Interinstitutional Metadata and Formats Committee

Metadata subgroup

IMMC schema compliance criteria's

Date:	27/01/2022
Doc. Version:	0.3

Document Control Information

Settings	Value
Document Title:	IMMC schema compliance criteria
Basis	
Doc. Version:	0.3
Date:	27/01/2022

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
		<Approve / Review>	

Document history:

Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
0.3	27/01/2022	Martin SCHERBAUM	Reworked section 3
0.2	24/01/2022	Marc VANDERPERREN	Review and formatting
0.1	21/01/2022	Martin SCHERBAUM	Initial version

Contents

1. INTRODUCTION

1.1. Purpose and audience of the document

In the context of the lifecycle of IMMC standardisation requests, a standardisation solution is provided to the members of the IMMC Metadata subgroup. The standardisation solution can include a preliminary schema published on CELLAR i.e. a schema provided along the documentation of the standardisation request in order to support its validation.

IMMC schemas intended for production are published on the EU Vocabularies portal.

The purpose of this document is to describe the criteria's that any IMMC schema must comply with, in order to ensure the long-term stability and usability of IMMC schemas by all stakeholders.

The criteria's include description of the (directory) structure of IMMC schemas and their presentations.

The members of the IMFC Metadata subgroup are
- The intended audience of the document and of the IMMC extensions overview,
- Invited to comment on this document.

2. DEFINITIONS

Term	Definition
IMMC series	Base line IMMC schemas e.g. IMMCv2, IMMCv3, IMMC public access
XML parser	A program which parses and validates a given XML instance against an XML schema by loading a specific XSD file
xsi:schemaLocation	An attribute on the root element of an XML schema, pointing to a URL indicating the parsing entry point.
Manifestation	Term from the FRBR methodology designating a grouping entity for files (can be understood as a directory containing files)
Authority table / AT	An authority table offers a list of concepts representing the same semantic scope, e.g. a list of country codes, a list of language codes; in scope of IMMC the authority tables are stored in form of enumerations in XSD files
Common Repository	The Common Repository is the EU institutions document dissemination storage holding both the content files and metadata for the documents. It is accessible via a http-REST interface. Also the IMMC schema distributions are available from there.
IMMC schema distribution	The IMMC schema in the strict sense is a collection of XSD files. However, for online use and download, the IMMC schemas (all IMMC series, all variants, all versions) are combined with documentation, samples and other materials. This compound of information is called a distribution.

3. IMMC SCHEMA COMPLIANCE CRITERIAS

3.1. Criteria 1: backward compatibility

All modifications to an IMMC series (IMMC v2, IMMC v3, IMMC public access) are done in a backward compatible way.

IMMC schema backward compatibility is fulfilled if any IMMC message instance validated against a schema series remains valid against the new release of the same schema series.

This criteria is attested by OP on each schema release by means of the provided sample IMMC descriptors to validate automatically against the new schema release.

3.2. Criteria 2: schema consistency

The parsing and validation of any IMMC message type is actually performed only by declared IMMC schema files (i.e. imported by each other, see explanation below). In other words, not all XSD files within an IMMC schema distribution are needed for the parsing and validation of an IMMC message type.

Every IMMC message starts with an XML root element with a name such as `xyz_transmission_request`.

The definition of those root elements is done in the (usually named) `xyz_transmission.xsd` file in the (also usually named) `xyz` domain specific subdirectory of the schema distribution.

The `xyz_transmission.xsd` includes the `cm_xyz_extensions.xsd` which in turn includes `core_metadata.xsd`, and so on. In consequence, only a couple of (XSD) files from the schema distribution are needed for the parsing of a specific message type. All other files in the distribution are irrelevant for the parsing of that IMMC message type.

The example below shows a schematic **inclusion order and depth** for the fictitious domain specific extensions "`myexta`" and "`myextb`". The first example assumes that `myexta` does nothing special, the second `myextb` is a more "ambitious" extension. All file names contain the path to the top level directory, file names in italic refer to core metadata related files.

```
        myexta/myexta_transmission.xsd
                cm_transmission.xsd
                core_metadata.xsd
                        at-corporatebodies.xsd
                        at-languages.xsd
                        [... further ATs ...]
                myexta/cm_myexta_extensions.xsd
```

```
        myextb/myextb_transmission.xsd
                cm_transmission.xsd
                        publication_request.xsd
                core_metadata.xsd
                        at-corporatebodies.xsd
                        at-languages.xsd
                        [... further ATs ...]
                myextb/cm_myextb_extensions.xsd
                        cm_common_extensions.xsd
                        myextb/at-myextb-local-at.xsd
```

For any defined root element like `xyz_transmission_request`, the OP attests that - independent from IMMC series or version - the complete set of needed XSD files is present

and can be imported properly. This does not only apply to the local variant of the schema distribution but also for the decoupled one (see section xxx).

3.3. Criteria 3: stable file system layout

3.3.1. *Purpose and structure*

An IMMC schema distribution serves several functions:

- Provide the different IMMC schema threads to allow the parsing / validation of IMMC messages.
- Provide documentation on the entirety of all contained schema threads but also for each domain specific extension.
- Provide IMMC message samples for the different domain specific extensions, illustrating the variety of possible implementations of one schema thread and allowing the validation of the given schema thread on development and/or testing.
- Provide additional information related to the specificities of the given IMMC distribution.

