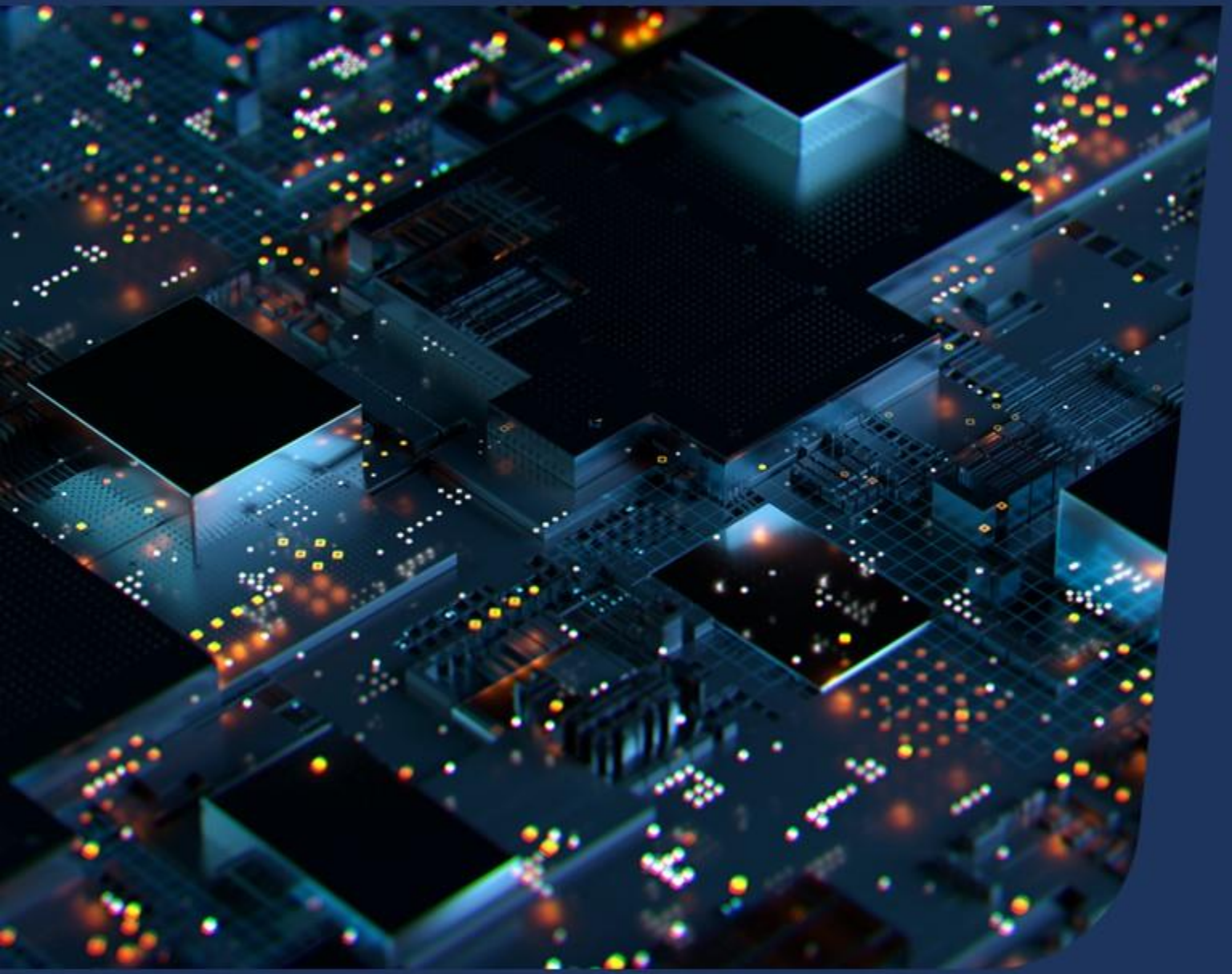


eArchiving Webinar #12

Putting the CITS content in the CSIP package

Start 10:00 (CET)

27 May 2021

Agenda

10:00 – 10:05

Welcome

Pawel Stech – CEF Stakeholder Management Office – DIGIT

10:05 – 10:15

eArchiving Building Block welcome

Dr Jaime Kaminski – eArchiving Building Block training lead

10:15 – 11:15

eArchiving: Putting the CITS content in the CSIP package

Karin Bredenberg – Kommunalförbundet Sydarkivera

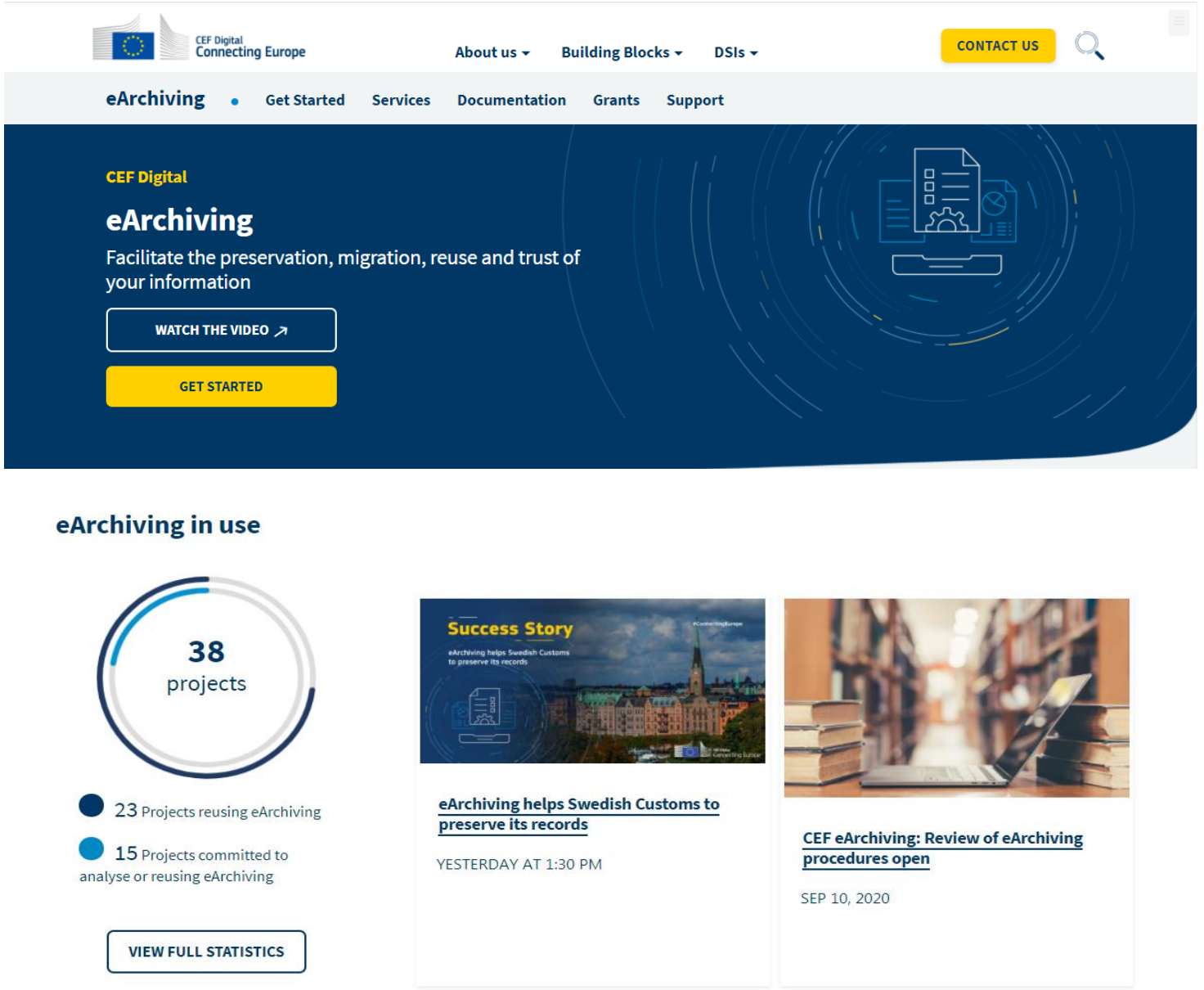
11:15

Q&A

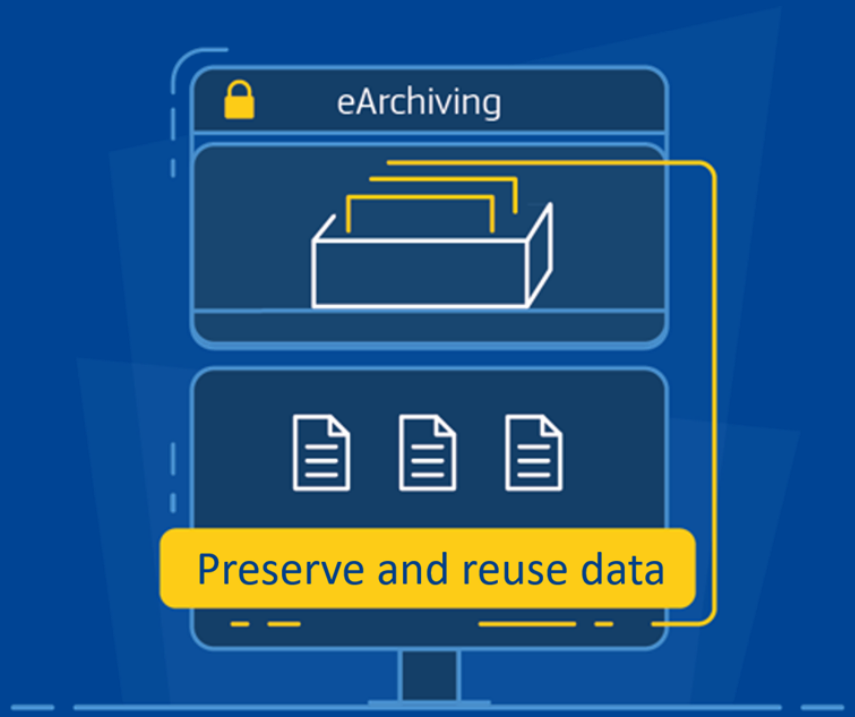
Welcome to the CEF eArchiving Building Block

Dr Jaime Kaminski
eArchiving Building Block training lead

eArchiving Building Block website



eArchiving Building Block



eArchiving services:

- Technical specifications
- Sample software
- Compliance/validation
- Service Desk
- Outreach/community engagement
- **Training**
  - Webinars
  - Videos
  - Moodle LMS training modules

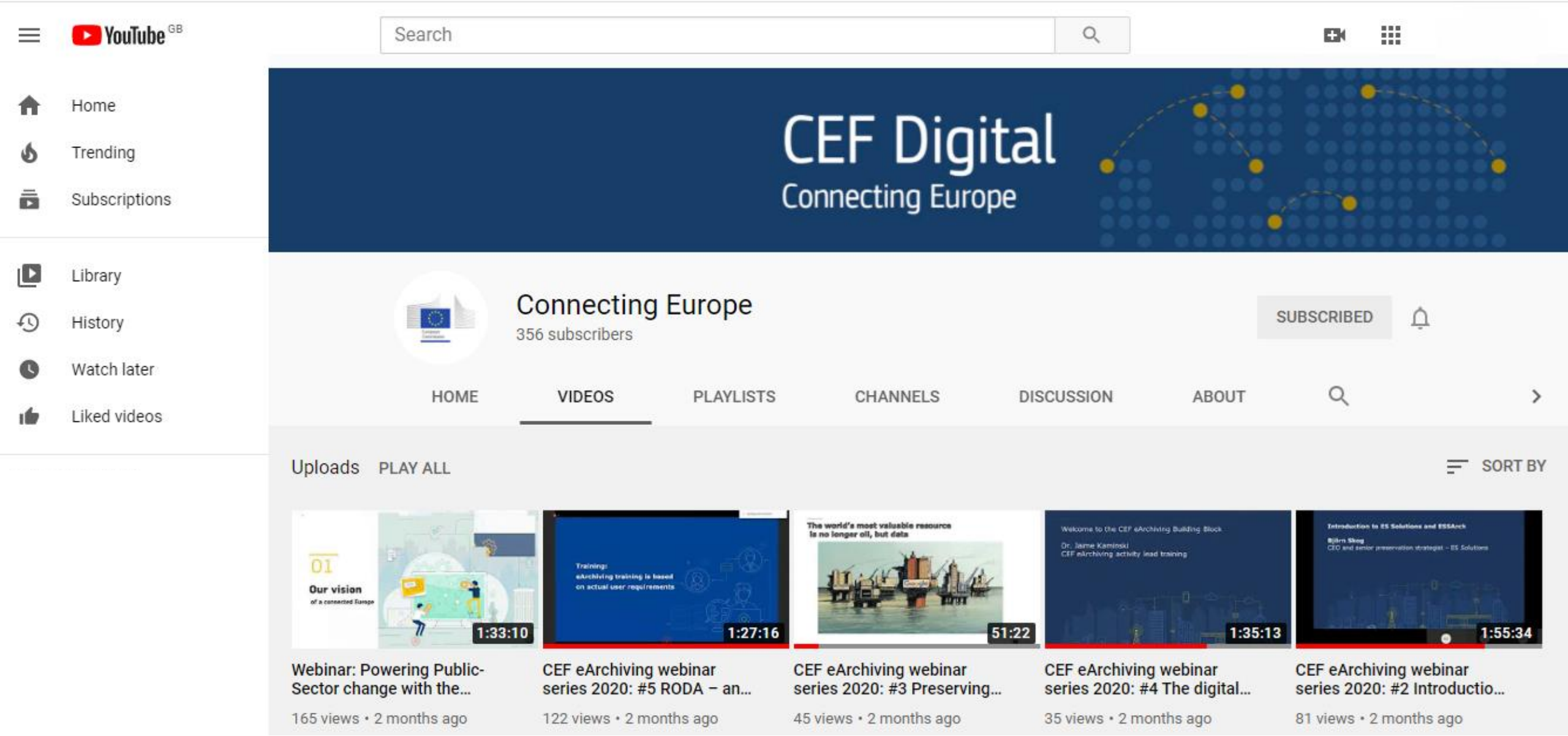
No.	Month	Date	Title	Lead(s)	Organisation
No. 13	June	10	Practical applications of digital archiving – (validation)	Björn Skog	ES Solutions
No. 14	June	24	Practical applications of digital archiving – (submission agreements)	Björn Skog	ES Solutions
Summer break					
No. 15	September	16	The eArchiving Reference Architecture	István Alföldi	Poliphon
No. 16	September	30	E-ARK Validation: What's inside the package?	Carl Wilson/Costas Simatos	OPF/CEF
No. 17	October	14	CITS Geospatial	Gregor Završnik	Geoarh

eArchiving training webinars 2021

WEBINARS : AGENDA & RECORDINGS

Session	What you'll learn	Date & Time*	Webinar presentation & recording	Link to Q&A
Webinar #1: Introduction to CSIP	• CEF eArchiving welcome • Why have a common standard? • Core principles for an information package • Elements and attributes used for describing a package • Extending CSIP to meet more needs • METS in E-ARK CSIP	• 27th February 2020: 10:00 - 11:00	 Recording: CEF webinar CSIP	CEF Webinar #1: Q&A
Webinar #2: Introduction to ESSArch – an open source-based solution for long-term preservation of digital information	• CEF eArchiving welcome • Introduction to ESS and ESSArch • Pre-Ingest and Ingest • Archival and Data Management • Access and Portal • Reports, Statistics, Monitoring and API • Configuration and Administration • ESSArch Installation procedures	• 26th March 2020: 10:00 - 13:00	 Recording: Part 1 - Part 2 - Part 3 - Part 4- Part 5	CEF Webinar #2: Q&A
Webinar #3: Preserving digital geospatial records	• CEF eArchiving welcome • Geospatial data and its role in organisations • How could you benefit form E-ARK specifications for geospatial data preservation? • Strategies for implementing an accessible geospatial records archive • Proactive preservation in new and existing systems	• 23rd April 2020: 10:00 - 11:15		CEF Webinar #3: Q&A

https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving+webinar+Series+2020

