Welcome to this live webinar on E-ARK validation

Start 10:00 (CET)

29 October 2020

Ground Rules for the Live Webinar

Click on “Connect audio” to hear the presenters but please mute your microphone throughout the webinar.

Submit your questions in writing by using the Webex chat function. We will answer some questions live during the webinar and provide written answers to all (within the coming days).

Please note that this webinar is recorded.

#ConnectingEurope

Agenda

10:00 – 10:05

Welcome

Thomas Fillis – CEF Stakeholder Management Office – DIGIT

10:05 – 10:15

CEF eArchiving welcome

Dr. Jaime Kaminski – CEF eArchiving Building Block training activity lead

10:15 – 11:15

E-ARK validation

Carl Wilson – Open Preservation Foundation

Costas Simatos – DIGIT

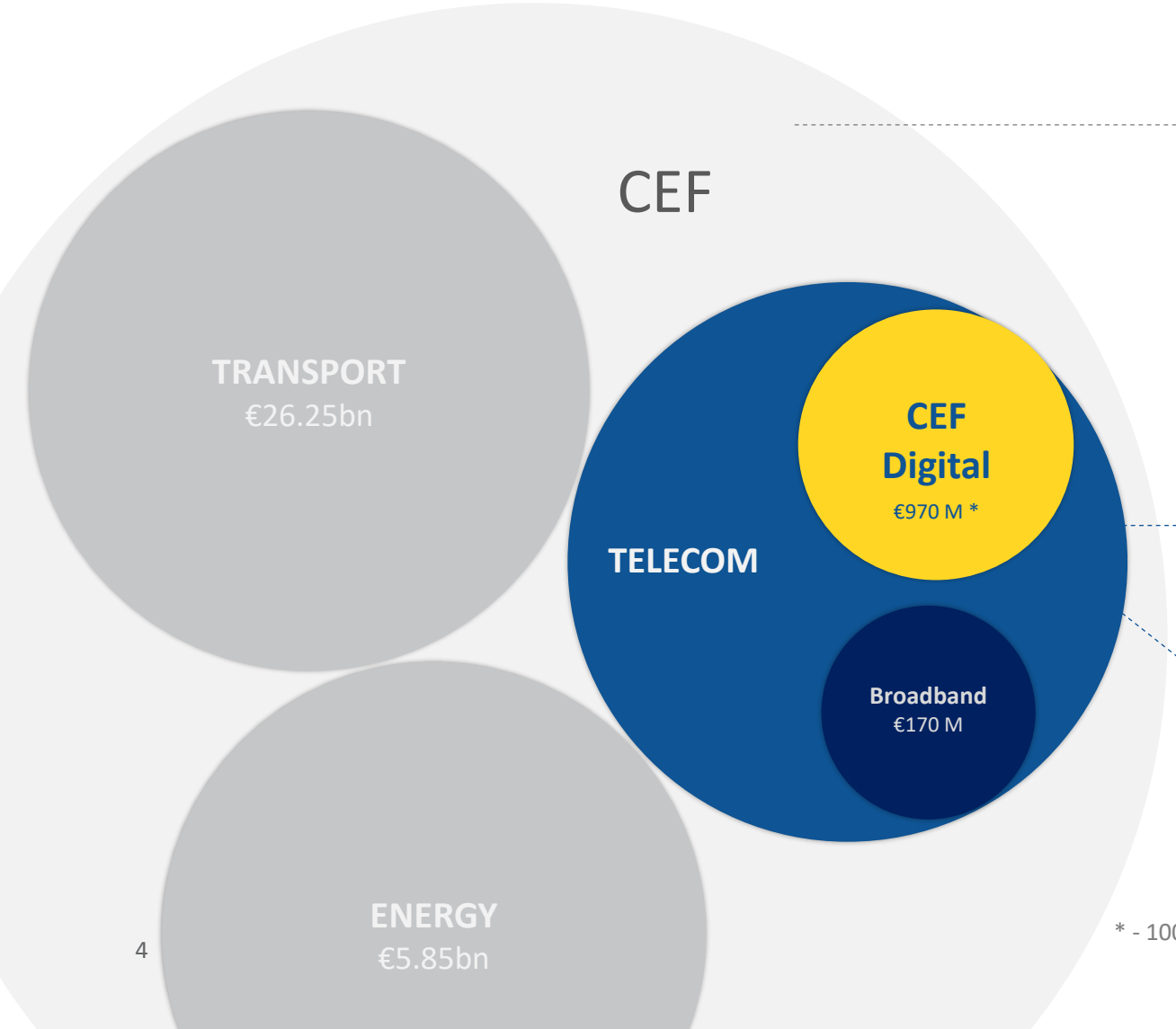
11:15 – 11:30

Q&A

Welcome to the Connecting Europe Facility (CEF) Building Blocks

Thomas Fillis
CEF Stakeholder Management Office, DIGIT

The CEF Building Blocks are funded by the Connecting Europe Facility



CEF Regulation
Defines how the Commission can finance support for the establishment of trans-European networks to reinforce an interconnected Europe.

CEF Telecom Guidelines
The CEF Telecom guidelines cover the specific objectives and priorities as well as eligibility criteria for funding of broadband networks and Digital Service Infrastructures (DSIs).

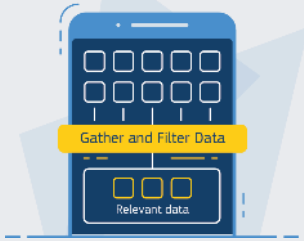
CEF Work Programmes
Translates the CEF Telecom Guidelines in general objectives and actions planned on a yearly basis.

* - 100 M Juncker Package

European Commission

Big Data Test Infrastructure

