IMMC Survival Kit

- **IMFC vs IMMC:** Today the term IMMC refers to the IMMC protocol which is based on the XML standard. The term IMFC (Interinstitutional Metadata and Formats Committee) refers to the governance body which created the IMMC standard.

- **What IMMC isn't:** IMMC is **not a means to broadcast messages** but is a scalable point-to point type of communication. For any given message, the purpose, sender and receiver must be defined a-priori. This means the range of specific items selected from the vocabulary to be used and the structure of the elements in the metadata file are known to the sender and receiver. Consequently, a (validating) schema can be used to check if the metadata message is complete and correct. This function can be delegated to and automated by information systems, allowing for more quality assurance along with reducing the need for tedious manual validation tasks.

- **What IMMC isn't:** Although both IMMC and FORMEX use XML documents to annotate and describe files, their purpose is very different. Indeed, IMMC **transports** metadata providing an additional standardised data layer alongside the content files (PDF, Word). The administrative data and metadata sent in the resulting descriptor **complements** the files and aids Search Engines in the retrieval of files. Through modelling (the Common Data Model, CDM), the OP also provides an additional semantic layer, which is useful in deriving links to similar items. **FORMEX** on the other hand is a standard to **describe the content files** (PDF, Word..) in a machine-readable way.

- **The workings of IMMC:** Schema-lines define the syntax (structure) of the metadata document and the vocabulary to be used in the metadata document. Each business purpose and sender and receiver couple agree on the specific transmission contract between the parties, which per se specifies the relevant and acceptable sets of metadata to be exchanged in a transmission. As the transmission contracts are defined a-priori and cannot change without the approval of the OP, each institution knows which type of IMMC message they will receive, from whom and how to handle it.

- **The workings of IMMC:** Both schemas have a similar architecture, in that they are designed for configurability and resilience. To date IMMC v2 and IMMC v3 are **stable** (little to no bugs occur) and the OP adapts the schema depending on the needs of the institutions (evolutive maintenance).

- **The workings of IMMC:** The IMMC Vocabulary is available and maintained on VocBench but is not published separately on the IMMC Vocabularies website.

- **The workings of IMMC:** *Common extensions can be reused by other extensions. Download the Technicalities for more information.*

- **The workings of IMMC:** We have four types of validation (2 axis of validation), see table to understand the capabilities automated validation confers. With regards to the ones in use we note:

  Structural validation of the IMMC descriptor – Manual. Corresponds to validation of the structure. These rules are validated automatically running the IMMC descriptor against the XSD schema (the implementation of the contract).

  Content (or business) validation of the message
  For the IMMC descriptor this validation is highly dependent on the context of the exchange and on the combination of metadata values sent. In the OP, once the transmission contract has been drawn, business validation on the IMMC descriptor is **mainly automatic**. The descriptor is however actively **manually** validated in the test campaigns (when validating requirements).

  With regards to the **automated business validation of any content file,** we note that after validating requirements with the affected stakeholders, these rules are transcribed and implemented through technical tools in the OP's information system (e.g. Schematron for FMX, additional rules for other files).

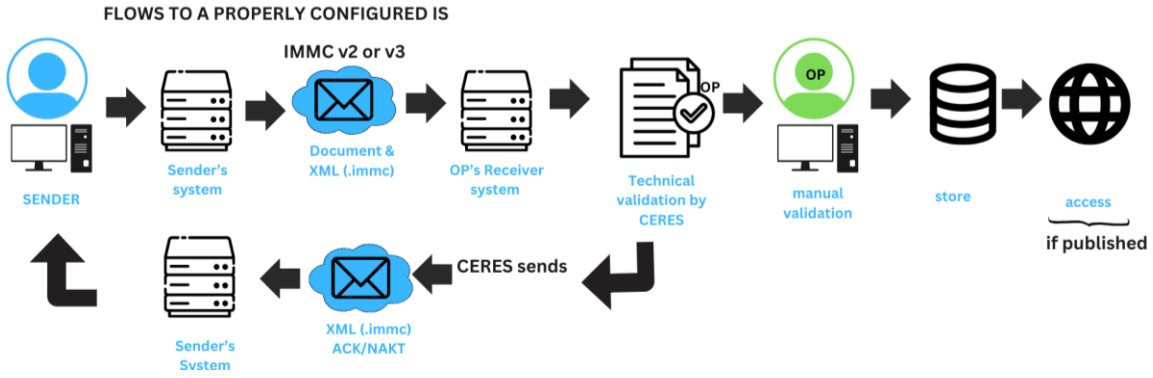
  Throughout the operations lifecycle, experts may flag metadata issues (e.g. incorrect metadata) to the OP. If this happens, the files concerned will be **manually inspected**, and if the issue is confirmed, the OP will **manually fix** the issue and update the rules of the receiving IS.