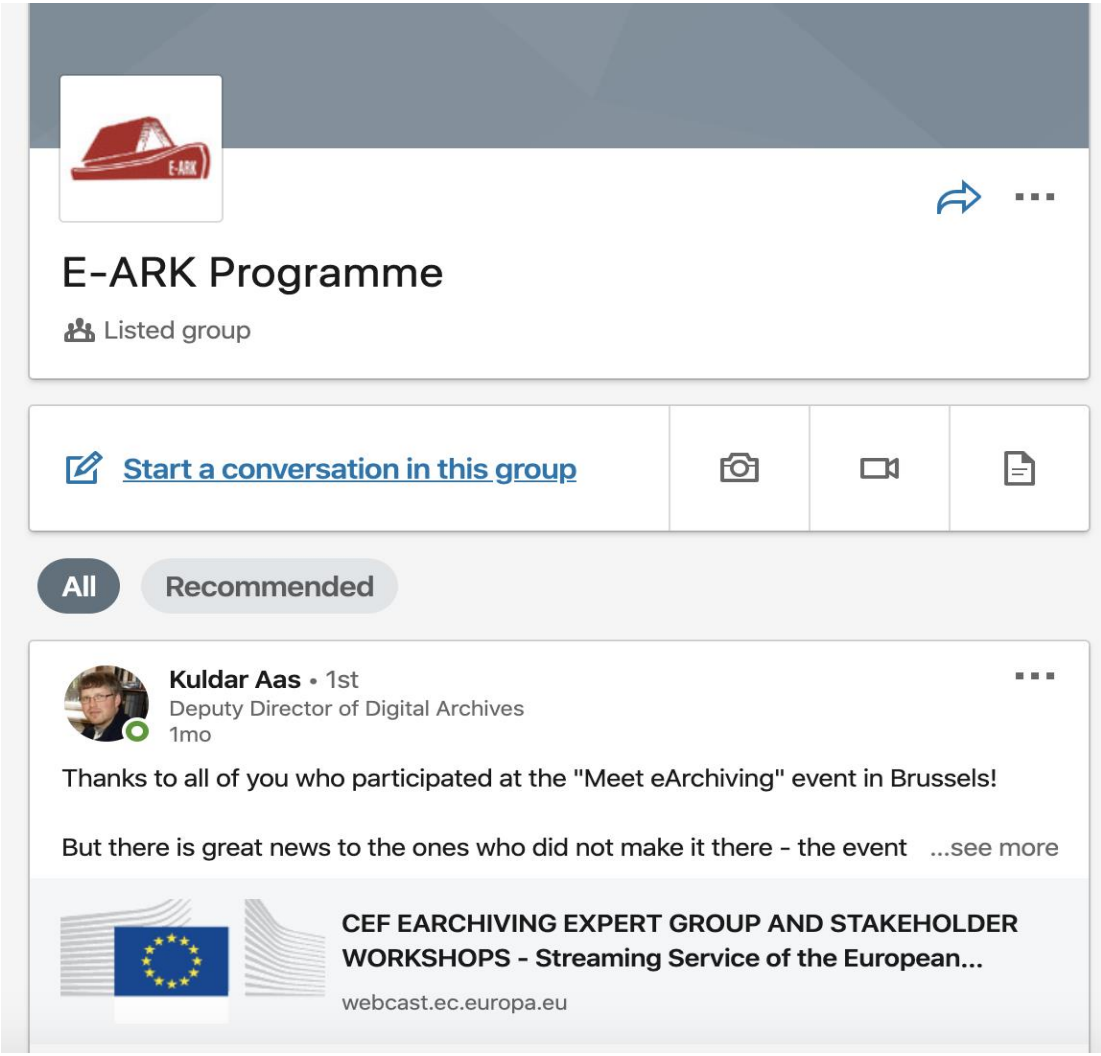


https://www.youtube.com/channel/UCaPOT_MBdE-kL5AJQzrCBDw/videos?view=0&sort=dd&flow=grid

eArchiving outreach

- Webinars
- Workshops
- LinkedIn group
- Twitter #EARKProject

LinkedIn Group: E-ARK Programme
https://www.linkedin.com/groups/8343650/



Our speaker

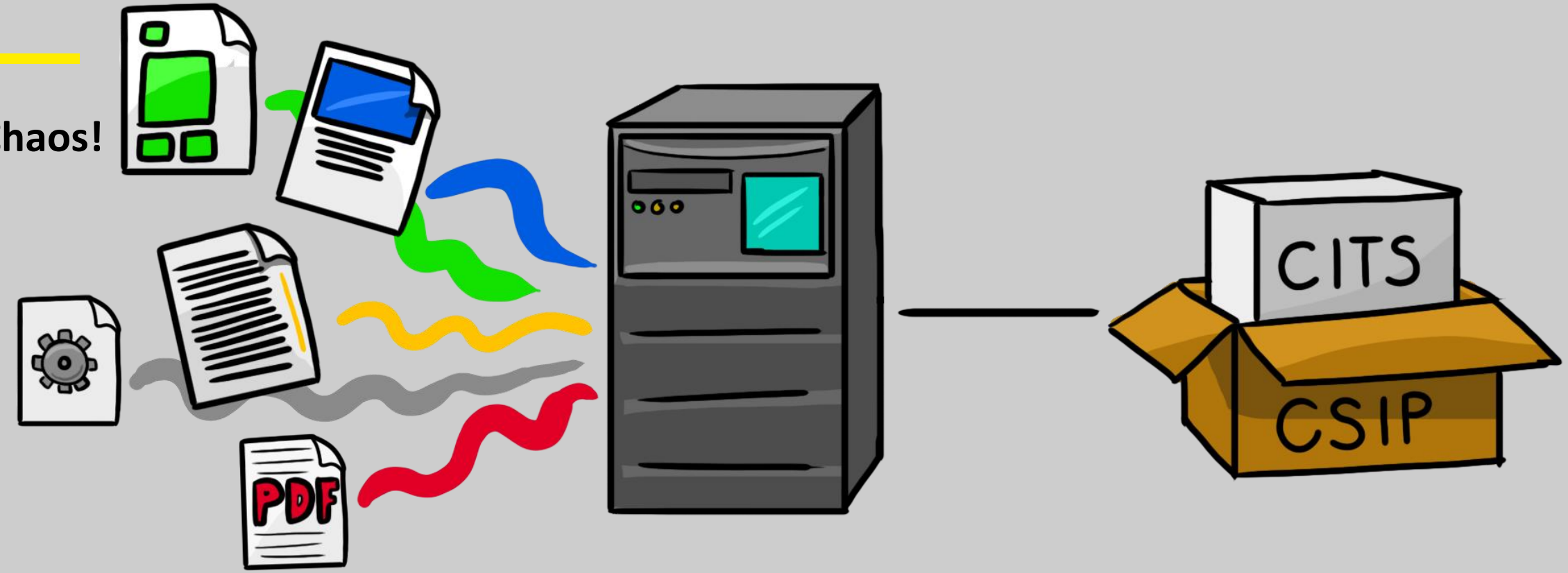
Karin Bredenberg
Kommunalförbundet Sydarkivera

01

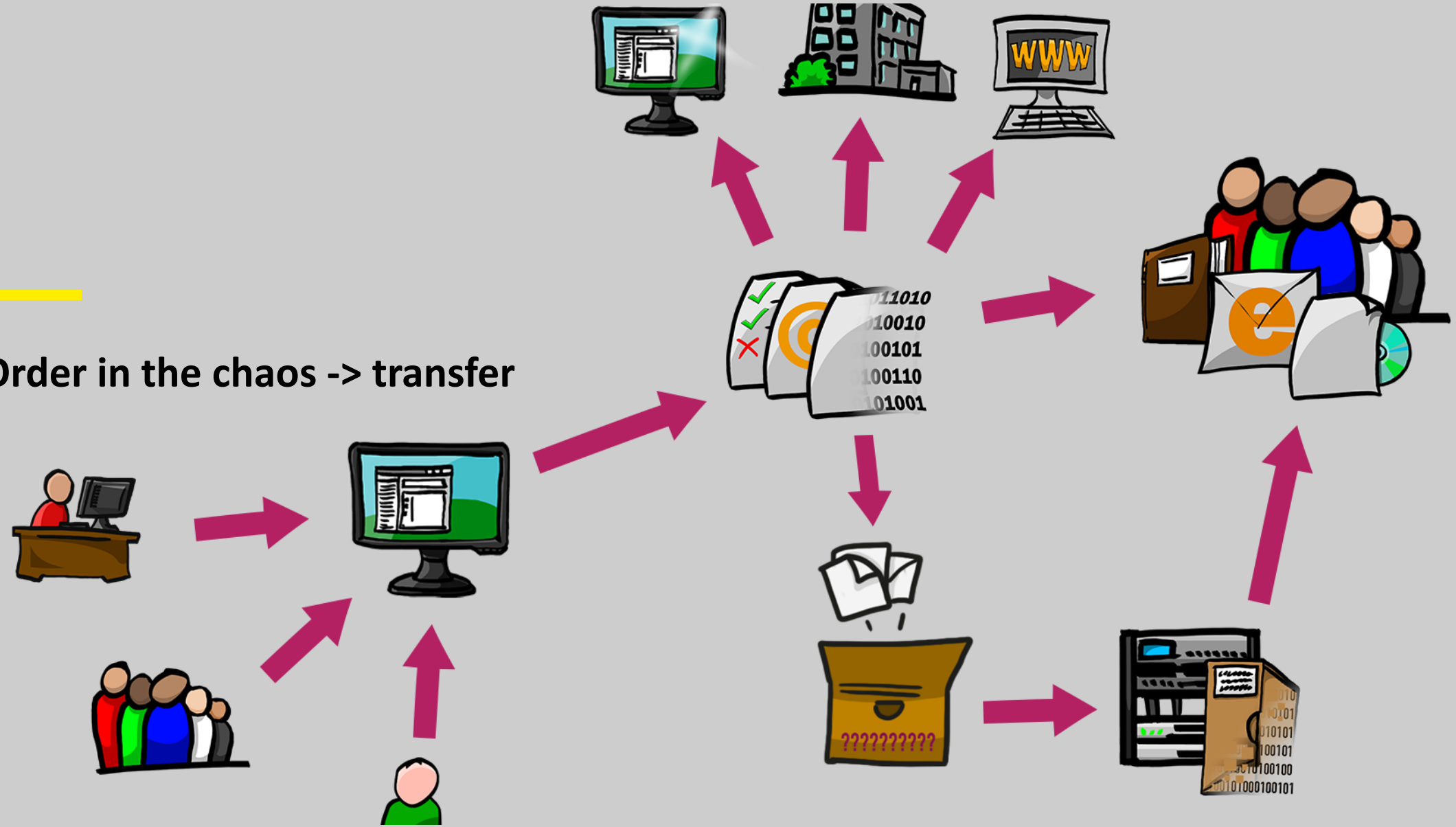
Specifications

Putting the CITS content in the CSIP package

Chaos!

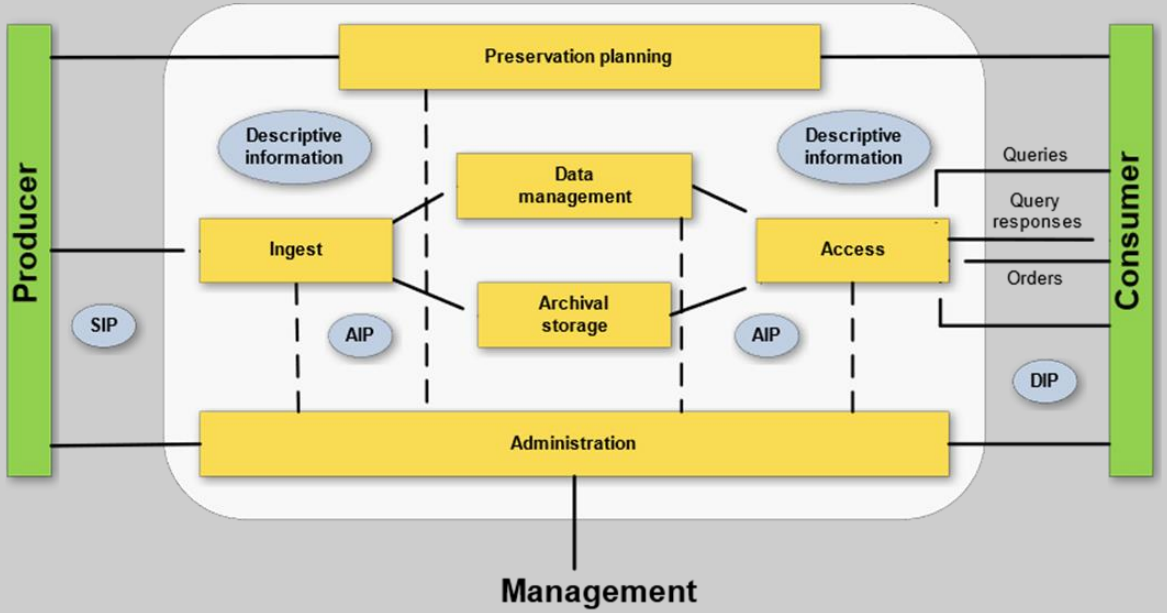


Order in the chaos -> transfer



The OAIS Reference Model

https://www.iso.org/standard/57284.html



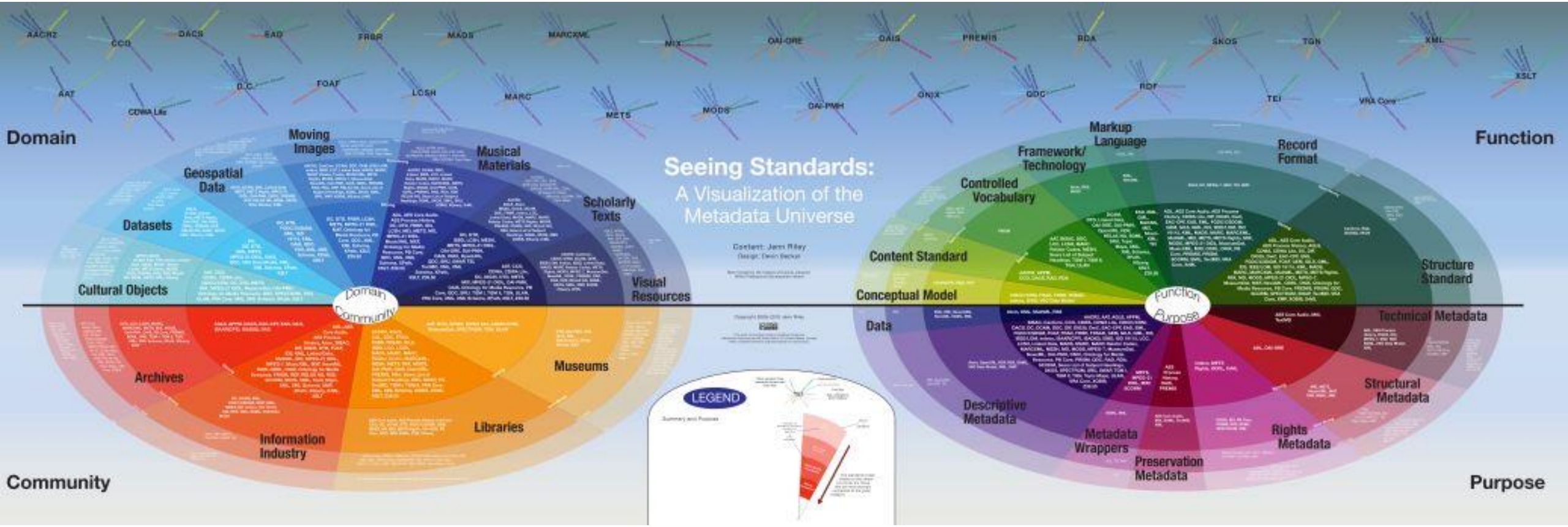
The **specifications**

https://github.com/DILCISBoard https://dilcis.eu/ https://dilcis.eu/reviews/current-reviews

The basics for both CS and CITS; using standards

There are plenty of standards to use for data transfer and conformance

http://jennriley.com/metadatamap/



Currently XML is used for the descriptions

We are monitoring what is happening and will evaluate alternatives as they emerge but it takes time, stability!