Explore and experiment with big data for improved performance and decision making

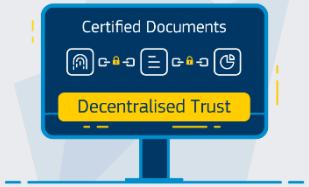


Context Broker

Analyze, manage and share data, in real time, at the right time, throughout Europe

eArchiving

Facilitates the preservation, migration, reuse and trust of your data

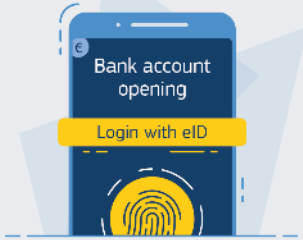


European Blockchain Services Infrastructure

Harness the power of a European-wide network of blockchain services, increasing trust through data security, privacy and transparency

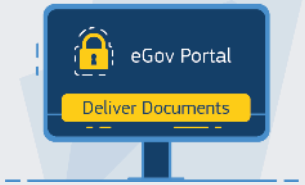
eInvoicing

Promote the implementation of the European standard for electronic invoicing across borders



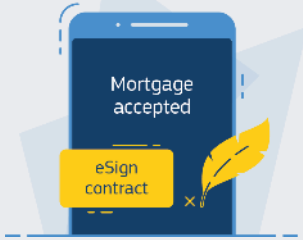
eID

Allow citizens to prove who they are across borders, making it easier to access online services in another EU Member State



eDelivery

Exchange online data and documents reliably and securely



eSignature

Create and verify electronic signatures between businesses and EU citizens

eTranslation

Offers machine translation to translate your documents and web content into any official EU language, Norwegian or Icelandic

How does CEF support projects to use the Building Blocks?

It provides free services to help you implement them in your system. There are a range of services across the building blocks but services typically include training, sample software, testing services.

Free services

Training sessions

Sample software

Testing services

How to use a Building Block?

Build, buy or reuse the Building Blocks on your own.

Co-develop the solution or partner with other parties.