	AUTOMATED VALIDATION	MANUAL VALIDATION
STRUCTURAL VALIDATION	**How:** XSD	**Not necessary**
CONTENT (OR BUSINESS) VALIDATION - VALUE BASED	**How:** OP's Information System performs technical validation on the contents of the descriptor and of the contents of the files (FMX, PDF..) sent along the descriptor. For example, for XMLs, Schematron rules are used. Other tools are used for other formats.	**How:** manually. Performed in the initial requirement validation phase and when issues occur (e.g. bug-fixing).

- **The workings of IMMC:** Although any IMMC descriptor must contain the minimal information specified in the IMMC core-metadata agreement, generally a sender-receiver will draw up a

more specific contract (transmission contract), which better satisfies their needs, providing for more flexibility and additional restrictions. IMMC messages must contain one IMMC descriptor (composed of the metadata) and may also contain (optional) disseminated content. Generally, for validation purposes or metadata update, the sender may choose only to send IMMC descriptors (without accompanying disseminating files). In this regard, IMMC descriptors are a way of exchanging information (and do not merely serve to describe the data contained within the disseminated files).

- **The workings of IMMC:** Both communication protocols allow for any receiving systems to send automatic notifications to the sender (e.g. sending an IMMC v3 message with feedback with regards to the documents they have received). Stakeholders therefore agree in advance whether notifications are manual or automatic. For example, technical acknowledgment (ACK) or non-acknowledged (NACK) IMMC messages are sent back to the sender if the Information System can unzip the package, parse the IMMC descriptor (thus find the unique file bearing the _immc.xml extension) and recognize the business-identification element. In IMMC v3, business acceptance/rejection messages are also sent automatically (after the system verifies if the message was incomplete or did not respect the agreed logical business rules by validating the message against the corresponding contract or XSD schema). Successful validation rules are optional though validation errors must be reported.



- **IMMC v2 vs IMMC v3:** As the IMMC v2 and v3 were designed to support different business needs, they are **not** compatible – an IMMCv2 descriptor cannot be validated with an IMMC v3 schema and vice versa. In other words, an IMMC v2 descriptor must be validated by an IMMC v2 schema, and an IMMC v3 descriptor **must** be validated by an IMMC v3 schema. Consequently, IMMC v2 and IMMC v3 have different namespaces.

- **IMMC v2 vs IMMC v3:** An interesting difference between IMMC v2 and IMMC v3, is the fact that in IMMC v3 values from the authority tables are prefixed with the name of the authority table. For example, the value of France in IMMC v2 is FRA while in IMMC v3 it is cou:FRA. This is because, at the time, some use-cases in IMMC v3 referenced values in the authority tables

which were not unique (e.g. FRA could refer to the corporate body authority Fundamentals Rights Agencies or the country France), and therefore IMMC v3 elements had to explicitly refer to the correct term (e.g. cou:FRA for the country, or cob:FRA for the agency). Nowadays, measures are taken to ensure that new values added to authority tables are unique.

- **IMMC – Domain-specific extension mechanisms:** It is important to understand that a domain-specific metadata extension may per-se import other domain-specific metadata extensions. Notable cases in IMMC v2 are:
    - Regarding metadata with respect to the production of **OJ-Acts and publications** elements from the domain-specific extensions of the following institutions (council, European parliament, council of regions and EESC, European court of Auditors and generic) can be used. As an example, the Council can send corrections of data with regards to the OJ, enriching the metadata with additional council-specific elements. Technically, Oj_cm_extension and oject_cm_extension import domain-extensions from the following domains: commission, council, European parliament, council of regions and eesc, European court of Auditors and generic.
    - Regarding metadata with respect to the Council flows and flows to the Publications Office (internal flows or OP-contractor flows), the Council (and OP) can make use of the elements in the common-extensions of JLP. This allows the Council and OP to send enriched EUR-LEX corrections of the data. Technically, the council_cm_extension imports jlp_cm_extension, and PO_cm_extension imports the council_cm_extension (e.g. finding the extension with the following relative path ../council/council_cm_extensions.xsd) and therefore jlp_cm_extensions.

