Eudor[5] as ingestion protocol for General Information Packages. It allows the ingestion and retrieval of publication data maintained by the EU

---

[5] Eudor: long term digital archive of the OP.

Publications Office. For the first version (Cellar METS v1) a profile has been registered at the Library of Congress (LOC) while for the second version (Cellar METS v2) it has not yet been done. More information about the Cellar METS v1 can be found in the following link:

http://www.loc.gov/standards/mets/profiles/00000036.xml.

Name	Status	Version(s)
Cellar METS v1	In production (registered at the LOC)	Based on 1.9 of the official METS schema
Cellar METS v2	In production (not yet registered)	Based on 1.9 of the official METS schema

MARC 21

MAchine-Readable Cataloging (MARC) standards are a set of digital formats for the description of items catalogued by libraries, such as books. It is developed by Henriette Avram MARC in the 1960s to create records that could be read by computers and shared among libraries. MARC encodes information about a bibliographic item, it is a metadata transmission standard.

In 1971, MARC formats had become the national standard for dissemination of bibliographic data in the United States and the international standard by 1973. The most successful version is MARC 21, created in 1999.

The MARC 21 family of standards now includes formats for authority records, holdings records, classification schedules and community information, in addition to the format for bibliographic records. Currently MARC 21 has been implemented successfully by The British Library, the European Institutions and the major library institutions in the United States and Canada.

At the Publications Office, MARC 21 is used based on the rules as found on the Library of Congress site (http://www.loc.gov/marc/bibliographic/). This is the 1999 Edition Update No. 1 (October 2000) through Update No. 22 (April 2016). However all the latest updates are not necessarily applied. It is used to catalogue all General Publications (some examples can be found at http://opac.publications.europa.eu).

Name	Status	Version(s)
MARC 21	In production	1999 Edition Update No. 1 (October 2000) through Update No. 22 (April 2016)

Dublin Core

The Dublin Core is a schema of generic metadata. It describes numerical and physical resources allowing relation with others resources. Dublin Core Metadata may be used for multiple purposes, from simple resource description, to combining metadata vocabularies of different metadata

standards, providing interoperability for metadata vocabularies in the Linked Data cloud and Semantic Web implementations.

It is used at the Publications Office to enable global research and to share a restricted set of metadata for each resource.

Name	Status	Version(s)
Dublin Core	In production	

ONIX

Online Information eXchange (ONIX) is an international standard metadata to represent and to communicate rich information in the book industry domain. It is a specific form that book bibliographic data can be sent in.

ONIX is not limited to a specific language or to a specific national book trade. It was created mainly by the Association of American Publishers (AAP) and the eDITEUR association in London in response to the industry's need.

ONIX is XML-based standard and its structure is very well suited to handle diverse information that can be conveyed for each product such as book descriptions, author biographies, catalog copy and even excerpts.

There are several versions of ONIX since 1998. The latest version is 3.0, but many retailers have not yet updated from version 2.1.

More details can be found on the following link http://onixedit.com/en-us/.

Name	Status	Version(s)
ONIX	In production	2.1/3.0

OP Core Metadata

In 2011 the Publications Office of the EU has defined and adopted its own OP Core metadata element set based on the Dublin Core metadata element set. It consists of 16 elements that represent resources managed and published by the Publications Office through its OP Portal. Some extensions are being defined and will be added for specific domains, such as the legal or the general publications domains. More details can be found in the following link:

https://op.europa.eu/en/web/eu-vocabularies/opcm.

Name	Status	Version(s)
OP Core Metadata	Available	0.10
OP Core Metadata	Under development	0.11

Premis

[PREservation Metadata: Implementation Strategies](#) (PREMIS) is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability.

It was defined in 2003 by an international group sponsored by [OCLC](#) and [RLG](#). Currently, PREMIS is maintained by the PREMIS Maintenance Activity, sponsored by the Library of Congress. The current version of the schema is 3.0 and the version of the ontology is 2.2, although the PREMIS Ontology Working Group has started working on the definition of a new version of the ontology, which will be based in version 3.0 of the schema.

At the Publications Office, PREMIS will be used in Cellar in the following months to define the provenance metadata (PREMIS Ontology 2.2). It is implemented in EUROR[6] 3 for defining Preservation Descriptive Information such as preservation metadata, provenance metadata, PREMIS 3 Data Dictionary.