Co-develop and partner
with other parties

Build
The solution from scratch based on a European standard

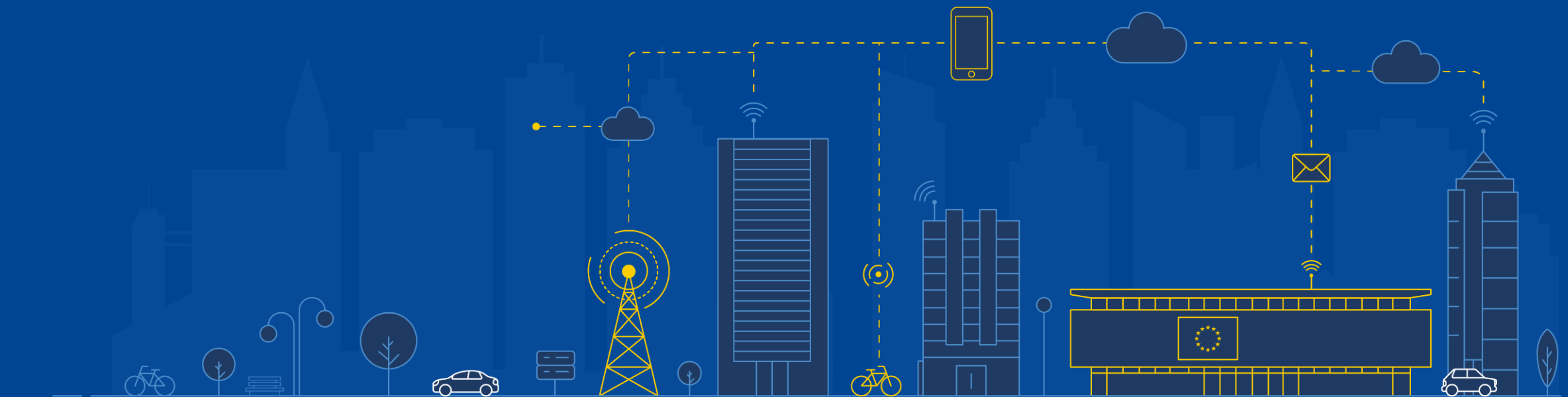
Buy
A compliant solution from the market

Reuse
Sample software available on CEF website

European Standards

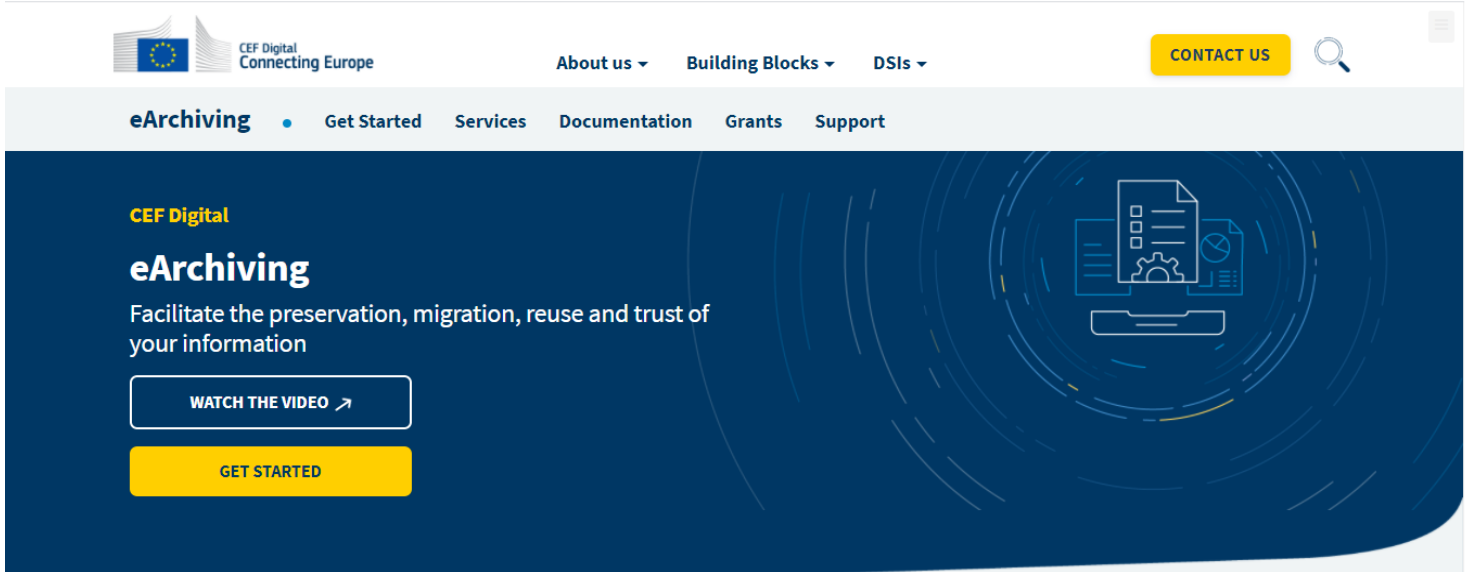
Welcome to the CEF eArchiving Building Block