IMMC schema distributions shall adhere to the file system layout described here, as minimal guaranteed standard:

3.3.1.1. Toplevel directory

- Core metadata related XSD files (e.g. `core_metadata.xsd`, `cm_transmission.xsd`, `cm_common_extensions.xsd`)
- General IMMC series related XSD files (e.g. `publication_request.xsd`, `nth-layer-extension.xsd`)
- Release notes as PDF document to describe the release
- One directory for each domain specific extension
- `deprecated` directory needed for the redirection XSDs of the deprecation/compatibility mechanism
- `documentation` directory containing e.g. transmission protocol descriptions as PDF documents
- `staging` directory for experimental schema extension XSDs (see "domain specific extensions")

3.3.1.2. Domain specific extension directories

- Common metadata extension XSD (e.g. `oib_cm_extensions.xsd`)
- Transmission extension XSD (e.g. `oib_transmission.xsd`)
- Possibly other domain specific files
- `samples` directory containing either sample IMMC descriptor XML files, subdirectories with the contents of sample IMMC transmission (including an IMMC descriptor XML file) or IMMC packages (ZIP files with the content of an IMMC transmission); the extent of the contents in these sample directories depends on the complexity of the concerned exchange domain and the number of potential addressees of the transmissions in that exchange domain

3.3.1.3. `deprecated` directory

The `deprecated` directory is part of the deprecation mechanism which is necessary to allow the replacement of singular elements or entire XSD files in a way that backward compatibility of the schema (see Criteria 1) is guaranteed.

The directory was used the first time for some elements in the generic extension to remediate a namespace URI fix. The directory is a valuable tool as well in the context of the gradual harmonisation between the IMMC series or domains specific extensions.

3.3.1.4. documentation directory

- General documentation about IMMC
- Domain specific documentation, e.g. transmission protocols

3.3.2. *Additional explanatory notes about the stable file system layout*

The **only necessary condition for the backward compatible functioning of the schemas** (see criteria 1) is that the domain specific transmission XSD file, i.e. the entry point of parsing/schema validation, is located in the domain specific extension directory. This XSD file includes all other schema components needed for this schema thread and consequently knows (by relative paths) where to find them in the schema directory structure; this allows to change the paths for implementation flexibility in the future. XML parsers watch only files explicitly imported/included by one of the involved XSD files, and ignore all other files not belonging to the given schema thread.

As a **good practice**, the entry point for parsing/schema validation should usually be declared in the IMMC message instance as the `xsi:schemaLocation` attribute on the root element.

The automated IMMC schema build process ensures that all domain specific parsing entry points are present and lead to a consistent schema loading for parsing/schema validation.

The directory structure of specific releases may vary e.g. by having additional files motivated by new features or extended business requirements of special IMMC releases.

3.4. Criteria 4: authority tables

The IMMC schemas allow to define element values to originate from name authority lists (authority table = AT), which list all acceptable (code) values for a given semantic concept scheme. For the use in the context of IMMC schemas, authority tables are converted into XSD files with an enumeration of values per table.

The ATs can be used for parsing/schema validation in
- A **decoupled way;** ATs are loaded remotely via http from the Common Repository – see the URLs in section 4.2), or
- A **local way**; ATs are stored along with the schema XSD files.

The EU vocabularies portal has been disseminating since 2018 two schema distributions per IMMC series:
- One with local ATs; this distribution is intended for download of the schema together with the ATs
- One with decoupled ATs; this distribution is intended for online validation (schema and ATs are used by the parser from the Common Repository, no local schema installation/maintenance)

Both distribution schemes offer a ZIP with the schema and the ATs for download, the decoupled distribution offers additionally all schema files for direct download.

4. STRUCTURE OF THE SCHEMA DISTRIBUTIONS

4.1. Explanatory notes

The distributions of IMMC schemas stored for download in the Common Repository follow a well-defined modelled structure which has evolved over time, in a way similar to the internal directory structure of the IMMC schemas.

IMMC schema distributions have been available from thonlinee Common Repository since September 2018; the download of IMMC schemas can be done directly from the Common Repository or via the EU vocabularies website[1].

The structure of schema distributions is expressed in a set of URI schemes for the download or addressing of the different IMMC schema related artefacts.

4.2. URI examples

Explanations about example URIs of this section:
- The example URIs below are referring to IMMCv2 schema series but can be applied similarly to other IMMC series as well,
- The release identifier, i.e. release-id, in the examples below is set to **20201216-0**. The current latest release a the time of writing this document,
- The leading URI part http://publications.europa.eu/resource/distribution has been replaced by <uri-root> for readability reasons.

4.2.1. *Download the decoupled schema distribution as XSD files:*

<uri-root>/immc_2_core_metadata_decoupled/20201216-0/xsd/schema_immc_v2

This URI will offer the current list of files contained in the manifestation for download one by one.

4.2.2. *Download the related schema distribution as ZIP archive:*

<uri-root>/immc_2_core_metadata_decoupled/20201216-0/zip/schema_immc_v2

4.2.3. *Download the related documentation as single files:*

<uri-root>/immc_2_core_metadata_decoupled_documentation/20201216-0/pdf/stddoc/

This URI will offer the current list of files contained in the manifestation for download one by one.

4.2.4. *Download the local schema distribution (together with ATs) as XSD files:*

<uri-root>/immc_2_core_metadata/20201216-0/xsd/schema_immc_v2

This URI will offer the actual list of files contained in the manifestation for download one by one.

4.2.5. *Download the related schema distribution (together with ATs) as ZIP archive:*

<uri-root>/immc_2_core_metadata/20201216-0/zip/schema_immc_v2

4.2.6. *Download the related documentation as single files:*

<uri-root>/immc_2_core_metadata_documentation/20201216-0/pdf/stddoc/

---

[1] E.g. https://op.europa.eu/en/web/eu-vocabularies/dataset/-
/resource?uri=http://publications.europa.eu/resource/dataset/immc_2_core_metadata