Governance - IMFC: Previously known as IMMC, the committee then expanded to include formats (e.g., AKN4EU and VocBench). It is responsible for the harmonisation and standardisation of metadata to facilitate information exchanges amongst actors at the European level.

IMMC basics - Communication protocol: A document which describes the IMMC schema-line, specifying the rules for the exchange of messages between institutions, Publications Office and contractors alike. It provides an overview of the principles followed when implementing the IMMC protocol, such as package naming conventions and common interactions between sender and receiver. Overall, it describes the core properties shared among all domains of that schema line.

IMMC basics - Data channel: For a given exchange domain, once the user selects the receiver of the message, they can communicate with them through a unique data channel. Indeed, the sender can send one or more messages through this channel, each uniquely identified by a **transmission identifier**. Generally, the messages are sent through secure protocol, e.g. SFTP, e-Trustex…

IMMC basics - Exchange domain: set of contracts that support one business function/process but may support different sender-receiver configurations. For example PRELEG is used by the Commission, but also by the Council of the Regions and the Parliament (among others). This means that one exchange domain can support different domain-specific transmission protocols and contains different domain-specific extensions.

IMMC basics – For a given **communication protocol** (generic agreed way of exchanging information), a sender can send messages (through a **data channel)** which follow a specific contract (delineated in the **transmission protocol** or its partial implementation, the domain-specific XSD) to the intended receiver. The messages can be partially automatically validated and must contain an **IMMC descriptor** (which must follow the rules delineated in the transmission protocol). The specific contract will depend on the business purpose of the exchange and the sender-receiver pair. The **IMMC dataflow inventory** provides an overview of all of the data channels using IMMC.

IMMC basics - IMMC (data)flow inventory: list of flows of information between a sender and a receiver in the context of a specific purpose (business domain) and IMMC schema version. Keeping track of all the flows, the inventory is a register of all of the contracts for the exchange of metadata (transmission protocols) specified between institutions, bodies and agencies. This document therefore implicitly identifies the domain specific extensions and metadata items to be used for a given business domain. By keeping track of this, the OP also has further visibility of

the dependencies of metadata for all business domains and therefore can better manage the evolution of the schemas. For example, if an element in a domain-specific extension is requested by other institutions, the OP may consider placing them in the unique cm-common-extension folder (thereby reducing duplication and maintenance efforts).

IMMC basics - IMMC descriptor: also known as XML schema instance. For a given purpose, sender and receiver the IMMC descriptors must comply with the transmission protocol. Descriptors are composed of items (selected by the vocabulary) which comply with syntactical rules (formatting of items, hierarchical structure of elements...) set out in the transmission protocol.

IMMC basics - IMMC exchanges: For a given domain, the IMMC exchanges are a list of contracts that can go from a sender-receiver. These exchanges are captured in the IMMC schema-lines communication protocol.

IMMC basics - IMMC message: also known as IMMC package or IMMC bundle depending on the sending/receiving IT systems. Messages exchanged between two IT systems must contain an IMMC descriptor file among other optional files referenced in the descriptor (e.g. manifestations such pdfs, pictures and Formex or files providing additional context/feedback to the transmission e.g. acknowledgement/receipt responses). When these files are packaged in ZIP format archives we talk about IMMC packages, while if they are bundled together (e.g. e-Trustex) we talk about IMMC bundles.

IMMC foundation - Controlled Vocabularies: provide a consistent way to describe data. Notable examples are code or named authority lists (see Authority Table) which harmonise codes and associated labels used in various environments, and thesauri and taxonomies (which provide hierarchical structures for concepts).

IMMC foundation - Transmission protocol: this document specifies the information which can or must be exchanged in the message for a given transmission. It is agreed upon a priori and is specific to a given purpose and sender-receiver couple. The protocol describes the different messages (instances) which can happen in an exchange domain. It must specify at least the following elements: sender, purpose of the exchange, recipient, acceptable values for the metadata, metadata items used and metadata schema-line in use

IMMC foundations - Authority Table: also known as **named** authority list (NAL). A named authority list or code list is a list of acceptable values for a given semantic concept (e.g. list of country codes, list of language codes). NALs contain named values (e.g. FR) while code lists

contain identifiers (e.g. c_c93e946d). The here possible, the IMMC protocol refers to authority lists (instead of strings).