Extensible Markup Language (XML)

Standards currently used

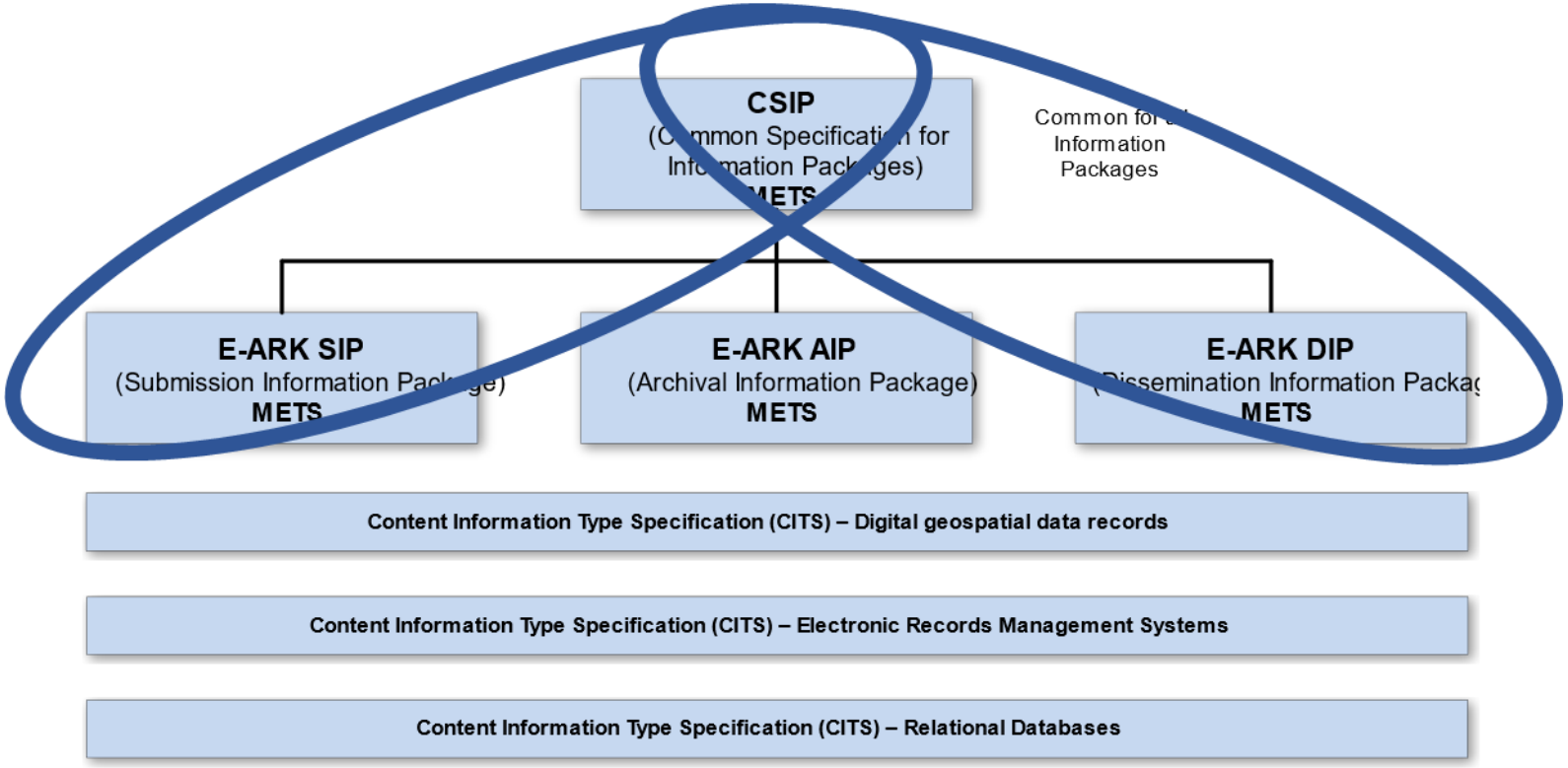
For the package we use the Metadata Encoding & Transmission Standard, METS, http://www.loc.gov/standards/mets/

The number of standards used increases with each additional CITS

There are two different types of specifications; the Information Package Specifications (CS) and the Content Information Type Specifications (CITS)

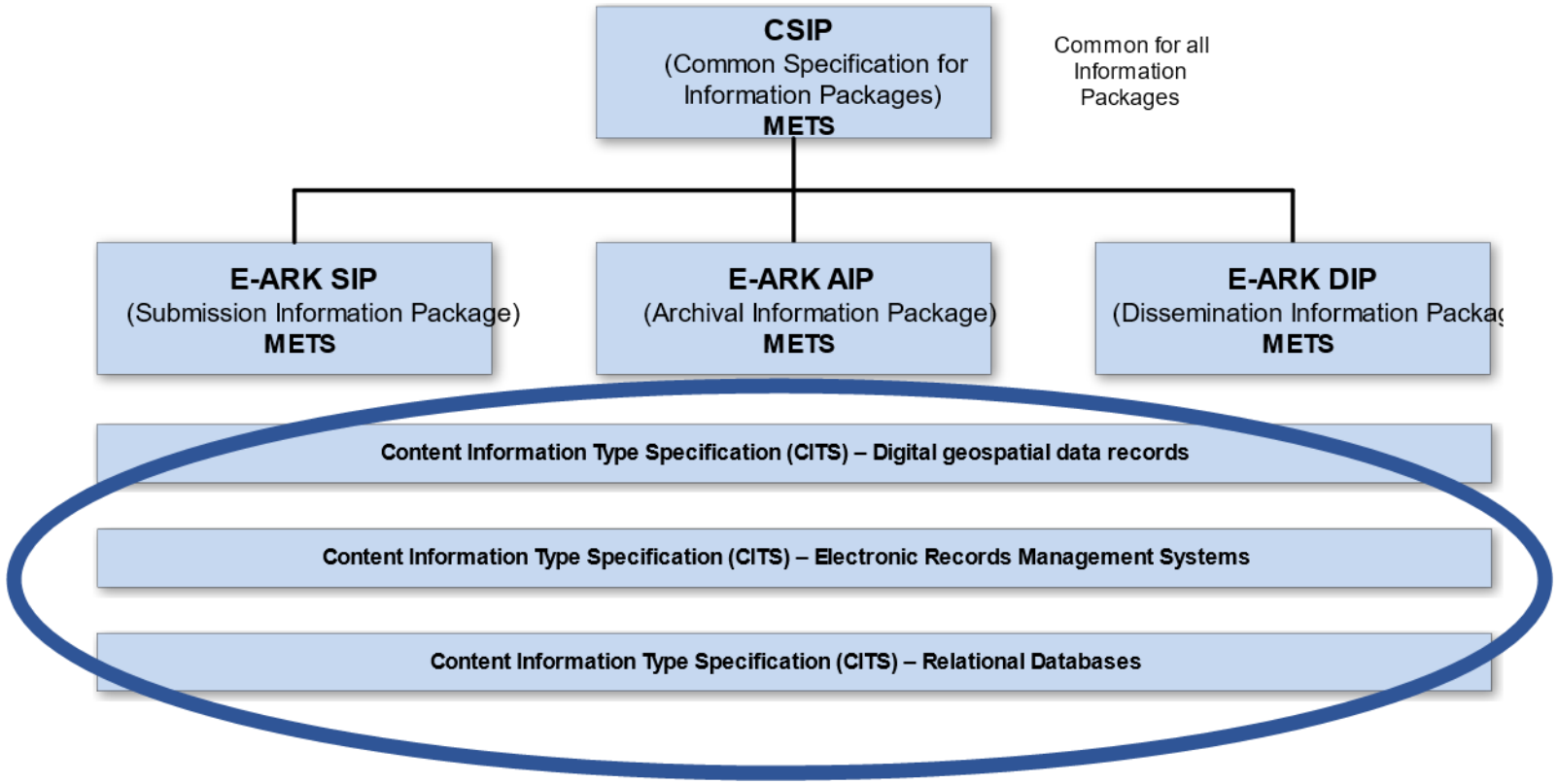
CSIP, SIP, AIP and DIP

The different Information Packages in the OAIS Reference Model



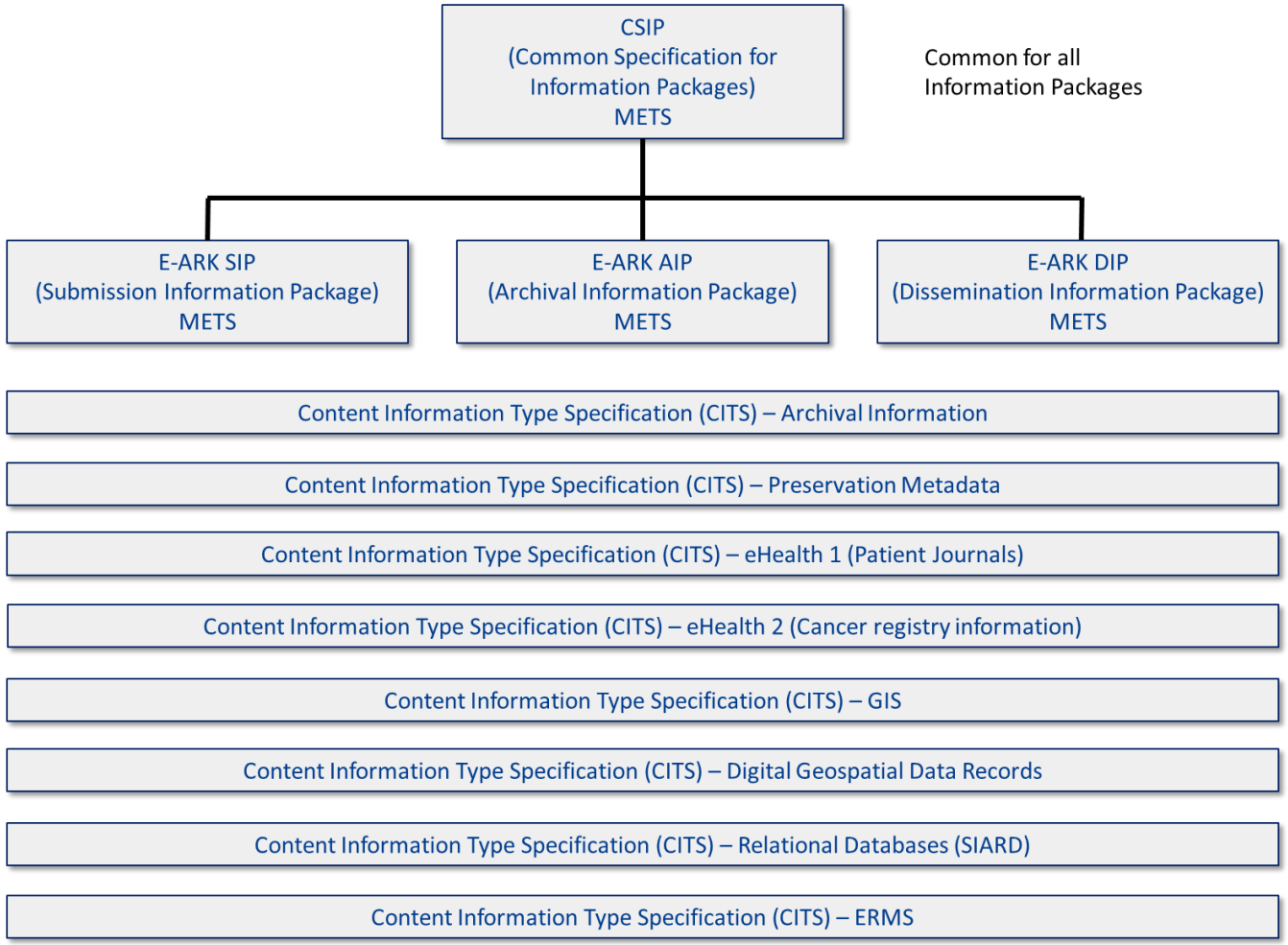
The Content Information Type Specifications (CITS)

The data/information/content in the package



The Content Information Type Specifications (CITS)

The ongoing work extends the number of CITS

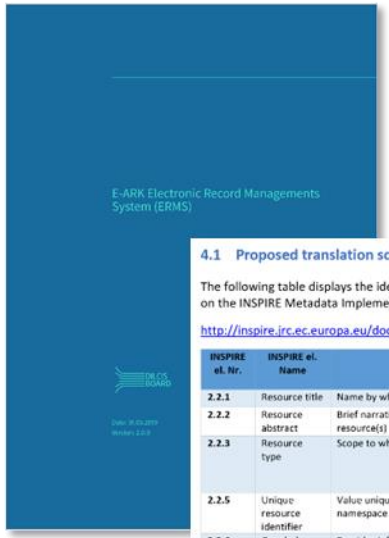


Reading a specification; a text with rules

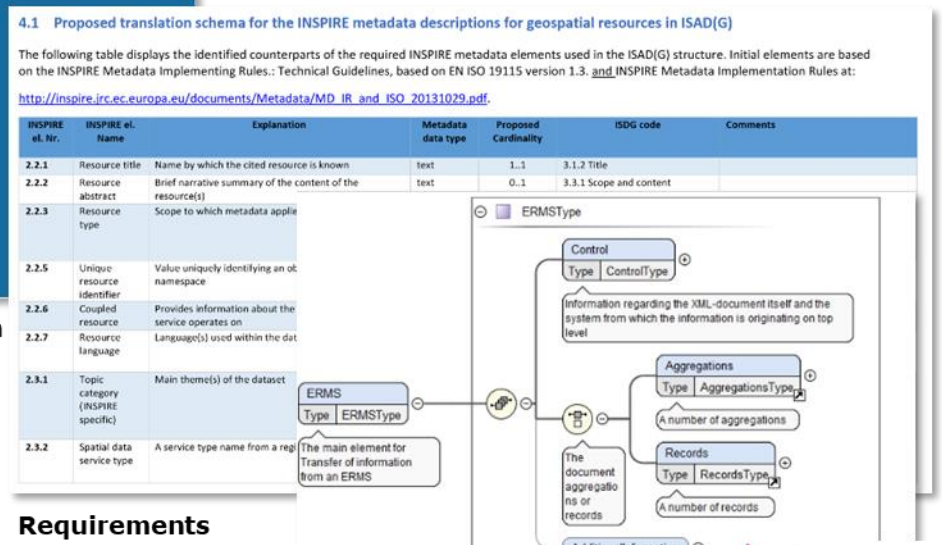
4

Document and files for a specification

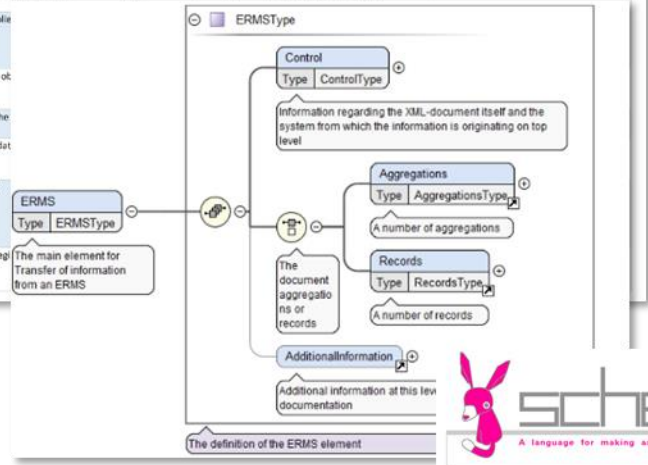
The text document and files providing the XML structure and validation rules