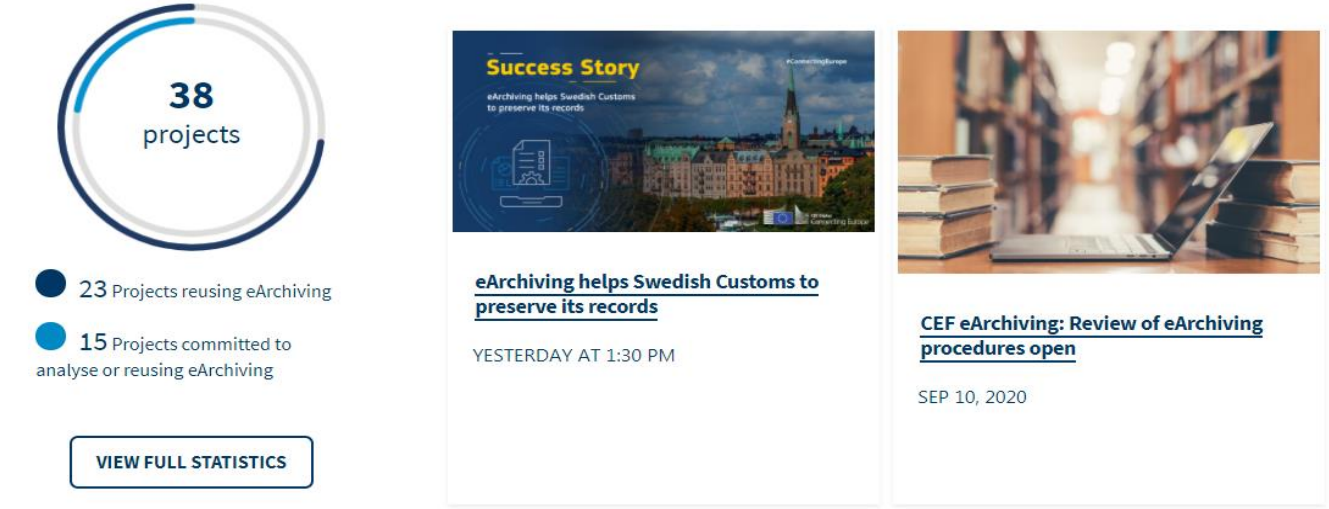
Dr. Jaime Kaminski
CEF eArchiving activity lead training



eArchiving Building Block

eArchiving Building Block website





https://ec.europa.eu/cefdigital/earchiving

European Commission

eArchiving Building Block

eArchiving services:

- Technical specifications
- Sample software
- Compliance/validation
- Service Desk
- Training
- Outreach/community engagement