IMMC foundations - FRBR: Functional Requirements for Bibliographic Records. Conceptual model developed by the International Federation of Library Associations and Institutions (IFLA). For digital items, **Group 1** entities are work, expression, manifestation (WEM). The **work** also known as the idea, or intellectual endeavor (e.g. directive), can be realized in one or multiple expressions (e.g. language). Each **expression** is then embodied in one or more **manifestations** (e.g. pdf, html). **Group 2 entities** are responsible for the custodianship of Group 1's endeavor. Examples of such are corporate bodies, CIDOC elements (sets of documents which can be related to events) and FRBRoo elements (temporal/circumstantial elements such as places, authors, timestamps).

IMMC foundations - IMMC Vocabulary: a collection of all the possible metadata concepts which can be used in an IMMC exchange. The concepts are composed of elements from rdf (e.g. properties), rdfs (subproperties), owl (equivalences) , skos (preferred labels and definition) and dcat(modified, created, ispartof). The concepts (available on **Vocbench**) either describe the metadata of documents or the processes of the institution. An example of such is the concept analysisDecision which is a subproperty of decision and is equivalent to V3 caselawext:analysis_decision, it was last-modified on 2021-10-19T12:19:44. As the IMMC vocabulary implicitly defines what can be transported/will be transported, these concepts (words in the vocabulary) form the basis for the elements in the Level 1 (core and common metadata), Level 2 (domain-specific metadata) and transmission documents.

IMMC implementation - IMMC schema-line: also known as IMMC series. To date there are two schema lines: IMMCv2, IMMv3. The legacy IMMC Public Access used by the Commission has been integrated into IMMCv2.

IMMC implementation- IMMC Schema: schema used to validate an IMMC descriptor against constraints formulated in XML schema notation. This file is the partial implementation of the transmission protocol.

IMMC structure - Common extensions: metadata placed at this level extends the core-metadata and is used by multiple stakeholders.

IMMC structure - Core metadata transmission (cmt): administrative metadata about the processing rules/actions to be performed (workflows). For example, the phase workflow element is used to facilitate the handling and dispatching of messages to the correct workflow, i.e.

depending on the specific phase of production. The cmt file can be extended for specific domains through inheritance (e.g. General Publication transmission) and/or by extending specific 'extension' elements specified in the core metadata.

IMMC structure - Core metadata: minimum set of metadata elements related to the legal decision-making process. These elements are applicable to all domains, used by all stakeholders and defined in the core-metadata (cm) extension.

IMMC structure - Domain specific extensions: contextual metadata about the transmitted documents or the events associated to that procedure.

IMMC structure - Domain specific transmission schema: also known as the implementing schema. Implements the transmission protocol in XML format, grouping together a domain-specific root element which extends the Core metadata transmission, with the FRBR-inspired extensions specific to the same domain.

IMMC validation - Business (content) validation: To date, this validation is both manual and automatic. Regarding **automatic validation** OP's receiving system validates the content of the IMMC descriptor and any data files included in the package. For example: for XMLs an external module to CERES(metadata controller program) validates the contents of the XMLs against rules set in the Schematron, while for PDFs, IT tools such as Callas verify the compliance of the delivered pdf (against technical requirements). Regarding **manual validation** both the IMMC descriptor and other files may undergo this check. For manuscripts and other files, the receiver performs various controls to accept/reject delivery package after the automatic validation step. For IMMC descriptors, this validation generally does not occur actively and is triggered when issues are raised: e.g. some values (or the combination of values) in the IMMC descriptor are not valid for the context of the specific transmission/use-case.

IMMC validation - Structural validation: Can be performed manually or completely automated. Verifies the structure of the IMMC package (e.g. mandatory contents, package naming and conventions, correct parsing of IMMC descriptor) and validates the descriptor against the schema (checking if the elements follow the pre-determined structure specified in the XSD schema).

IMMC validation - Manual validation: can occur after the packages are ingested in CELLAR. Involves people manually checking the manuscripts or documents which were sent (e.g. Directive to be published on the OJ).