More detailed information can be found in the following links.

https://www.loc.gov/standards/premis/ontology/index.html and
https://www.loc.gov/standards/premis/v3/index.html.

Name	Status	Version(s)
Premis Ontology	Under evaluation	2.2
Premis	In production	3

4.2 Transmission protocols

IMMC

The Inter-institutional Metadata Maintenance Committee (IMMC[7]) is a transmission protocol based on XML schema use, it is in charge of the harmonisation and standardisation of metadata to facilitate the exchanges of information between and amongst the Institutions at European level. This Committee manages the transmission protocol (also named IMMC) which has been approved by the EU institutions.

The IMMC [Core metadata](#) is the minimum set of metadata elements that are used in all the data exchange related to the decision making process between the institutions and the Publications

---

[6] EUDOR 3 in the long-term digital preservation archive of the Publications Office. It is based on the RODA software (http://www.roda-community.org) and on international and OP standards, such as OP Core Metadata, PREMIS, and OAIS. EUDOR 3 archives all the publications of the Office, but it will also archive additional material, such as the Web Archives of the EU.

[7] The IMMC abbreviation is used either for the committee or for the transmission protocol.

Office. The IMMC transmission protocol includes this Core metadata and also extensions have been defined for specific metadata elements in particular context. The IMMC transmission protocol makes extensive use of authority tables.

More detailed information can be found in the following links.

https://op.europa.eu/en/web/eu-vocabularies/immc and https://op.europa.eu/en/web/eu-vocabularies/authority-tables.

Name	Status	Version(s)
IMMC	In production	2
IMMC	In production	3(beta)

OJEEP

Official Journal Electronic Exchange Protocol (OJEEP) was defined by the Publications Office of the European Union to manage the data exchange between PlanJO, the production follow-up system for the Official Journal of the European Union and the contractors in charge of printing the Official Journal. The data managed by OJEEP are manuscripts, models, instructions, proofs, orders, etc.

To cover all business requirements, the Current version (OJEEP version 07.07-20160315) consists of the several schemas: common elements schema, general schema, legal value schema, MSG schema, supplier schema.

Other specific versions and more information can be found in the following link: https://op.europa.eu/en/web/eu-vocabularies/ojeep .

Name	Status	Version(s)
OJEEP	In production	07.07

4.3 Controlled vocabularies

NALs

Named Authority Lists (NALs) are controlled vocabularies listing metadata values with their authority code and label(s) in the 24 official languages of the European Union. They are also called Authority tables (ATs). A number of NALs are part of the IMMC exchange protocol used in the data exchange between the institutions involved in the legal decision making process. These NAL sis under governance of the Interinstitutional Metadata Maintenance Committee (IMMC) and maintained by the Publications Office of the EU in its EU Vocabularies website. The list and the documentation of the Authority Tables are available on https://op.europa.eu/en/web/eu-vocabularies/authority-tables.

The [released NAL versions](#) are also available on the [European Union Open Data Portal](#) (ODP). The following resources are considered as well NALs at the Publications Office: EuroVoc Thesaurus, ATTO tables.

Name	Status	Version(s)	Release
NALs	In production	Not Applicable	20191211

NALS are also known as Common Authority Tables (CATs), controlled vocabularies or value lists.

EuroVoc

EuroVoc is the EU's multilingual thesaurus. It is a multilingual, multidisciplinary thesaurus covering the activities of the EU. It contains terms in 23 EU languages[8], as well as in the language of three candidate countries.

EuroVoc is managed by the Publications Office and used by several users such as the European Union Institutions including the Publications Office, national and regional parliaments in Europe, plus national governments and private users around the world.

The EuroVoc resource is available for [download in EU Vocabularies](#) and on the [EU Open Data Portal](#) under two distributions (SKOS and XML). Alignments with other thesauri can be found there as well.

For more information, please visit the [EuroVoc website](#).

Name	Status	Version(s)
EuroVoc	In production	4.9.2
EuroVoc (new release)	Under publication	4.10
EuroVoc (next release)	In development	4.11

5. Semantic Technologies and Data Models

The Office's central content and metadata repository, the Cellar, contains all legislative texts and increasingly other types of texts. A recent workshop (http://publications.europa.eu/en/web/access-reuse-conference/reuse-workshop) gives a set of useful links. The Cellar SPARQL endpoint is at http://publications.europa.eu/webapi/rdf/sparql.