Specification



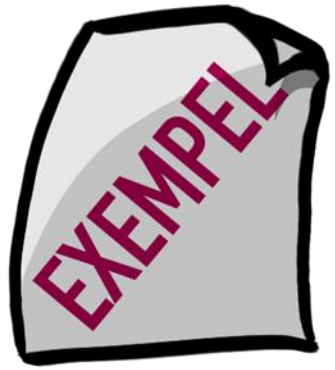
Requirements



XML-Schema/Schemas

Expression of requirements not possible to make in the XML-schema

Guideline



Example

Tables are used

The tables are used to express the requirements

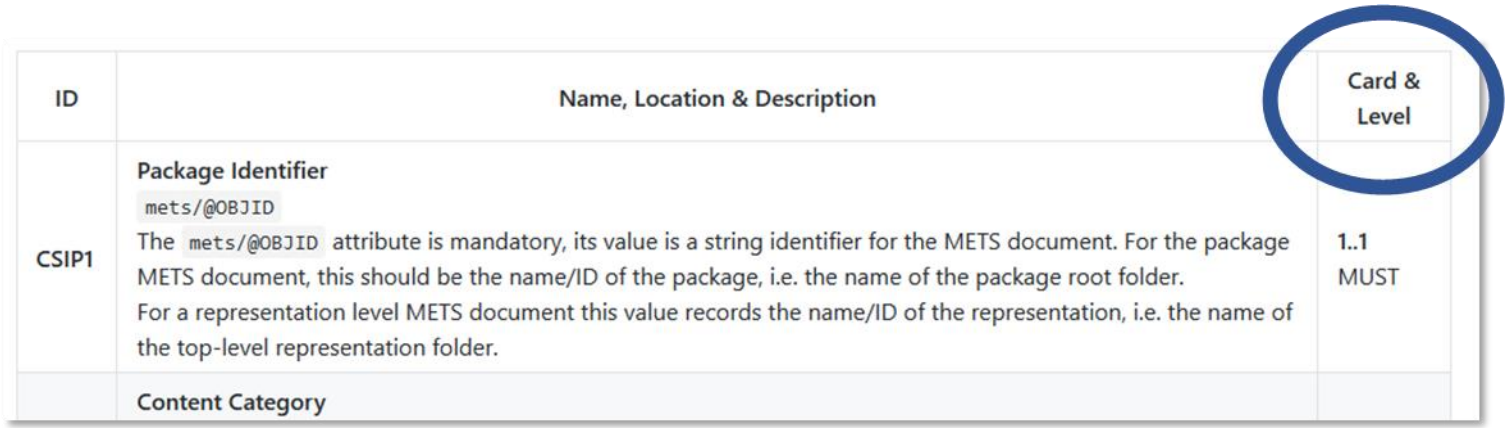
ID	Name, Location & Description	Card & Level
CSIP1	**Package Identifier** `mets/@OBJID` The `mets/@OBJID` attribute is mandatory, its value is a string identifier for the METS document. For the package METS document, this should be the name/ID of the package, i.e. the name of the package root folder. For a representation level METS document this value records the name/ID of the representation, i.e. the name of the top-level representation folder.	**1..1** MUST
CSIP2	**Content Category** `mets/@TYPE` The `mets/@TYPE` attribute MUST be used to declare the category of the content held in the package, e.g. book, journal, stereograph, video, etc.. Legal values are defined in a fixed vocabulary. When the content category used falls outside of the defined vocabulary the `mets/@TYPE` value must be set to "OTHER" and the specific value declared in `mets/@csip:OTHERTYPE`. The vocabulary will develop under the curation of the DILCIS Board as additional content information type specifications are produced. **See also:** Content Category	**1..1** MUST
CSIP3	**Other Content Category** `mets[@TYPE='OTHER']/@csip:OTHERTYPE` When the `mets/@TYPE` attribute has the value "OTHER" the `mets/@csip:OTHERTYPE` attribute MUST be used to declare the content category of the package/representation. **See also:** Content Category	**0..1** SHOULD
CSIP4	**Content Information Type Specification** `mets/@csip:CONTENTINFORMATIONTYPE` Used to declare the Content Information Type Specification used when creating the package. Legal values are defined in a fixed vocabulary. The attribute is mandatory for representation level METS documents. The vocabulary will evolve under the care of the DILCIS Board as additional Content Information Type Specifications are developed. **See also:** Content information type specification	**0..1** SHOULD
CSIP5	**Other Content Information Type Specification** `mets[@csip:CONTENTINFORMATIONTYPE='OTHER']/@csip:OTHERCONTENTINFORMATIONTYPE` When the `mets/@csip:CONTENTINFORMATIONTYPE` has the value "OTHER" the `mets/@csip:OTHERCONTENTINFORMATIONTYPE` must state the content information type.	**0..1** MAY
CSIP6	**METS Profile** `mets/@PROFILE` The URL of the METS profile that the information package conforms with.	**1..1** MUST

Example: METS root element showing use of `csip:@OTHERTYPE` attribute when an appropriate package content category value is not available in the vocabulary. The `@TYPE` attribute value is set to OTHER.

```
<mets:mets OBJID="uuid-4422c185-5407-4918-83b1-7abfa77de182" LABEL="Sample CSIP Information Package" TYPE="OTHER" OTHERT
</mets:mets>
```

The key terms when reading a specification

Understanding occurrence and obligation



ID	Name, Location & Description	Card & Level
CSIP1	**Package Identifier** `mets/@OBJID` The `mets/@OBJID` attribute is mandatory, its value is a string identifier for the METS document. For the package METS document, this should be the name/ID of the package, i.e. the name of the package root folder. For a representation level METS document this value records the name/ID of the representation, i.e. the name of the top-level representation folder.	**1..1** MUST
	Content Category	

- Card = Cardinality answers:
  - How many times can an element or attribute occur?
  - Is the element or attribute mandatory or optional?
- Level answers:
  - If the element or attribute is not mandatory what are the recommendations?

The Information Package **CS**

The package principles and requirements

The principles of a package

What makes a package a package and how do we identify it and its content?

3.2. Identification of the Information Package

3.2.1. Principle 2.1:

The Information Package OAIS type (SIP, AIP or DIP) ***MUST*** *be clearly indicated.*

One of the first tasks in analysing any Information Package is to identify its current status in the overall archival process. Therefore, any Information Package must explicitly and uniformly identify itself as a SIP, AIP or DIP.

3.2.2. Principle 2.2:

Any Information Package ***MUST*** *clearly identify the Content Information Type(s) of its data and metadata.*

As stated in Principle 1.1, any Information Package MUST be able to include any kind of data and metadata. At the same time, we have introduced in earlier Sections the concept of Content Information Types which allow users to achieve more detailed control and fine-grained interoperability. As such, any CSIP Information Package MUST include a statement about which Content Information Type Specification(s) has been followed within the Information Package, or on the contrary, indicate clearly that no specific Content Information Type Specification has been followed.

The practical implication of principles 1.1, 2.1 and 2.2 is that, once these have been followed in implementations, it is possible to develop modular identification and validation tools and workflows. While generic components can carry out high-level tasks regardless of the Content Information Type, it is possible to detect automatically which additional content-aware modules need to be executed.

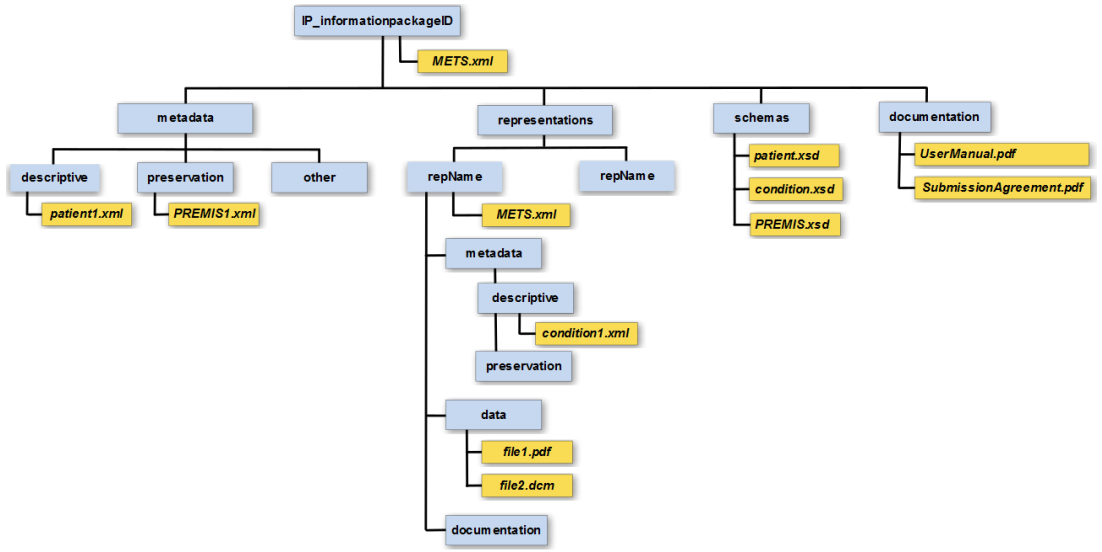
The folder structure of a package

If we do not have a manifest, we still need to be able to understand the package

4. CSIP structure

4.1. Folder structure of the CSIP

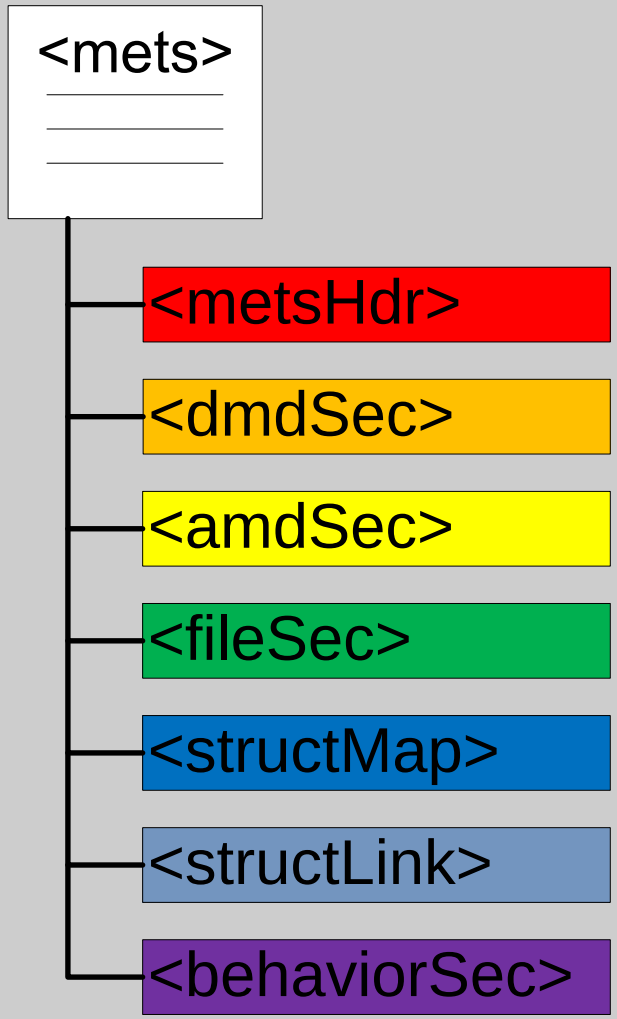
4.2. Implementing the structure



METS - different packages CSIP, SIP, AIP and DIP and their relationships

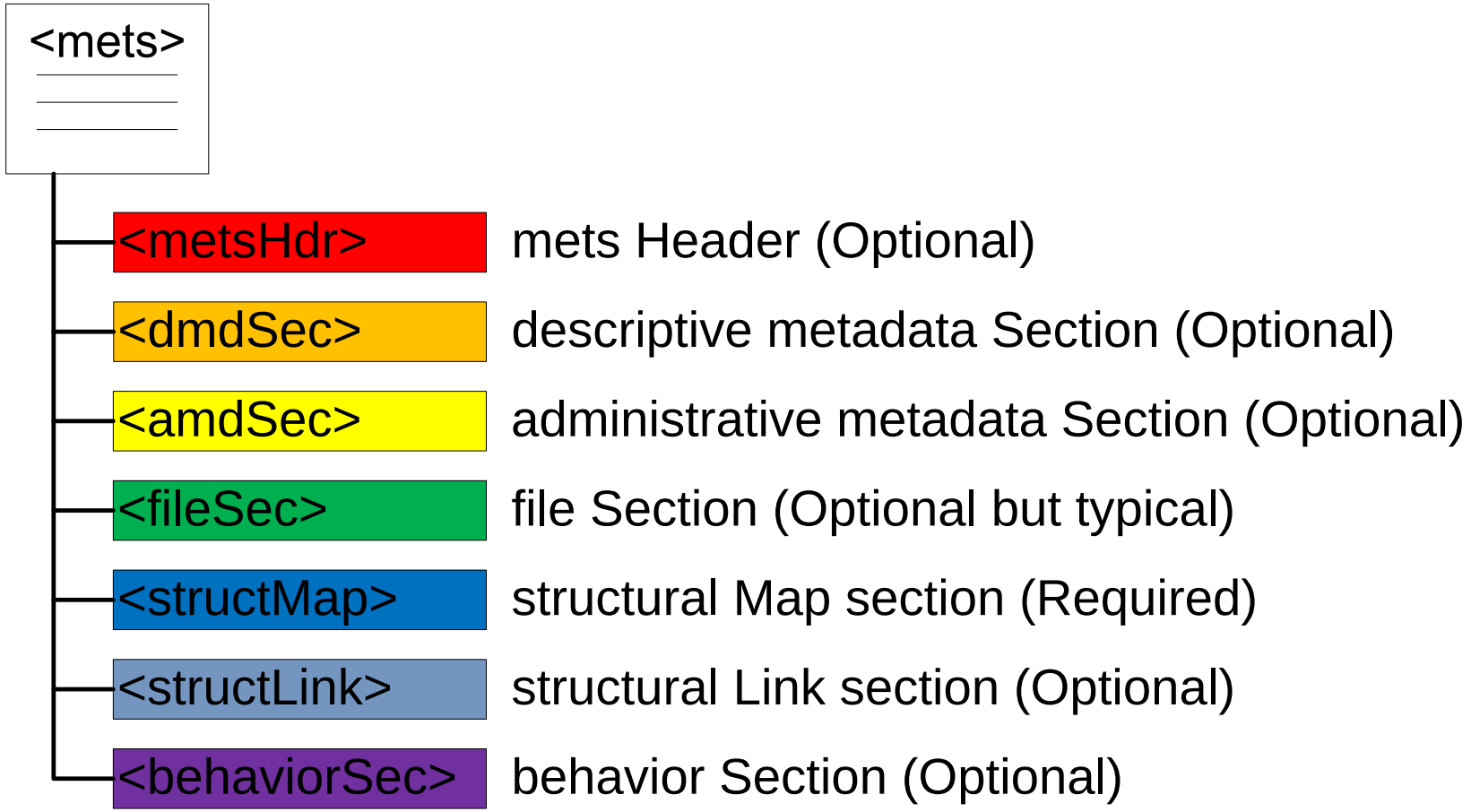
The Metadata Encoding and Transmission Standard (METS) used in CSIP