IMMC validation – automatic validation: Regarding the transmitted metadata, XSD schemas are run against the descriptors to validate them. For other files, OP's receiving system is configured to automatically validate technical features (e.g. checksum..) and detect incoherences with the descriptor.

Information Systems - CELLAR: OP's common data repository storing all metadata and digital content (pdfs..) managed by the OP in a harmonized and standardized format. It can be accessed via REST interfaces or through the EUR-LEX and OP portal.

Information Systems - CERES:. This IT system receives all the incoming IMMC packages, validates them, reroutes them to the correct workflow and may load the content received by the OP in CELLAR (through IMMC flows). Monitoring this system is done by the OP-RECVAL system.

Other protocols - OJEEP: Legacy protocol used between OP and the contractors in the context of OJ publications. The packages which were sent consisted of preformatted messages implemented in XML (such as reception messages) along with content files. Depending on the business use-case, the sender would select the correct preformatted message, and the OP performed technical validation on the message by validating it against the schema.

Metadata - Administrative metadata: metadata which covers the following aspects: provenance metadata (sender, recipient, date of transmission), technical metadata (e.g. checksums), intellectual property metadata and process metadata (e.g. to processing information to direct to the correct destination). Administrative metadata is used to manage resources and can provide details related to technical domains (e.g. decoding/rendering files), preservation (e.g. archiving) and long-term rights and use domains, answering to the following questions how the asset was created, who the owner of the asset is and how can the asset be used/until when). In IMMC, administrative metadata is specified at the transmission-level, and added on top of the descriptive metadata.

Metadata - Descriptive metadata: Descriptive metadata is used to discover, identify and select resources. It can include elements such as title, author, subject. In IMMC examples of such as work, procedure, session, dossier.

Metadata - Level 1 Metadata: mainly comprises of core and common-metadata. This metadata is specified at the root of the folder and is shared between at least some domains.

Metadata - Level 2 Metadata: also known as domain-specific files. In the domain-specific folder, two XML files cover Level 2 Metadata which comprises of

- Level 1 metadata extended or re-defined
- Level 2 metadata uniquely relevant for this business domain/domain-specific transmission.

Metadata - Structural metadata: describes the internal structure of resources, for example – how pages in a document are organised (e.g. headings..). IMMC does not contain structural metadata elements as it does not describe the structure of the assets. On the other hand Formex is used for such purposes.

Metadata: data providing information about aspects of the data (e.g. purpose, means of creation..)

Syntax: Data structures formulated in the IMMC schemas.

Systems - EUR-LEX: web portal providing access to EU Law.

Terminology – IMFC : Interinstitutional Metadata and Formats Committee

Terminology - IMMC: or IMMC protocol. Interinsitutional Metadata Maintenance Committee protocol.

Terminology - IS: Information Systems

Terminology - NAL: Named Authority List (see Authority Tables)

Terminology - OJEEP: Official Journal Electronic Exchange Protocol.

Terminology - OP: Publication Office/Office de Publication.

Terminology - URI: Uniform Resource Identifier (path to allow the IS to identify a given resource)

Terminology - URL: Uniform Resource Locator (path to allow the user to access a given resource, sub-part of URI).

Terminology - XML: Extensible Markup Language

Terminology – CERES: Central Electronic REception System of the OP

Tools - "IMMCBuilder" tool: a service, offered by the OP to institutions, which constructs well-defined and valid IMMC packages. Once the user has entered the data in the web form and submitted it, the tool constructs IMMC packages.

Tools – metadata controller tool: Logic-based validation module which makes use of Schematrons to validate XMLs. Used by OP and more specifically CERES.

Tools - Vocbench: Web-based multilingual collaborative development platform to manage ontologies and thesaurus among others (notably the IMMC vocabulary).

Tools - XML Validating Parser: A program (application) which can read and extract the necessary data from the XML document (which will then be compared against the specific XML validating schema). The parser will start reading the IMMC message and look for the file against it has to validate the descriptor (by looking at the schemaLocation attribute). The **schemaLocation** attribute should point to the exact version of the contract the message should conform with (*transmission.xsd). If this attribute is not mentioned, the parser will automatically select a default validating schema and try to validate the message.

Transmission identifier: Identifier for the IMMC descriptor (must be unique for a given exchange channel) to ensure that no message is lost or overwritten.