- **IMMC – domain-specific transmission schema:** In validating the XML message, the receiving system will first parse the domain-specific transmission XSD which specifies the complete set of relevant XSD files to be imported. This document (which follows the xyz_transmission_protocol.xsd protocol) imports the complete set of *relevant* XSD files. For example, the file may specify that the domain-specific core-metadata extension is to be imported (cm_xyz_extensions.xsd), along with the core_metadata and cm_transmission file. In turn these files may import other files. For example, the domain-specific core-metadata extension may import other files (other cm_extensions), the core_metadata file generally imports authority-table values, and the cm_transmission file may import cm_common_extensions file among others.

- **IMMC – changes:** When a contract is amended, all affected stakeholders (e.g. the sender and the receivers) must adopt the new schema. If only one party (e.g. the sender) updates to the latest schema, the receiver (which is still using the previous schema) will not be able to fully validate the IMMC message (as the previous contract establishes fewer rules than the current). In other words, in this case the receiver will not be able to check that the messages sent comply with the latest ruleset (added features).

- **Exchange flows:** Each exchange domain is linked to one IMMC schema - IMMC v2 or v3.

- **Preliminary schemas and samples:** When a Standardisation Request (SR) is at the solution proposal stage, the OP releases an accompanying preliminary schema, to allow the concerned stakeholders to provide their views on the proposed solution. Although the link to the preliminary schema follows a specific format it is not available on EU Vocabularies. Examples of URI formats it follows are: https://publications.europa.eu/resource/distribution/immc_x_core_metadata/releasedate-isr-ID/schema_immc_vx/cm-releasedate-isr-ID.zip, where x stands for the IMMC schema-line (e.g. 2 or 3), and the release date has the YYYYMMDD format, and the ID appends an incremental number to the YYYMMDD format leading to YYYYMMDDN format (the date of when the SR was registered). Once the solution has been accepted, the OP will provide the official release of the schema on the EU Vocabularies page, releasing accompanying samples (valid descriptors) to support the institutions in performing testing.

- **Exceptions to the directory structure:** *unlike other transmission protocols, the transmission protocol used between the EC and OP does not follow the conventional folder structure. Looking into the history:*
    - *In early 2013, the Commission used a derivation of IMMC v1 for its internal use (under the PUBLIC ACCESS project) and developed & maintained the schema independently. The IMMC v1.1 schemas were specific to the Commission and catered to inter-institutional transmissions (ec-transmissions-v1.1.xsd) or transmissions for publications between the European Commission and the OP (ec-publications-v1.1.xsd). The two schemas mixed Level 1 and Level 2 metadata together in the same schema and containing two different schemas: one for transmissions to other bodies and other for transmissions to the OP for the purpose of publication)*
    - *As the IMMC v2 project evolved it adopted the modern file structure which separated "common and core metadata" w.r.t domain-specific data (Level 1 metadata in the root folder and Level 2 metadata in the domain specific folder). Moreover, further developments in IMMC v2 implied that the transmission (for publications) to the OP refer to elements in the publication_request.xsd schema or enrich these by defining extensions at the level of the specific-domain-level transmission (e.g. t_publication_request_extension at the cortrans level).*
    - *By 2022 the schemas were aligned to IMMC v2 and only three differences remained:*
        - *two schemas (the v1.1) are in the public_access eu folder (as, in line with the stable URI feature and retro-compatible messages, they could not be removed.*
        - *The domain-specific extension publications.xsd schema extends the transmission schema and is implemented as a separate schema (instead of as an element in the domain-specific transmission schema).*
        - *The elements used in IMMC v1.1 (a subset of those in IMMC v2) are described in the Inter-institutional Transmission Format, not in the communication protocol.*