https://www.loc.gov/standards/mets/



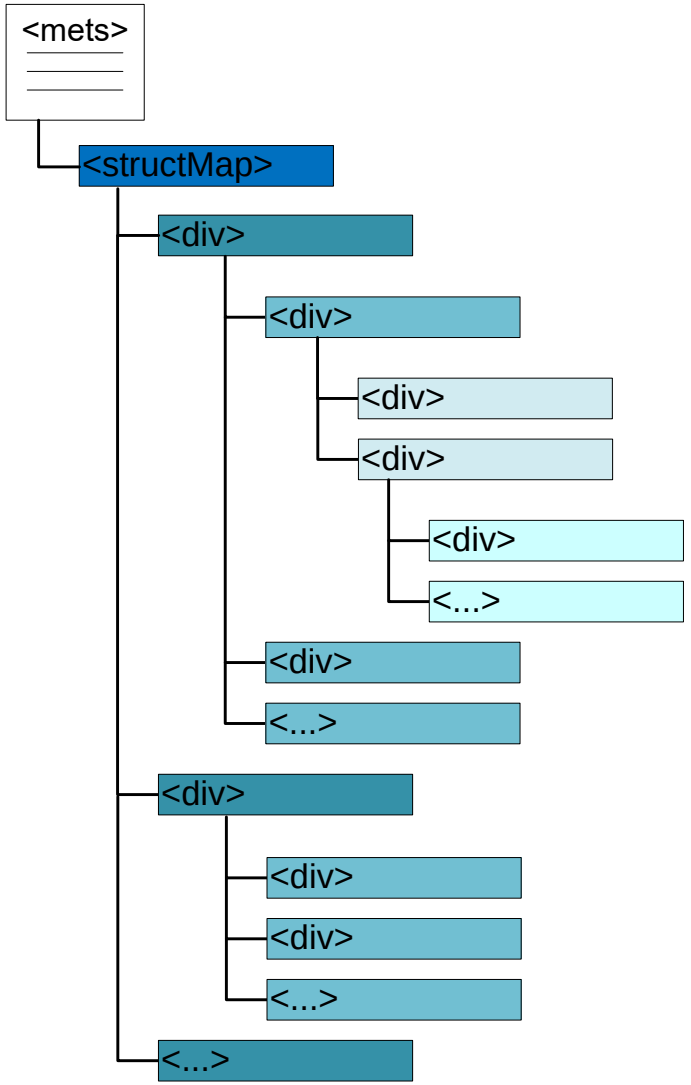
The standard METS

The sections of METS, a short overview



The standard METS and its core element

The structural map in METS

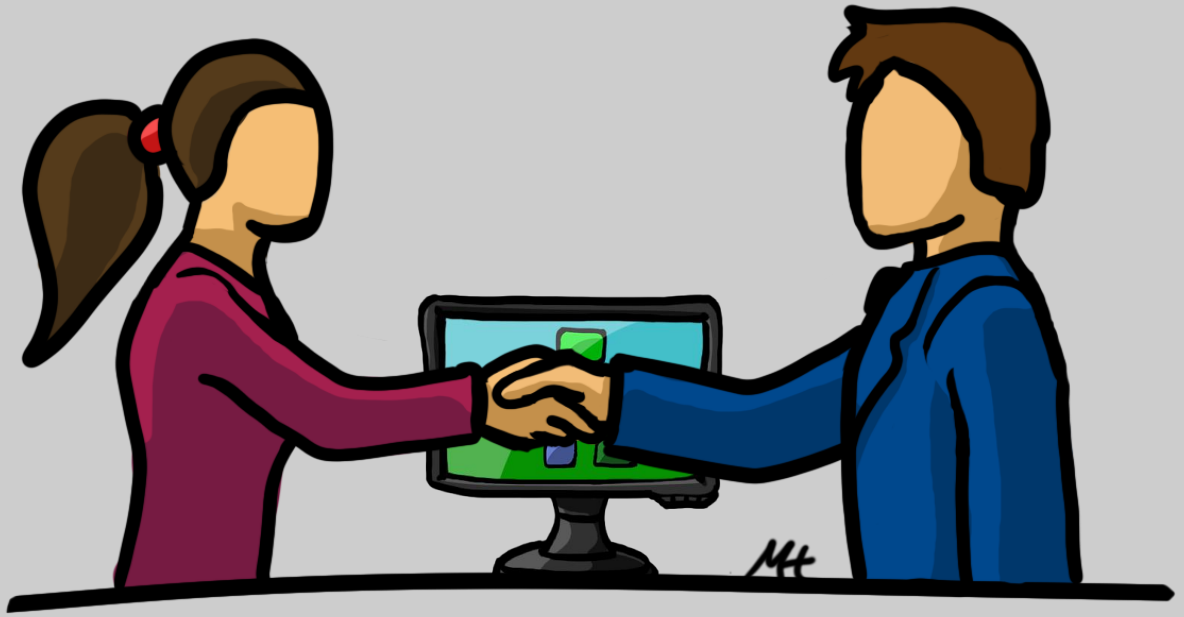


The standard METS and its core element

The structural map used in CSIP

```
<mets:structMap ID="struct-map-example-1" TYPE="PHYSICAL" LABEL="CSIP">
  <mets:div ID="struct-map-example-div" LABEL="csip-mets-example">
    <mets:div ID="struct-map-metadata-div" LABEL="Metadata" ADMID="digiprov-premis-file-1 digiprov-premis-file-2" DMDID="dmd-ead-file">
    </mets:div>
    <mets:div ID="struct-map-doc-div" LABEL="Documentation">
      <mets:fptr FILEID="file-ptr-doc">
      </mets:fptr>
    </mets:div>
    <mets:div ID="struct-map-schema-div" LABEL="Schemas">
      <mets:fptr FILEID="file-grp-schema">
      </mets:fptr>
    </mets:div>
    <mets:div ID="struct-map-reps-sub-div" LABEL="Representations">
      <mets:fptr FILEID="file-grp-rep-subdata">
      </mets:fptr>
    </mets:div>
  </mets:div>
</mets:structMap>
```

Connecting CSIP with METS



Common Specification for Information Package (CSIP)

The common elements and attributes used in the transfer described in a requirements table and a METS profile

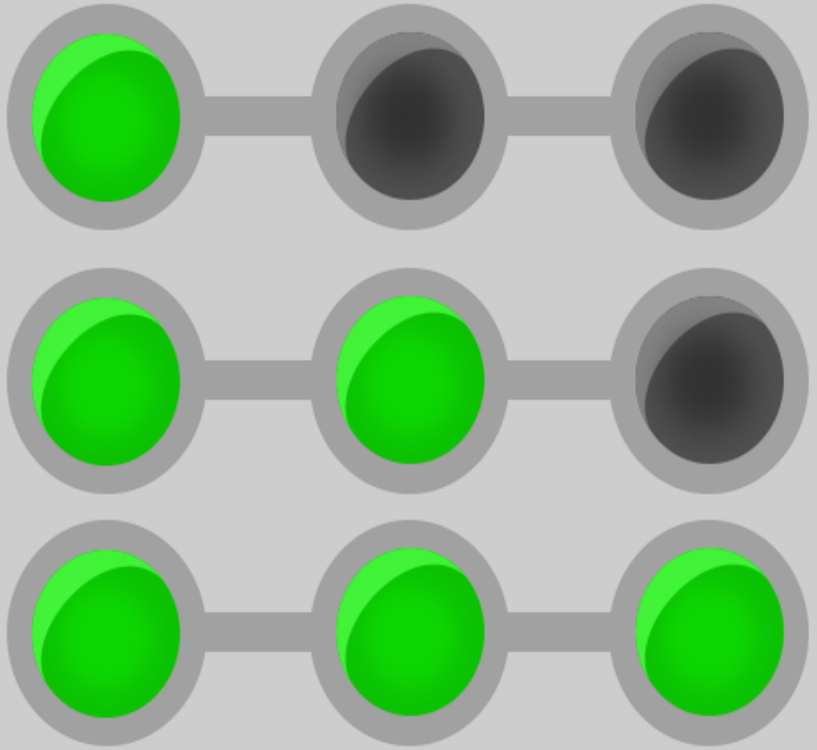
ID	Name, Location & Description	Card & Level
CSIP1	**Package Identifier** `mets/@OBJID` The `mets/@OBJID` attribute is mandatory, its value is a string identifier for the METS document. For the package METS document, this should be the name/ID of the package, i.e. the name of the package root folder. For a representation level METS document this value records the name/ID of the representation, i.e. the name of the top-level representation folder.	1..1 MUST
CSIP2	**Content Category** `mets/@TYPE` The `mets/@TYPE` attribute MUST be used to declare the category of the content held in the package, e.g. book, journal, stereograph, video, etc.. Legal values are defined in a fixed vocabulary. When the content category used falls outside of the defined vocabulary the `mets/@TYPE` value must be set to "OTHER" and the specific value declared in `mets/@csip:OTHERTYPE` . The vocabulary will develop under the curation of the DILCIS Board as additional content information type specifications are produced. **See also:** Content Category	1..1 MUST
CSIP3	**Other Content Category** `mets[@TYPE='OTHER']/@csip:OTHERTYPE` When the `mets/@TYPE` attribute has the value "OTHER" the `mets/@csip:OTHERTYPE` attribute MUST be used to declare the content category of the package/representation. **See also:** Content Category	0..1 SHOULD
CSIP4	**Content Information Type Specification** `mets/@csip:CONTENTINFORMATIONTYPE` Used to declare the Content Information Type Specification used when creating the package. Legal values are defined in a fixed vocabulary. The attribute is mandatory for representation level METS documents. The vocabulary will evolve under the care of the DILCIS Board as additional Content Information Type Specifications are developed. **See also:** Content information type specification	0..1 SHOULD
CSIP5	**Other Content Information Type Specification** `mets[@csip:CONTENTINFORMATIONTYPE='OTHER']/@csip:OTHERCONTENTINFORMATIONTYPE` When the `mets/@csip:CONTENTINFORMATIONTYPE` has the value "OTHER" the `mets/@csip:OTHERCONTENTINFORMATIONTYPE` must state the content information type.	0..1 MAY
CSIP6	**METS Profile** `mets/@PROFILE` The URL of the METS profile that the information package conforms with.	1..1 MUST

Example: METS root element showing use of `csip:@OTHERTYPE` attribute when an appropriate package content category value is not available in the vocabulary. The `@TYPE` attribute value is set to OTHER.

```
<mets:mets OBJID="uuid-4422c185-5407-4918-83b1-7abfa77de182" LABEL="Sample CSIP Information Package" TYPE="OTHER" OTHERT
</mets:mets>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Draft of E-ARK CSIP METS Profile 2.0 -->
<METS_Profile xmlns="http://www.loc.gov/METS_Profile/v2"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:mets="http://www.loc.gov/METS/"
    xmlns:csip="https://DILCIS.eu/XML/METS/CSIPExtensionMETS"
    xmlns:xlink="http://www.w3.org/1999/xlink"
    xsi:schemaLocation="http://www.loc.gov/METS_Profile/v2 http://www.loc.gov/standards/mets/profile_docs/mets.profile.v2-0.xsd h
    <URI LOCTYPE="URL" ASSIGNEDBY="local">https://earkcsip.dilcis.eu/profile/E-ARK-CSIP.xml</URI>
    <title>E-ARK CSIP METS Profile</title>
    <abstract>This base profile describes the Common Specification for Information Packages (CSIP) and the implementation of METS
        This will enable repository interoperability and assist in the management of the preservation of digital content.
        This profile is a base profile which is extended with E-ARK implementation of SIP, AIP and DIP.
    The profile can be used as is, but it is recommended that the supplied extending implementation are used. Alternatively, an o
    <date>2020-01-08T12:00:00</date>
    <contact>
        <institution>DILCIS Board</institution>
        <address>http://dilcis.eu/</address>
        <email>info@dilcis.eu</email>
    </contact>
    <related_profile>This profile has no related profiles</related_profile>
```

The connection between the different Information Package specifications



E-ARK SIP, E-ARK AIP and E-ARK DIP

Special use cases of the CSIP

- E-ARK SIP + E-ARK DIP extends CSIP
  - Defines the Submission Information Package and the Dissemination Information Package
  - Some extra attributes
  - Defines some values to use like informing about where we are in the OAIS reference model; SIP and DIP
- E-ARK AIP description of how to handle a package in an electronic archive