The Office's website https://op.europa.eu offers both a SPARQL end-point and a "user-friendly" SPARQL search option to the contents of the Cellar. The Linked Data Query Wizard uses the web-

---

[8] Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish.

based open source SPARQL editor, Flint, to construct a query without having to be a SPARQL expert – and you can export the query it composes into the SPARQL end point, called the "expert search".

The European Union Open Data Portal makes the catalogue metadata available as Linked Data. It offers a SPARQL endpoint which allows queries on the RDF descriptions of datasets: http://data.europa.eu/euodp/sparqlep. Also a graphical user interface is provided for entering SPARQL queries http://data.europa.eu/euodp/en/linked-data

Sample queries are offered to facilitate retrieval of information about data and metadata stored in the EU ODP triple store.

5.1   LOD

Linked Open Data (LOD) is a method of using the Web to publish structured data. As shown in Figure 2, using the Web is the best practice to connect data, pieces of data, information and knowledge on the semantic Web. Semantic Web technologies (RDF, OWL, SKOS, SPARQL, etc.) provide an environment where applications can query that data.
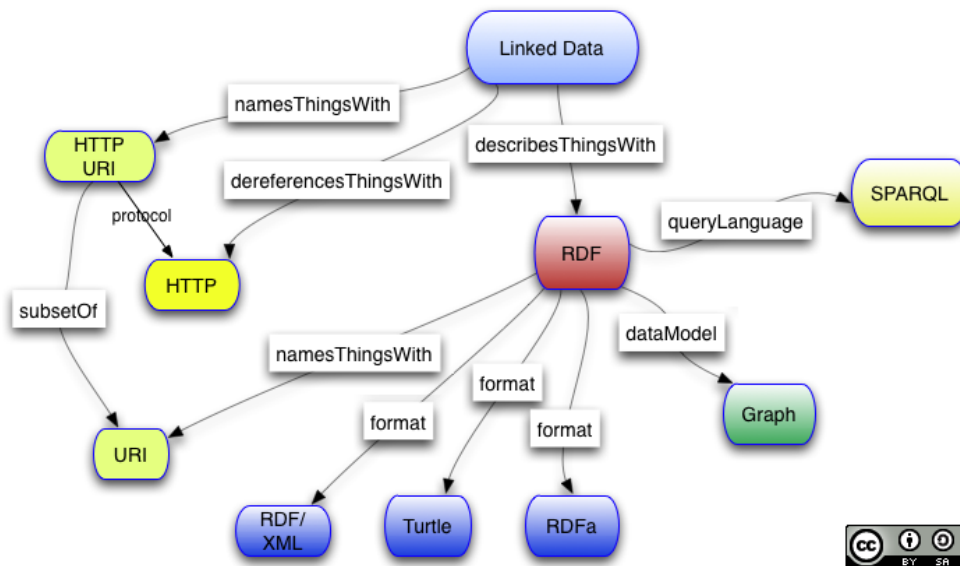

Figure 2: Linked Open Data

Using Linked Open Data, links were made between related resources and datasets can be explored and queried through the standard query language SPARQL to facilitate the accessibility and the reusability of data at various levels of complexity.

The Cellar architecture at the Publications Office is compatible with the Linked Open Data principals. More details can be found in the Cellar interface description. The EU Open Data Portal makes the catalogue metadata and some datasets available as LOD as well.

Name	Status	Version(s)
LOD	In production	Not Applicable

5.2   RDF

The Resource Description Framework (RDF) is a standard model for data interchange on the Web. RDF facilitates structured and semi-structured data to be mixed, exposed and shared across different applications even under different schemas. It specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF enables users to link, analyze and discover information horizontally, as shared services, across a variety of enterprise and Web sources. RDF is a language for representing information about resources in the World Wide Web.

RDF is a W3C standard, originally designed as a metadata data model. It is used also as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialisation formats. The RDF data model is similar to classical conceptual modeling approaches such as entity relationship or class diagrams, as it is based upon the idea of making statements about resources in the form of subject-predicate-object expressions.

RDF is also used in knowledge management applications. It was adopted as a W3C recommendation in 1999. The RDF 1.0 specification was published in 2004, the RDF 1.1 specification in 2014.