Training:
eArchiving training is based on actual user requirements

EUSurvey

Create better online surveys and forms

Login

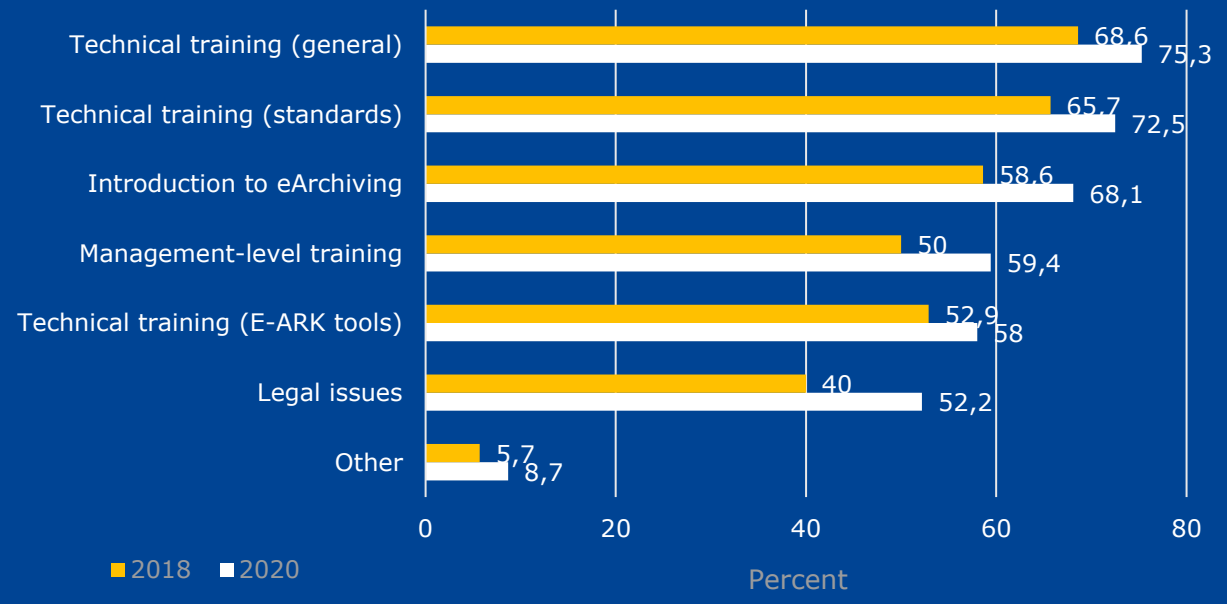
E-ARK3 is currently conducting a Training Needs Analysis for digital archiving as part of the Connecting Europe Facility (CEF) eArchiving Building Block. The need for eArchiving training has never been greater. In the last two decades, governments, businesses and individuals have become increasingly digital. So much so that ninety per cent of the data in the world today has been generated in the last two years.

Data on this scale brings opportunities but it also brings challenges. The problem for memory institutions is that archiving even a small fraction of this data will become a massive burden. Long-term repositories dealing with this deluge must implement appropriate specifications, tools and workflows capable of processing huge quantities of information in an increasing variety of formats. Organizations in the archiving ecosystem need to be aware of the different approaches, systems and formats that can be implemented.

The aim of E-ARK3 is to promote a deeper engagement with and access to, eArchiving services. One of the ways that we will achieve this is by offering training. The following Training Needs Analysis will help ensure that the training developed is appropriate for our target communities, and meets their actual needs.

Your contribution to this questionnaire will help shape this training. Please encourage your colleagues and staff to respond. It should take about 10 minutes to complete.