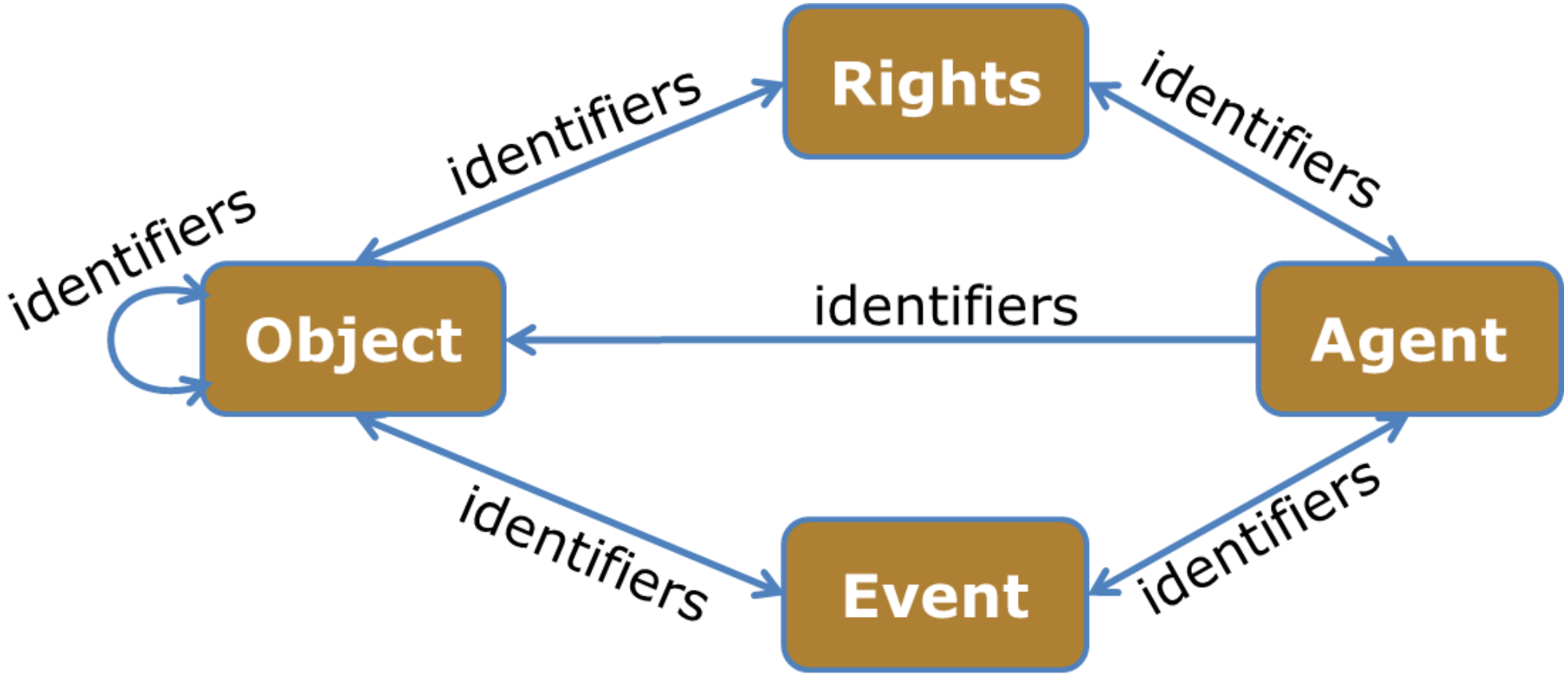
Connected to the content of the package

Archival Information

Information regarding the creator and finding aid

Preservation metadata

Preservation metadata for the content

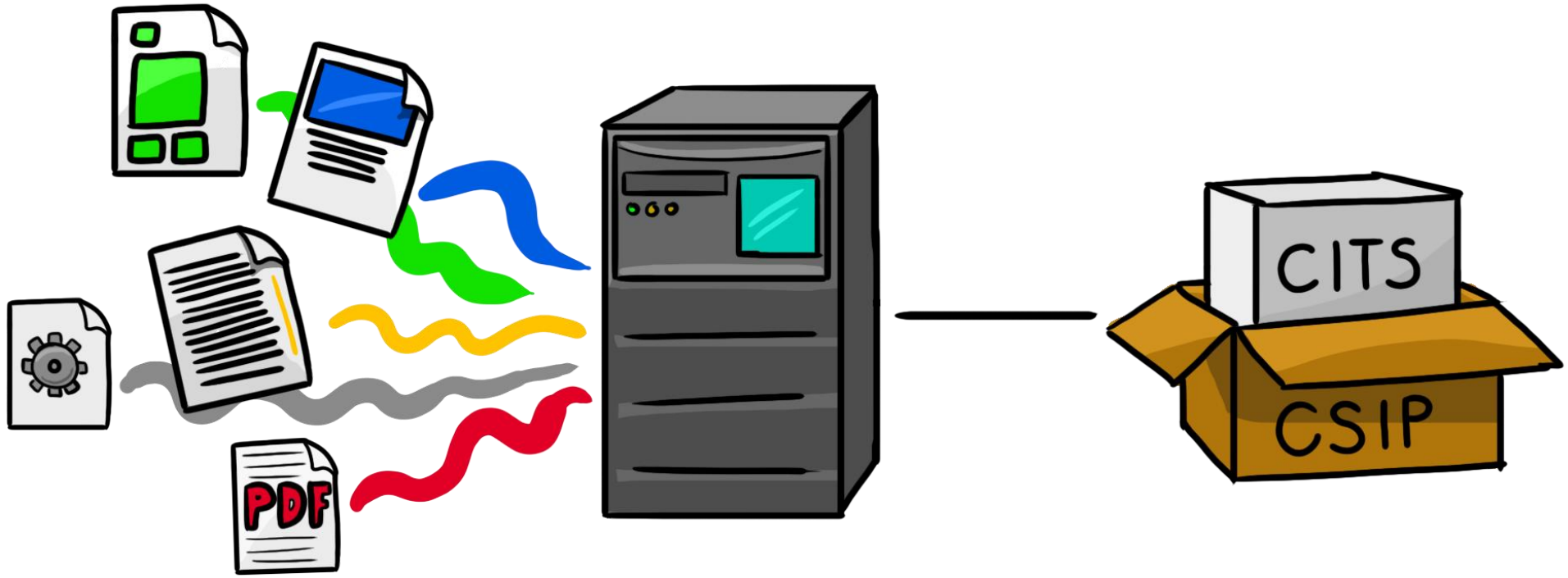


The Information Type Specifications **CITS**

The Content Information Type Specifications; ways of reaching the goal of a CITS

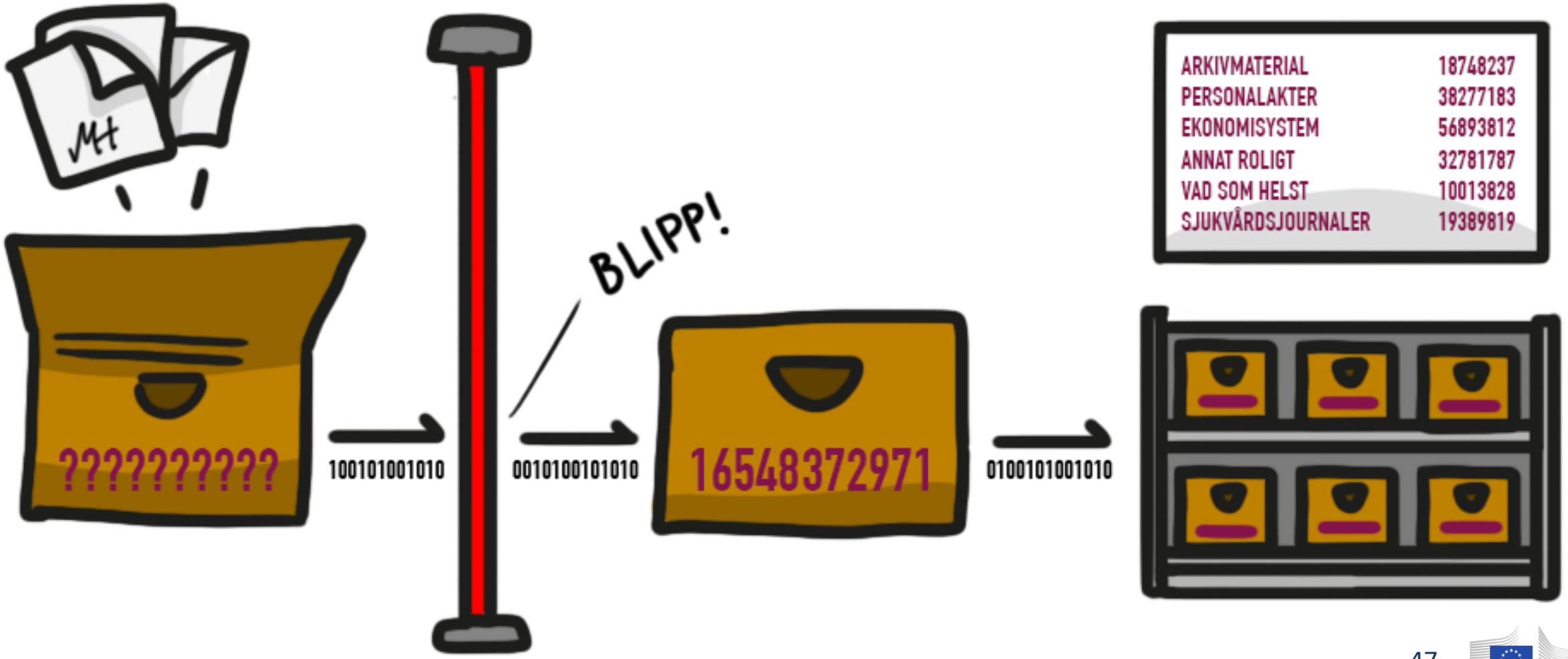
Endorsement and placement

A specification is already used and gets placed in a package



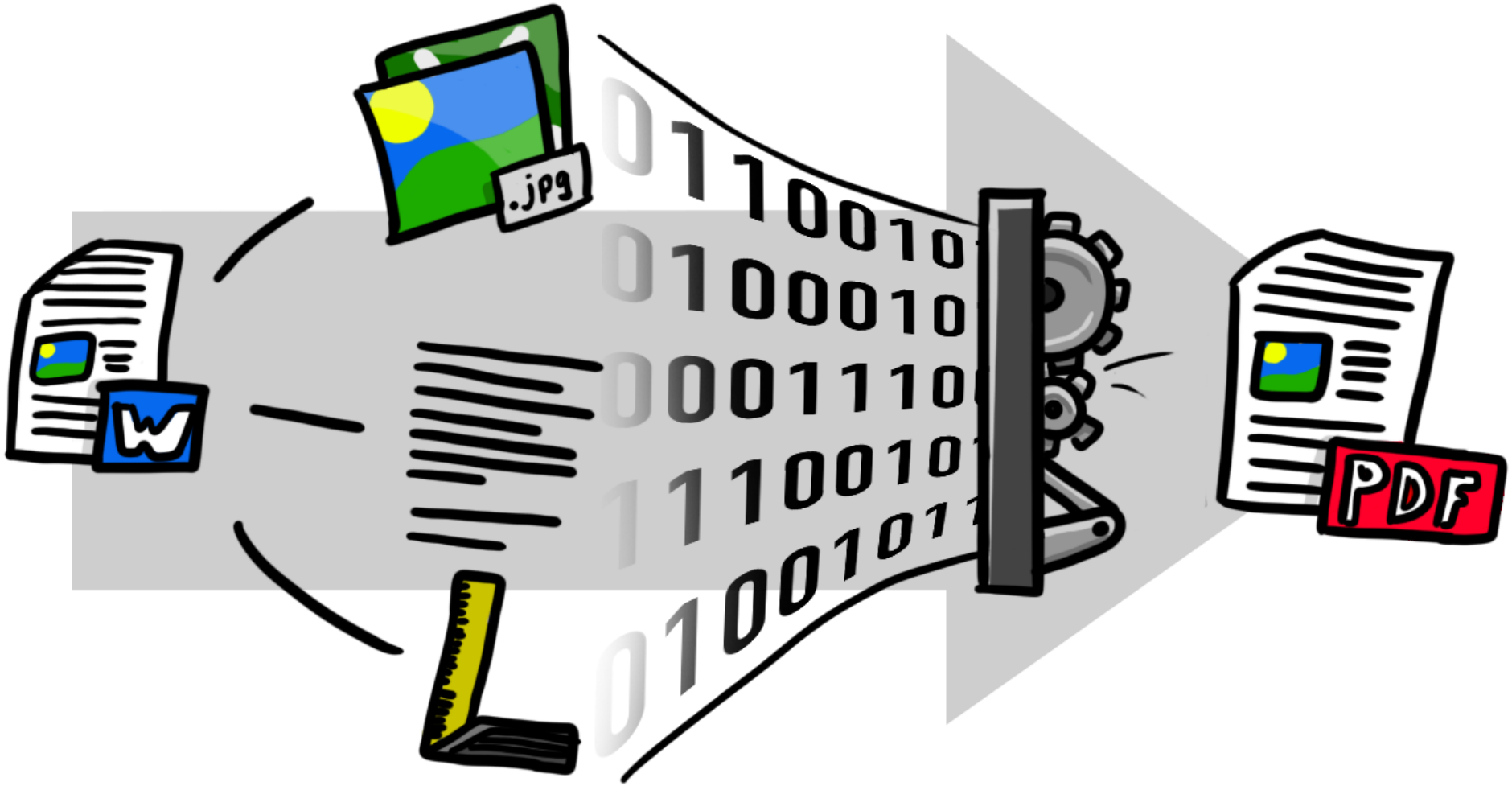
Giving all the requirements for the content

All information needed for the content is provided along with how to place it in a package



Giving the requirements for the content and all the files

All requirements needed for the content are given including all files that should be present and how to place them in a package



Introducing the specifications; the different ways to create a CITS

The E-ARK Content Information Type Specification for Relational Databases using SIARD (CITS SIARD)

Using a standard and focus upon placement in the package (equals an endorsement)

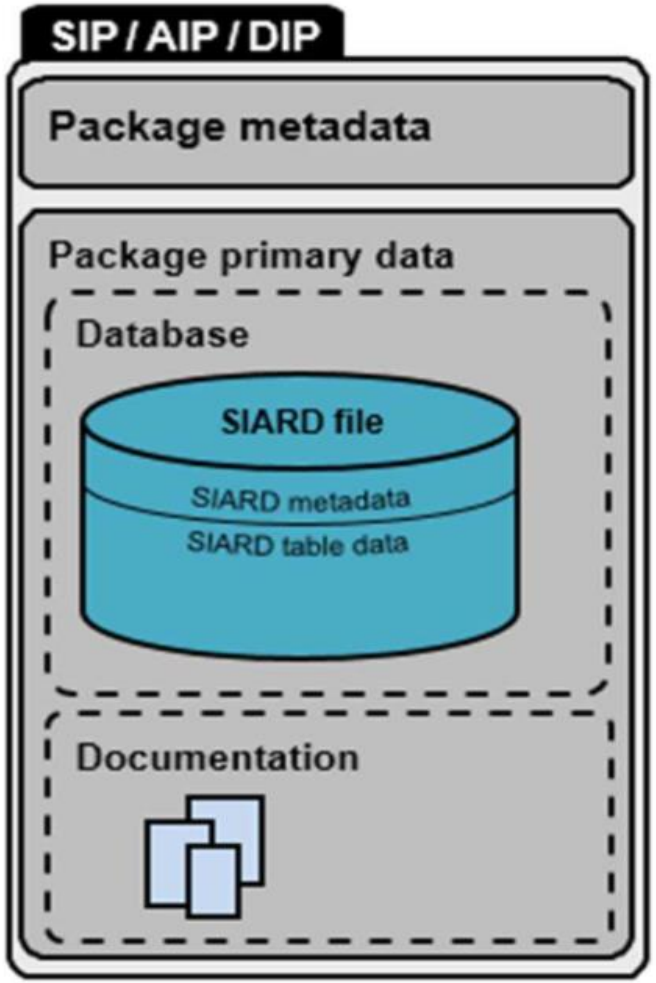
CITS SIARD

The standard

https://dilcis.eu/content-types/siard

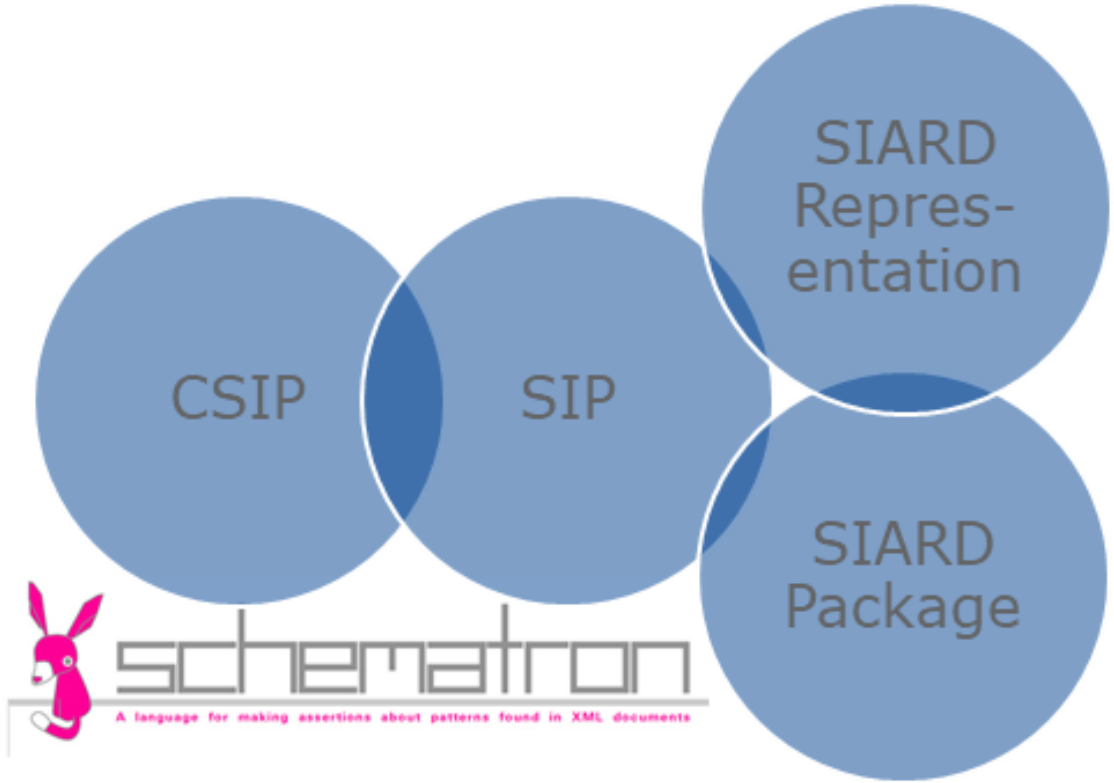
CITS SIARD

A SIARD package



CITS SIARD

In the package

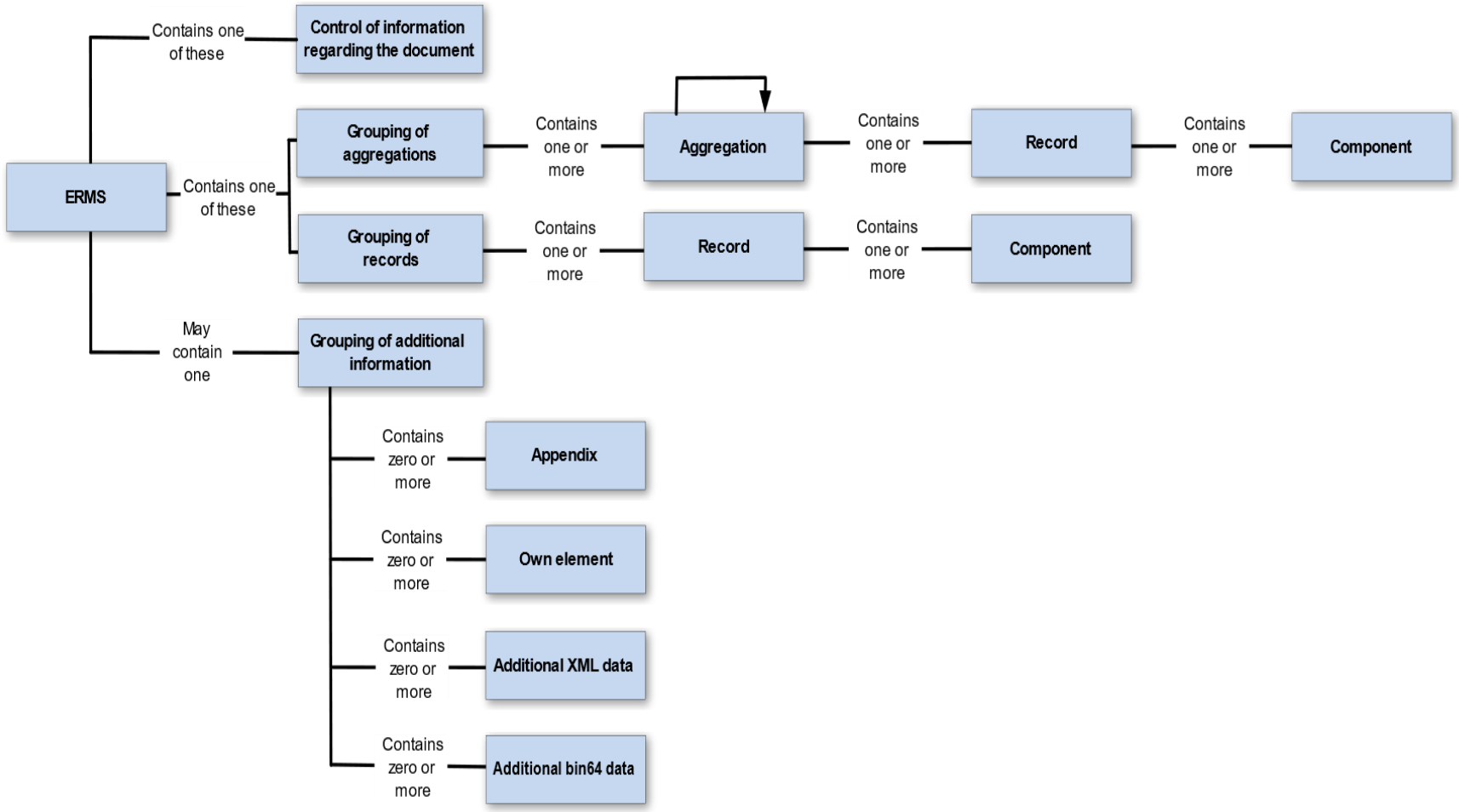


The E-ARK Content Information Type Specification for Electronic Records Management Systems (CITS ERMS)