At the Publications Office, RDF-XML is mainly used to describe the Common Data Model (see section 5.4). An RDF based interface is used for upload of metadata, stored in triple store and accessible through SPARQL endpoint.

Name	Status	Version(s)
RDF	In production	1.0/1.1

5.3   OWL

Web Ontology Language (OWL) is a semantic web language designed to represent rich and complex knowledge. It extends RDF and RDFS[9]. It aims to bring the expressive and reasoning power of description logic to the semantic web.

OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit. OWL documents, known as ontologies, can be published in the World Wide Web and may refer to or be referred from other OWL ontologies. OWL is part of the W3C's Semantic Web technology stack, which includes RDF, RDFS, SPARQL, etc. The current version of OWL (OWL 2) was developed by the W3C OWL Working Group and published in 2009, with a Second Edition published in 2012. OWL 2 is an extension and revision of the OWL first edition published in 2004 (OWL 1).

Name	Status	Version(s)
OWL	In production	1/2

---

[9] Resource Description Framework Schema (RDFS) is a set of classes with certain properties using the RDF extensible knowledge representation data model.

5.4   SPARQL

SPARQL is an RDF query language. It is a semantic query language for RDF databases, able to add, modify or delete RDF data stored in RDF format. It is one of the main technologies of the semantic web. It was made a standard by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium. SPARQL 1.0 became an official W3C Recommendation in 2008 and SPARQL 1.1 in 2013.

SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL is able to query required and optional graph patterns along with their conjunctions and disjunctions. It supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be results sets or RDF graphs.

At the Publications Office, SPARQL is used to mainly query metadata. In fact, Cellar regroups all content and metadata created or disseminated by the Publications Office. It also drives the Publications Office's major portals, including EUR-Lex and OP Portal. Cellar resources are semantically described by the CDM (Common Data Model), an FRBR-compliant OWL ontology, which also serves as the basis for the Publication Office's ELI implementation. It is publicly accessible and allows all interested parties to retrieve content and its RDF structured metadata for reuse. A [first public SPARQL endpoint](#) is maintained by the Cellar team and made available to query the metadata store.

The EU Open data portal provides a [second SPARQL endpoint](#). It allows queries on the RDF descriptions of datasets. Documentation can be found at http://data.europa.eu/euodp/en/linked-data, including the URL to access it, ontology, etc.

Name	Status	Version(s)
SPARQL (Cellar)	In production	1.1
SPARQL (EU Open Data Portal)	In production	

5.5   CDM

The [Common Data Model](#) (CDM) is used by the Publications Office to publish resources. It is an OWL ontology structuring the metadata of all of Cellar's texts. It is based on the [FRBR model](#), able to represent the relationships between the resource types managed by the Publications Office and their views according to the FRBR model in terms of Work, Expression, Manifestation and Item.

CDM establishes a data model at core level, but there are extensions that captures and brings together all of the OP's metadata such as metadata related to legal resources and General Publications, and will include metadata for TED and CORDIS, etc.

The Common Data Model is represented as ontology (concepts and the relationships) according to Web Ontology Language (OWL) format. The ontology formally defines the various classes and properties and assigns unique URIs to them.

Name	Status	Version(s)
CDM	In production	4.3.4
CDM	In development	4.3.5

5.6   DCAT-AP

DCAT Application profile for data portals in Europe (DCAT-AP) is a specification based on the Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe. The elaboration of the DCAT-AP was a joint initiative of DG CONNECT, the EU Publications Office and the ISA Programme. The specification was elaborated by a multi-disciplinary Working Group with representatives from 16 European Member States, some European Institutions and the US.

DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. It was developed under the responsibility of the Government Linked Data Working Group at W3C. The work on DCAT was initiated at the Digital Enterprise Research Institute (DERI) and the Greek National Centre for Public Administration and Decentralization.

DCAT-AP basic use case is to enable a cross-data portal search for data sets and make public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of data sets among data portals.

The latest version is accessible under the following link: https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/releases

At the Publications Office the European Union Open Data Portal is adopting the DCAT-AP profile with some local extensions for its metadata catalogue.