Giving the requirements for the content

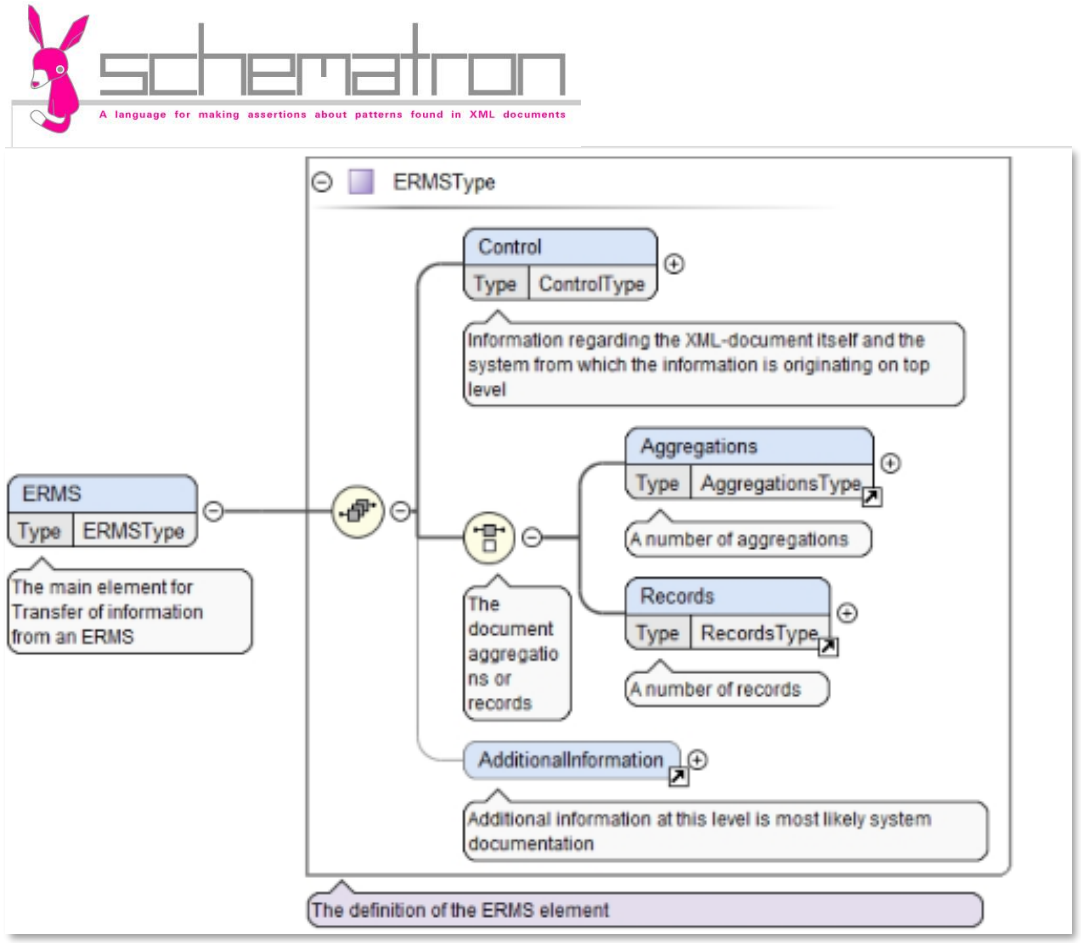
CITS ERMS

Content focused



CITS ERMS

The specification has a schema with Schematron rules



CITS ERMS

Placement in the package

3.3.2 Placement of data in a CSIP Information Package

The ERMS document is placed as a representation file following the instructions in CSIP.

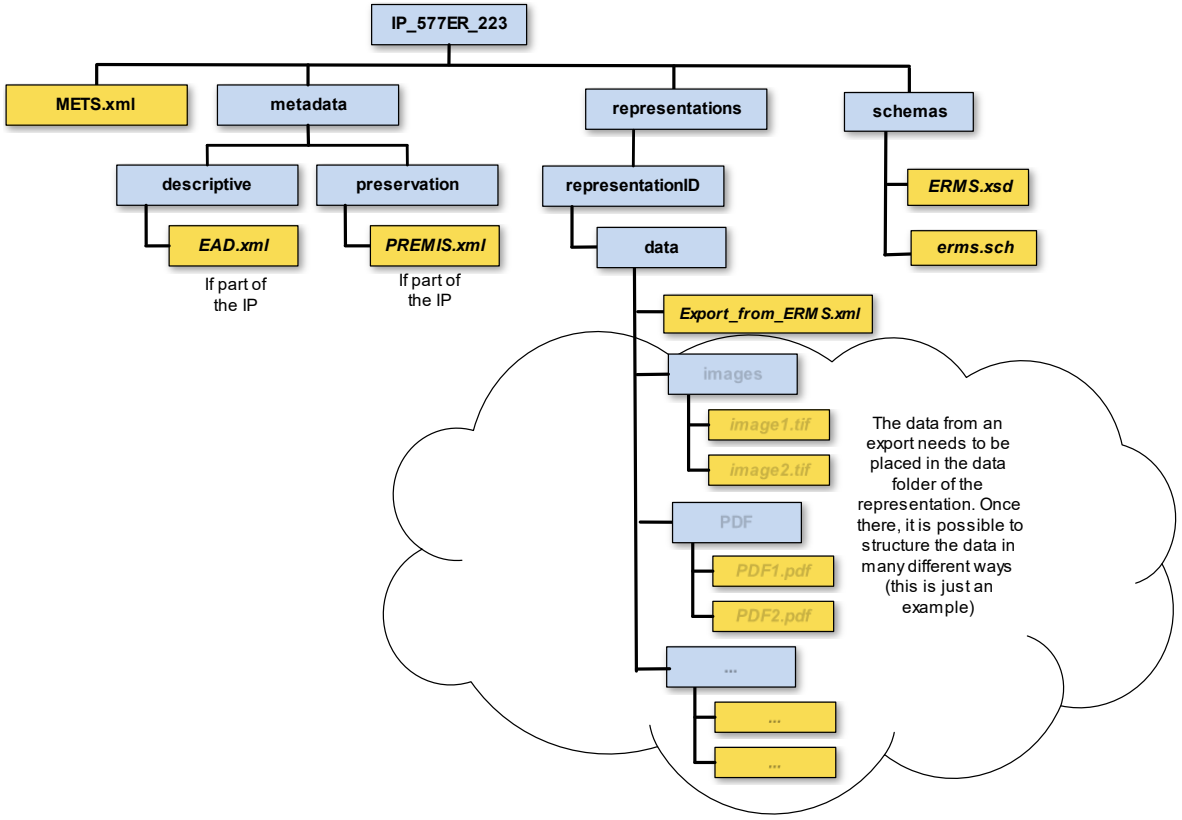


Table 1: Specific fields to use in CSIP

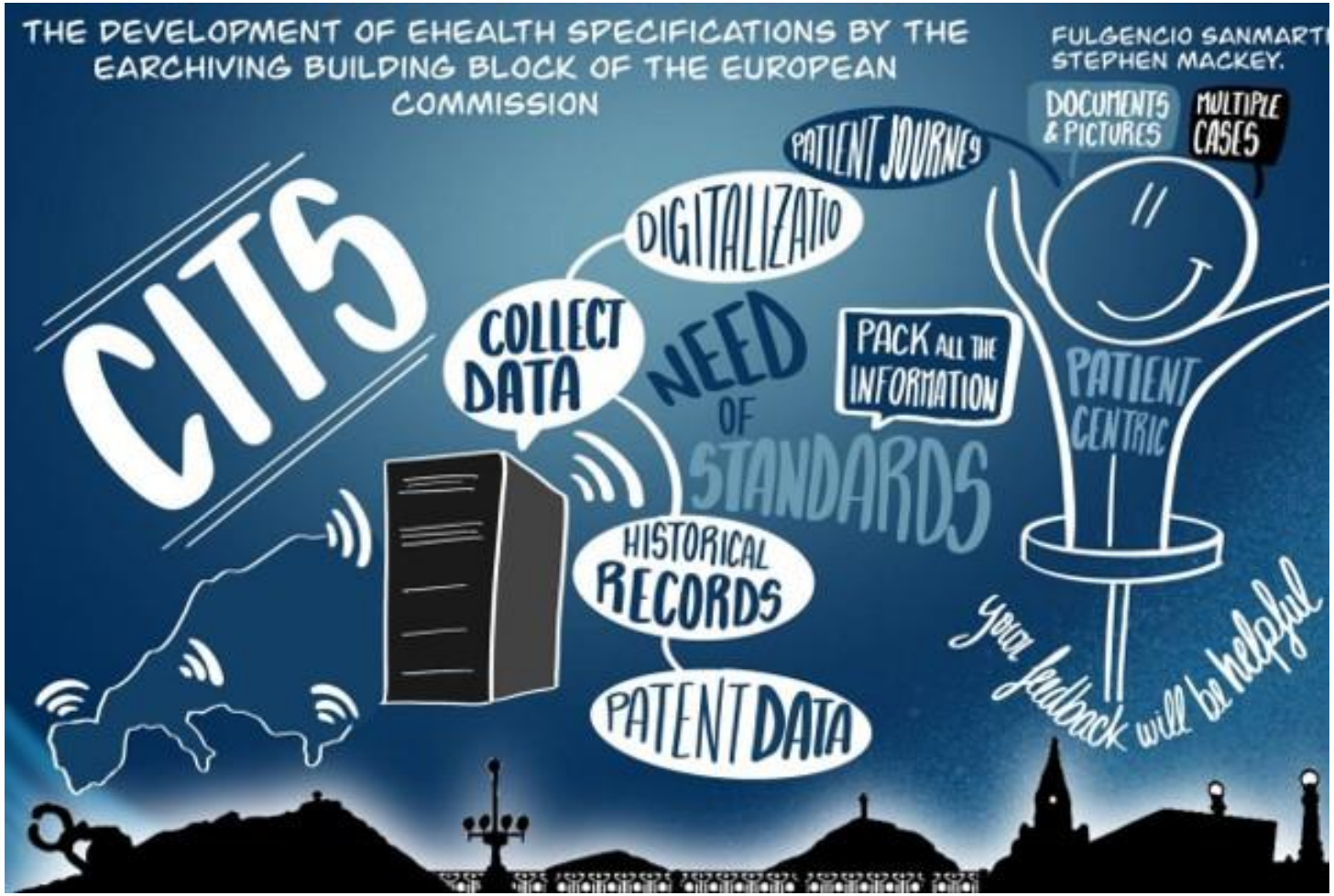
Element name	METS path	Value
General content type	mets/@TYPE	Dataset
Specific content type	mets/@csip:CONTENTINFORMATIONTYPE	ERMS
Specific content type	fileGrp/@csip:CONTENTINFORMATIONTYPE When the FileGrp describes a Representation	ERMS

The E-ARK Content Information Type Specification for eHealth1
(CITS eHealth1)