Name	Status	Version(s)
DCAT-AP	In production	2.0.0

6. Presentation formats

6.1   PDF

The Portable Document Format (PDF) is a file format used to present documents independently of software, hardware and operating systems. It was developed in the early 1990s as a way to share documents, including text formatting and images. PDF was controlled by Adobe, then released as an open standard on 2008 and published by the International Organisation for Standardisation.

Three types of PDF are used at the Publications Office:

- PDF/A-1a is a constrained PDF version 1.4 intended to be suitable for long-term archiving. The ISO standard [ISO 19005-1:2005] was developed by a working group with representatives from government, industry and academia and active support from Adobe Systems Incorporated. The version accepted by the Publication Office is ISO 19005.

- PDF/x is a subset of the PDF ISO standard developed to facilitate graphics exchange and it therefore has a series of printing related requirements which do not apply to standard PDF files. The used norm by the Publications Office is ISO 15930. The general accepted version is PDF/X-1a.

- PDF/A is an ISO-standardised PDF version specialised for the digital preservation of electronic documents. PDF/A includes features unsuited to long-term archiving, such as font linking (as opposed to font embedding). The ISO requirements for PDF/A file viewers include color management guidelines, support for embedded fonts and a user interface for reading embedded annotations. The Publications Office uses it as optimised version for transferring documents to websites so that some features have to present for example the activation of links even in the table of content of the document, etc.

Name	Status	Version(s)
PDF/A-1a	In production	1.4 (ISO 19005)
PDF/x	In production	(ISO 15930), PDF/X-1a

6.2   ePub

ePub is an e-book file format that can be used on devices like smartphones, tablets, computers or e-readers. It is a free and open standard published by the International Digital Publishing Forum in September 2007, replacing Open eBook standard. ePub is the most widely supported vendor-independent XML-based e-book format.

At the Publications Office, ePub is mainly used to publish General Publications and recently for legislative documents as well.

OP as member of the IDPF (International digital publishing forum), a forum created by private and institutional publishers and W3C to develop and maintain formats, such as ePub and HTML, is closely following-up on developments of new formats – notably interoperable formats and accessibility questions – and adapts his offer to the development of new formats.

Name	Status	Version(s)
ePub	In production	2.x/3.x

6.3   HTML, XHTML

HyperText Markup Language (HTML) is the standard markup language used to create Web pages to be read by Web browsers. It can embed scripts such as JavaScript or Cascading Style Sheets (CSS) to modify the behaviour or to define the look of HTML web pages.

Extensible Hypertext Markup Language (XHTML) is part of the XML family. XHTML was developed to make HTML more extensible and increase interoperability with other data formats. While HTML was defined as an application of Standard Generalised Markup Language (SGML), a flexible markup language framework, XHTML is an application of XML, a more restrictive subset of SGML.

XHTML 1.0 became a W3C Recommendation in 2000, followed by XHTML 1.1 in 2001. The standard known as XHTML5 is being developed as an XML adaptation of the HTML5 specification.

Name	Status	Version(s)
HTML	In production	4/5
XHTML	In production	1.0/1.1

6.4   XForms

XForms allows a flexible presentation of forms, including classic XHTML forms to be attached to an XML form definition; it separates between sections that describe what the form does and how it looks. XForms collects data expressed as XML instance data following an XForms Model and using an XForms Submit Protocol. More details can be found in the following link: http://www.w3.org/MarkUp/Forms/#waXForms

At the Publications Office, XForms is used with the eNotices application.

Name	Status	Version(s)
XForms	In production	1.1

6.5   RTF

Rich Text Format (RTF) is a proprietary document file format developed and maintained through several versions by Microsoft Corporation for cross-platform document interchange with Microsoft products.

RTF is used by various European institutions including the Publications Office to convert XML documents mainly for editing purposes.

6.6   Images

The following list was approved as standards regarding current digital imaging guidelines and best practices developed and employed by various European institutions including the Publications Office.

1. TIFF (Tagged Image File Format): is one of the most popular and flexible file formats for storing raster graphic images, popular among graphic artists, the publishing industry and photographers. It is supported by image-manipulation applications, by publishing and page layout applications and by scanning, faxing, word processing, optical character recognition and other applications.

2. JPEG (Joint Photographic Experts Group): is the most common image format on the internet, commonly used for compression for digital images.

3. PNG (Portable Network Graphics): is a raster graphics file format. It is the most used lossless image compression format on the Internet. It was created as an improved, non-patented replacement for Graphics Interchange Format (GIF).