Giving overall requirements for the content and how to place it in a package

eHealth1

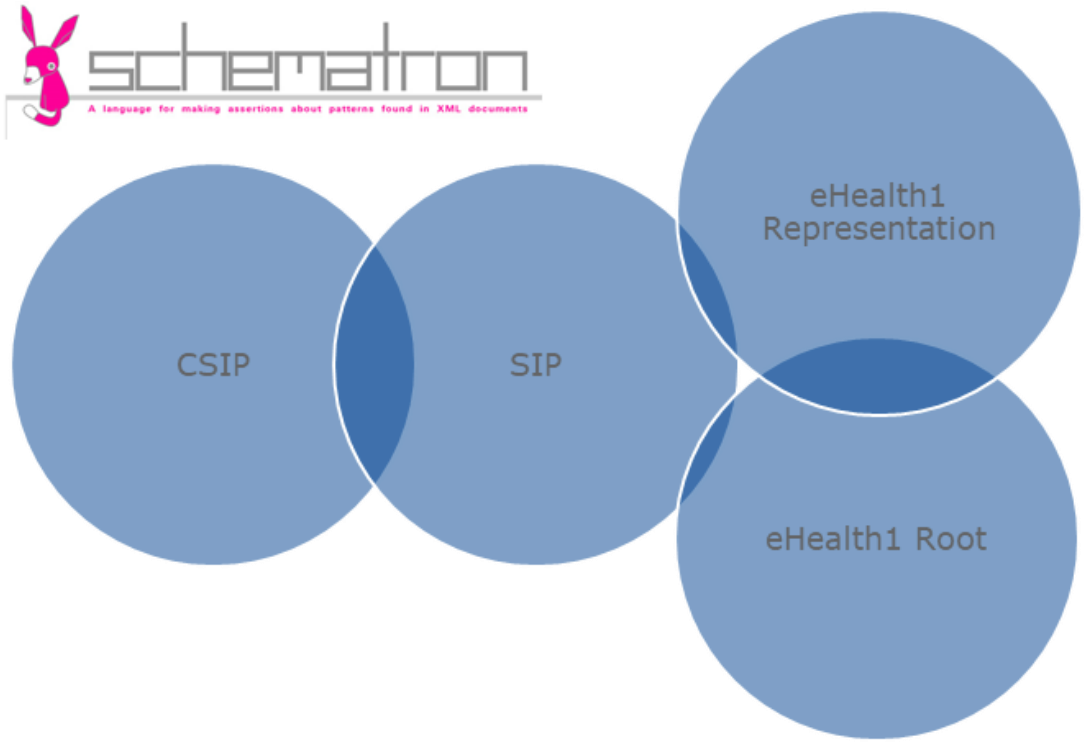
The use cases and data



https://twitter.com/MarisaMerinoH

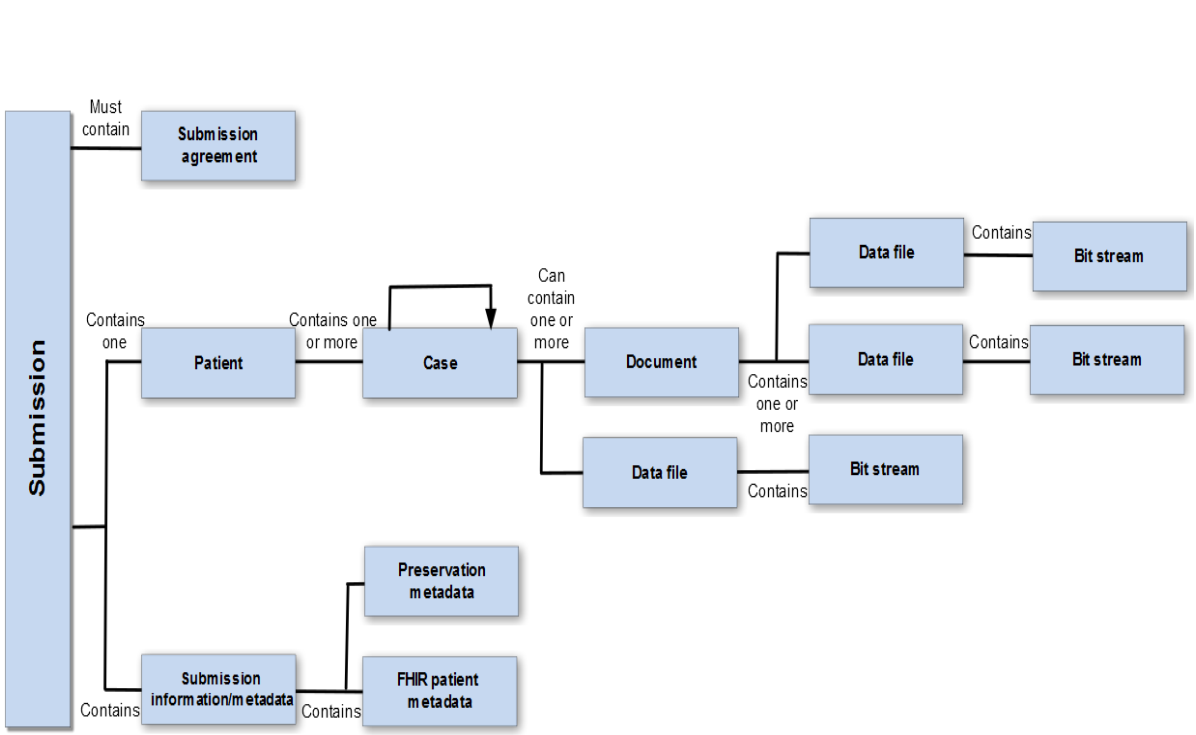
eHealth1

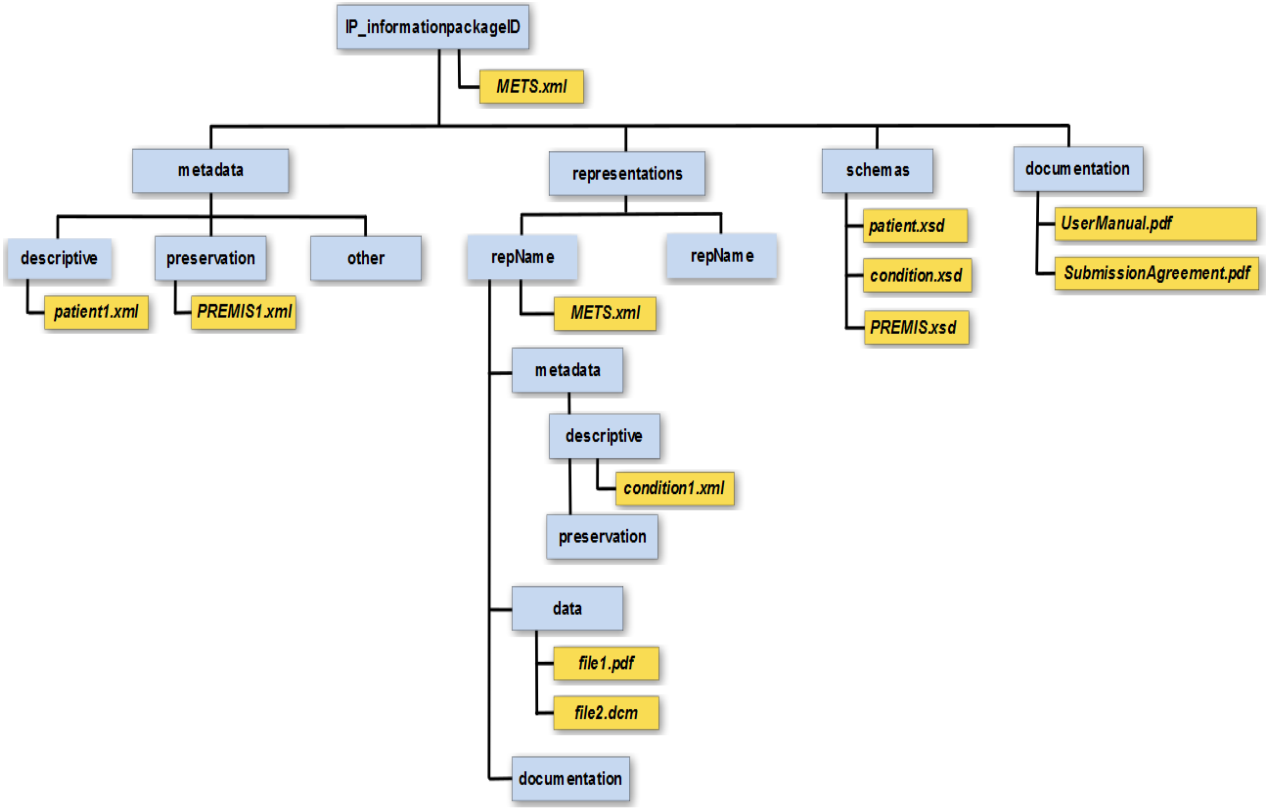
The specification itself



eHealth1

A closer look at the structure





Tools;
ways of reaching the goal of a CITS

Creation of files following a specification

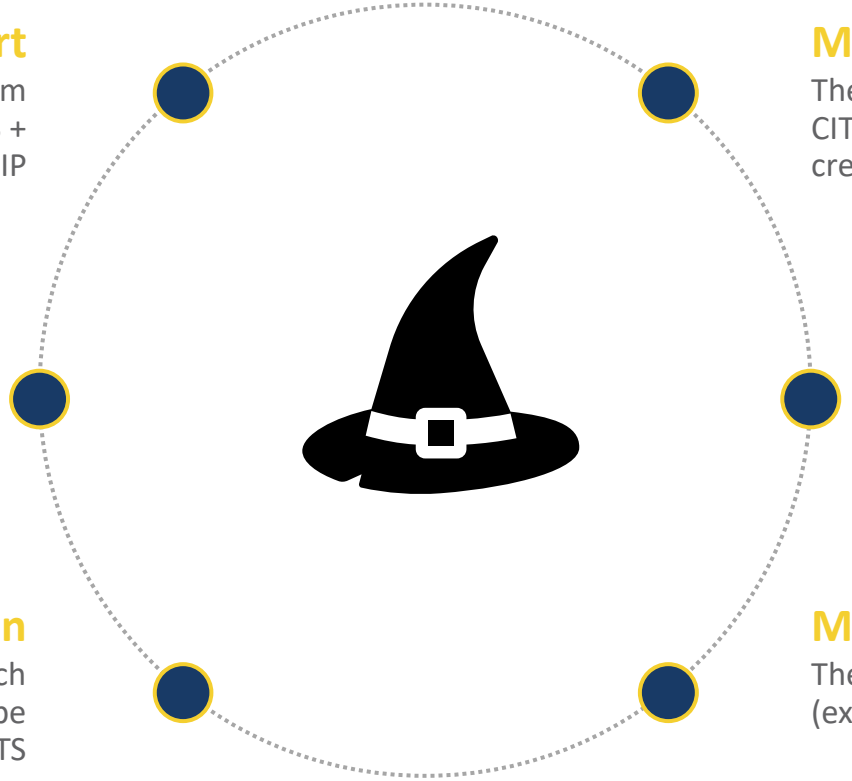
There are many ways to get the resulting content and put it into a package

Built in export

There might be an export built into a system which creates a package following a CITS + CSIP

Mapping

There might be a need to do a mapping to a CITS format when performing an export to create the export file(s)



Package creation tools

There might be tools that create packages (examples: RODA-IN, ESSArch)

Export tool

There might be export tools available or developed (examples: DBPTK, SIARD Suite, eHealth1 SIP creation tool)

Export and transformation

There might be a built in export which creates an XML document which can be transformed to follow a CITS

Manual creation

There might be manually created files (examples: an XML Editor is used)

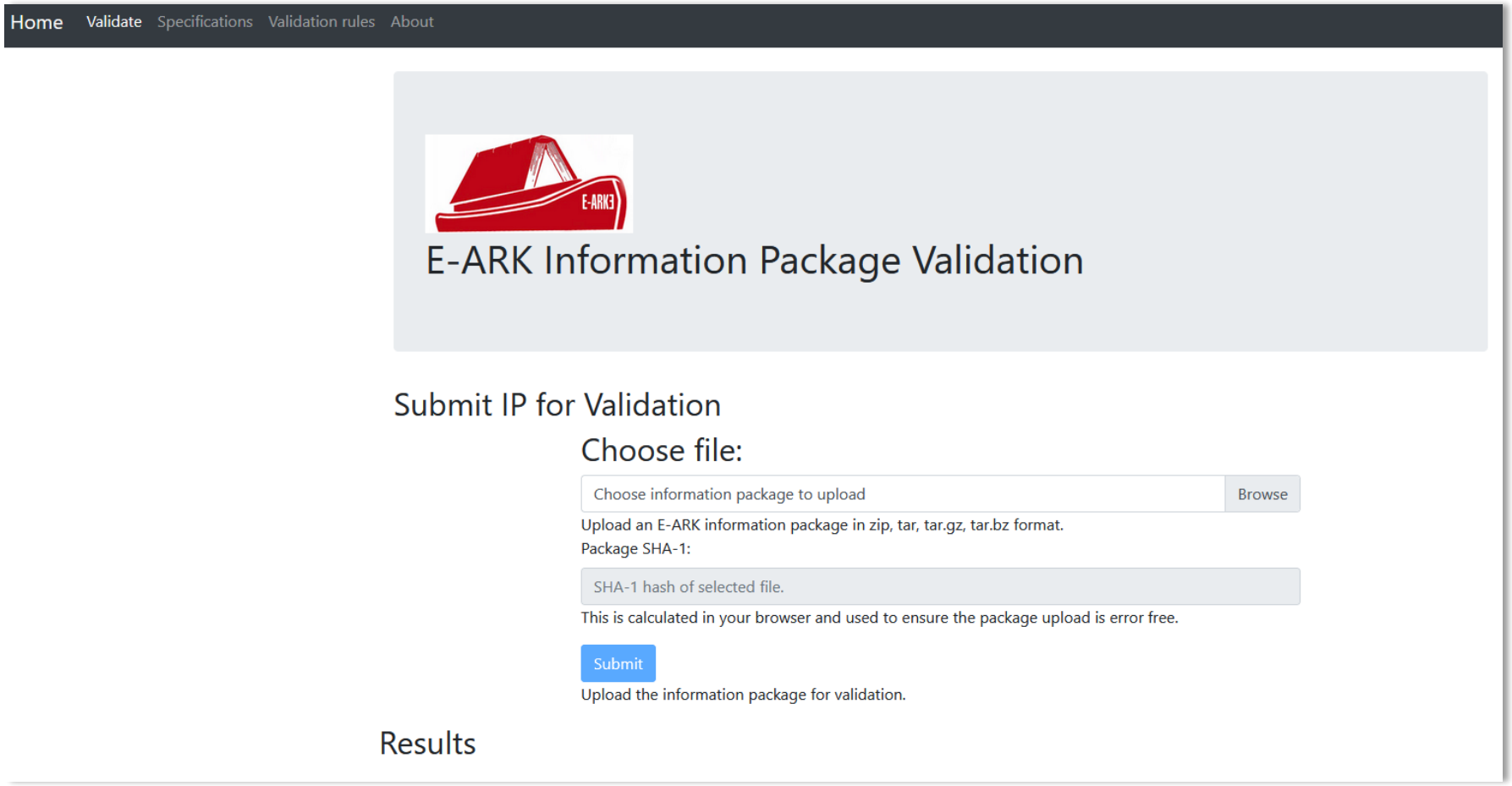
CITS ERMS

ERMS and tools currently rely upon mapping in the export moment. In the future?

Observe that the specifications do not give a general export to any ERMS software because how an export is facilitated depends on the ERMS

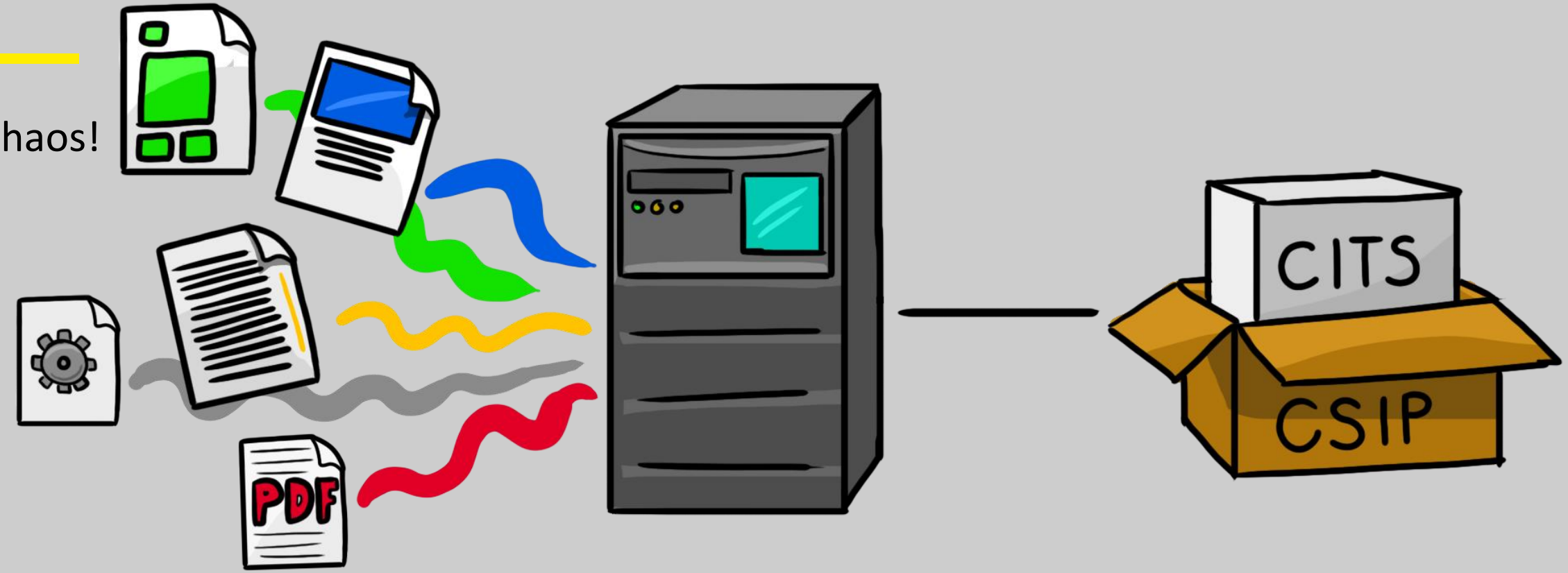
CSIP Validation

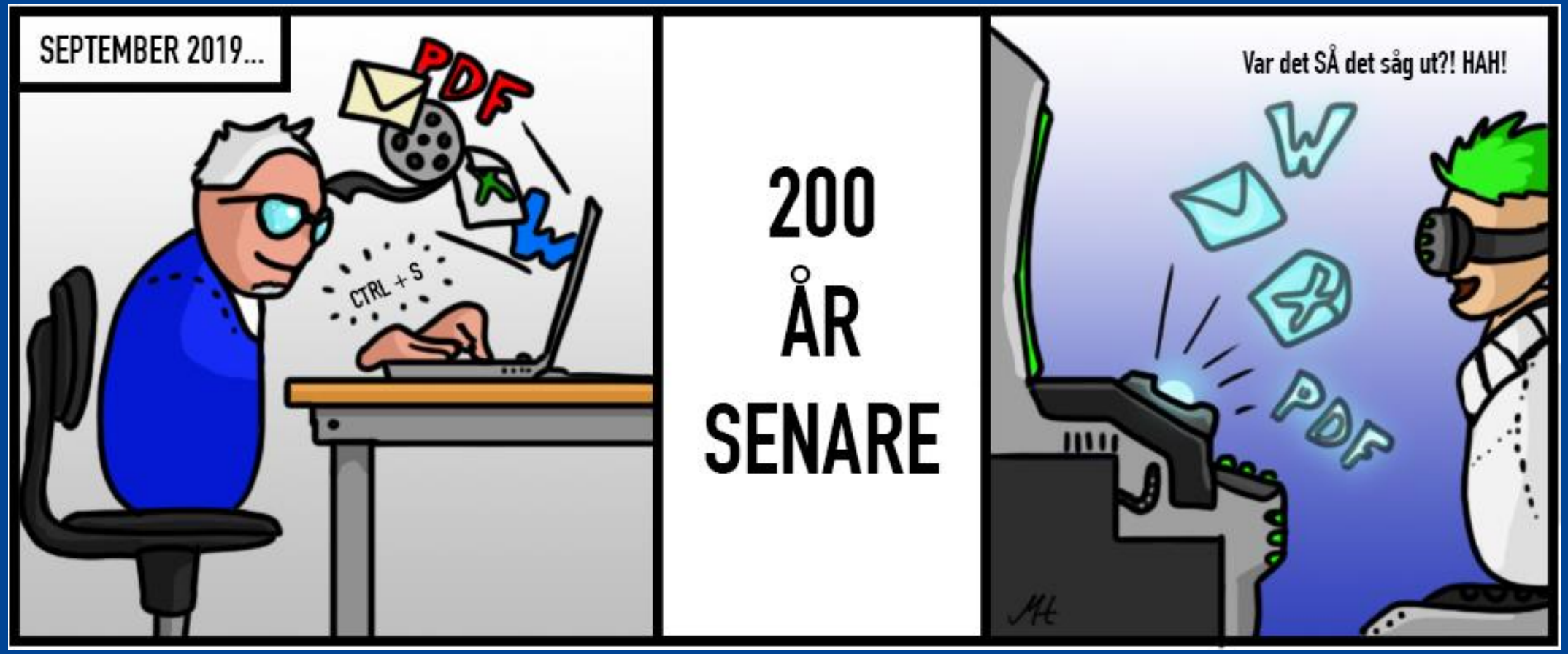
A tool for validating CSIP have been developed



https://eark.openpreservation.org/

Chaos!




SEPTEMBER 2019...
PDF
W
CTRL + S
200
ÅR
SENARE
Var det SÅ det såg ut?! HAH!
W
X
PDF

Links

https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving
https://dilcis.eu/
https://github.com/DILCISBoard
https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Sample+Software+Portfolio
https://github.com/E-ARK-Software
https://joinup.ec.europa.eu/collection/interoperability-test-bed-repository/solution/interoperability-test-bed/news/itb-and-cef-earchiving
https://riksarkivet.se/intro-fgs
https://www.loc.gov/standards/mets/
https://github.com/SAA-SDT
https://www.loc.gov/standards/premis/

Questions?

Karin Bredenberg
Kommunalförbundet Sydarkivera
karin.bredenberg@sydarkivera.se

E-ARK Programme
LinkedIn: www.linkedin.com/groups/8343650/
Twitter: #EARKProject

Ready to get started?

Find out more at:
ec.europa.eu/cefdigital

Contact us:
cef-building-blocks@ec.europa.eu

Thank you!