4. BMP: known as bitmap image file or Device Independent Bitmap (DIB) file format or simply a bitmap, it is used to store bitmap digital images, independently of the display device.

5. SVG (Scalable Vector Graphics): is an XML-based vector image format for two-dimensional graphics optimised for interactivity and animation. The SVG specification is an open standard developed by the W3C since 1999.

Name	Status	Version(s)
TIFF	In production	6.0
JPEG	In production	2
PNG	In production	1
BMP	In production	5
SVG	In production	1

7. Framework

7.1 FRBR

The Functional Requirements for Bibliographical Records (FRBR) is a conceptual model for the bibliographic universe. It is used by the Common Data Model (see section 5.5) and by the IMMC (see section 4.2) at the Publications Office to offer a fresh perspective on the structure and relationships of bibliographic and authority records in terms of Work, Expression, Manifestation and Item (WEMI). It clarifies other rich relationships that enable collaboration between related items and allows navigation through the complex network of the bibliographic universe.

The service allows the user to search the RDF (Resource Description Framework) content of the given object which can be a work, an expression, a manifestation, a dossier, an event, a top level event or an agent.

At the Publications Office, FRBR is integrated in Cellar to represent a publication at different levels of abstraction. The Cellar realises the WEMI pattern through several different hierarchies, each with its own level of abstraction:

- Hierarchy work-expression-manifestation-content stream: composed by a work (document as W role), several expressions (linguistic version as E role), several manifestations (specific format as M role), several content streams (physical file as I role). The Cellar contains works from OP's primary domains of work (Legislative data published in EUR-Lex, General Publications published in [EU Publications](#)) and will probably contain tender documents published in the TED portal and research documents published in Cordis portal. The WEMI model is applied throughout for works from all domains. However, the abstract classes such as work are then concretised for the various domains in the Cellar's Common Data Model (CDM) by subclassing them. They are documented in the CDM's wiki page http://www.cc.cec/wikis/display/OP/CMR+Common+Data+Model:

- Hierarchy dossier-event: composed by a dossier (W role) and several events (E role)

- Hierarchy event: called top level event hierarchy and composed by an event (W role)

- Hierarchy agent: solely composed by an agent (W role)

The realisations of the WEMI pattern are the basis of the Cellar's definition and data layer (ontology).

Name	Status	Version(s)
FRBR	In production	Not Specified

7.2 TOGAF

The Open Group Architecture Framework (TOGAF) is the de facto global standard for Enterprise Architecture. It is the process for better doing architecture. It brings business, information and infrastructure architecture together.

The TOGAF standard is developed and maintained by the [Open Group Architecture Forum](#), comprised of more than 200 enterprises. The TOGAF framework enables organizations to effectively address critical business needs by Ensuring that everyone speaks the same language, avoiding lock-in to proprietary solutions by standardizing on open methods for Enterprise Architecture, saving time and money, and utilize resources more effectively and by achieving demonstrable ROI.

At the Publications Office, TOGAF framework is used inside ARIS 7.2 mainly to describe OP IT environment. Several models describing the IT applications, the technological platforms and the technologies used are available in the Publications Office ARIS repository and accessible via ARIS Publisher.

Name	Status	Version(s)
TOGAF	In production	9

7.3   OAIS

The Open Archival Information System (OAIS) reference model is a conceptual framework for an archival system dedicated to preserve and maintain long term access to digital information for a designated community of users. OAIS helps archival and non-archival institutions to be familiarized with the preservation procedures, by providing the fundamental concepts for preservation and its related definitions aiming to avoid any confusion on the used digital preservation terminology. Although most of the organizations implement it to support the digital resources preservation, it can fit to the preservation of analogous resources as well. The OAIS is also published as an ISO standard, the ISO 14721.

At the Publications Office, OAIS is used as the basis for the functioning of EUDOR 3, the long-term digital preservation archive of the office.

Name	Status	Version(s)
OAIS/ISO 14721:2012	In production	

8. Useful Links

1. Interinstitutional style guide: http://publications.europa.eu/code/

2. The International Organisation of Standardisation (ISO): http://www.iso.org/iso/

3. Report prepared for the ISA program (Interoperability Solutions for European Public Administrations):
    https://joinup.ec.europa.eu/sites/default/files/42/b6/1e/ISA_LEOS_Final_Results_Final_Version.pdf

4. The Legal Entity Identifier (LEI): https://www.gleif.org/en/lei-focus/what-is-an-lei

    https://joinup.ec.europa.eu/sites/default/files/ckeditor_files/files/7-Rowley_2016-05-12_SEMIC%20GLEIF-Slides_v1_0.pdf

    The Global Legal Entity Identifier Foundation (GLEIF): https://www.gleif.org/

9. Under construction

9.1   IFC-IMFC

  The Interinstitutional Formats Committee (IFC) has been created in order to facilitate the production, exchange, publication, preservation and reuse of EU legislative documents. It facilitates the interoperability of documents between EU institutions including validation rules for the common structure by defining XML based common semantic structures for the documents (in particular legislative documents) content and respective validation rules.

The IFC is an interinstitutional platform which allows collaboration and knowledge exchange and guarantees the maintenance of the specifications. In order to safeguard consistency with IMMC (see section 4.2) a close collaboration between the two standards is ongoing.

Since 2018, IFC merged with IMMC to become IMFC (Interinstitutional Metadata and Formats Committee). The IMFC has been tasked with developing and maintaining common interinstitutional standards for the secured and automated exchange of data based on machine-readable structured formats for content and metadata.

Akoma Ntoso for European Union (AKN4EU) is the future machine-readable structured format for the exchange of legal documents in the EU decision-making process. It is based on XML, and more specifically on Akoma Ntoso, an OASIS standard, and should serve as the specification for all future exchanges of legal documents. The scope of the current version includes legal acts adopted through the ordinary legislative procedure (regulations, directives and decisions), as well as the respective legislative proposals. The Interinstitutional Metadata and Formats Committee (IMFC) will further develop and maintain AKN4EU as the interinstitutional standard for the exchange of structured content.

Name	Status	Version(s)
Common Vocabulary (CoV)	ongoing	
AKN4EU*	ongoing	2.0/2.1

*: Formerly Common Exchange Model (CEM).

9.2   GenDoc

GenDoc is a document formatting standard developed by the Court of Justice of the European Union (CJEU) based on the ISO standard Open XML and used by the OP for the production of CJEU Case Law.

The Office Open XML standard, known as the Open XML is a document format developed by Microsoft in December 2006 that consists mainly of XML documents stored in a ZIP file.

Open XML is also based on the "Open Packaging Conventions" standard (commonly known as OPC) to store XML, binary, and other parts in a package at large (mostly as a ZIP file).

Adopted by ISO in November 2008 as the ECMA-376 2nd Edition (ISO / IEC 29500), the Open XML standard is divided into four parts:

• Part 1 ("Fundamentals and Markup Language Reference") consists mainly reference materials around the XML structure defined by the standard. It shall include XML schemas.

• Part 2 ("Open Packaging Conventions") contains a description of the OPC standard. This part of the standard is used by other standards, including XPS.

• Part 3 ("Markup Compatibility And Extensibility") contains a description of the elements and attributes that enable applications to have other means for content description.

• Part 4 ("Transitional Migration Features") is attractive to compatibility issues settings as well as support graphical markup language VML.

GenDoc scrupulously respects this standard for document generation based on Open XML SDK including 2.6, the SDK (Software Development Kit) Official Microsoft for document generation in OpenXML format.

GenDoc uses the "WordProcessingML" Open XML package for the generation .docx compliant MS Word 2010/2013.

Name	Status	Version(s)
GenDoc	ongoing	1.2

9.3   eProcurement ontology

The e-Procurement Ontology (ePO) is a common data standard under development by the publications Office for publishing procurement data. It allows data from different sources to be easily accessed and linked, and consequently exchanged and reused under the same practice.

ePO aims to harmonise the vocabularies and the semantics introduced by different initiatives coming from the public sector, the industry and academia, to align the phases of public procurement, and to use the same technologies. The objective is to put a commonly agreed OWL ontology that will conceptualise, formally encode and make available in an open, structured and machine-readable format data about public procurement, covering it from end to end, i.e. from notification, through tendering to awarding, ordering, invoicing and payment.

An informative report grouping all the necessary information to proceed with the specification of the ePO is under preparation and can be requested.

Name	Status	Version(s)
ePO	ongoing	