Next



eArchiving training plans

Core training will be delivered as:

- **Webinars**
- **Video training**
- **Moodle LMS training modules**

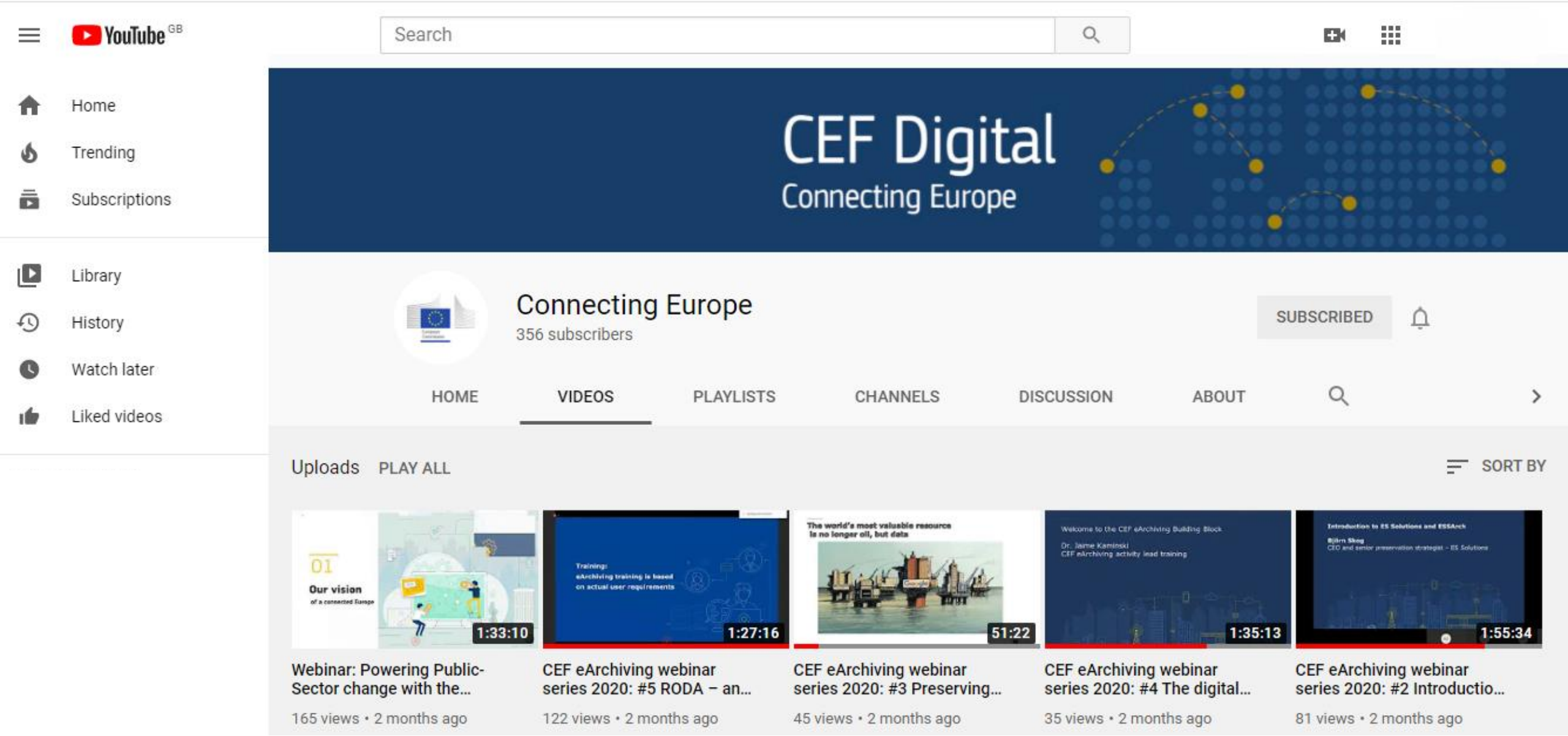
Content will include:

- The core specification (CSIP)
- Key Content Information Type Specifications (CITS):
  - Database (SIARD)
  - Geospatial data
- Two end-to-end eArchiving systems
  - ESSArch
  - RODA
- Database preservation (DBPTK)
- Validation/compliance
- The E-ARK Web user interface guide

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



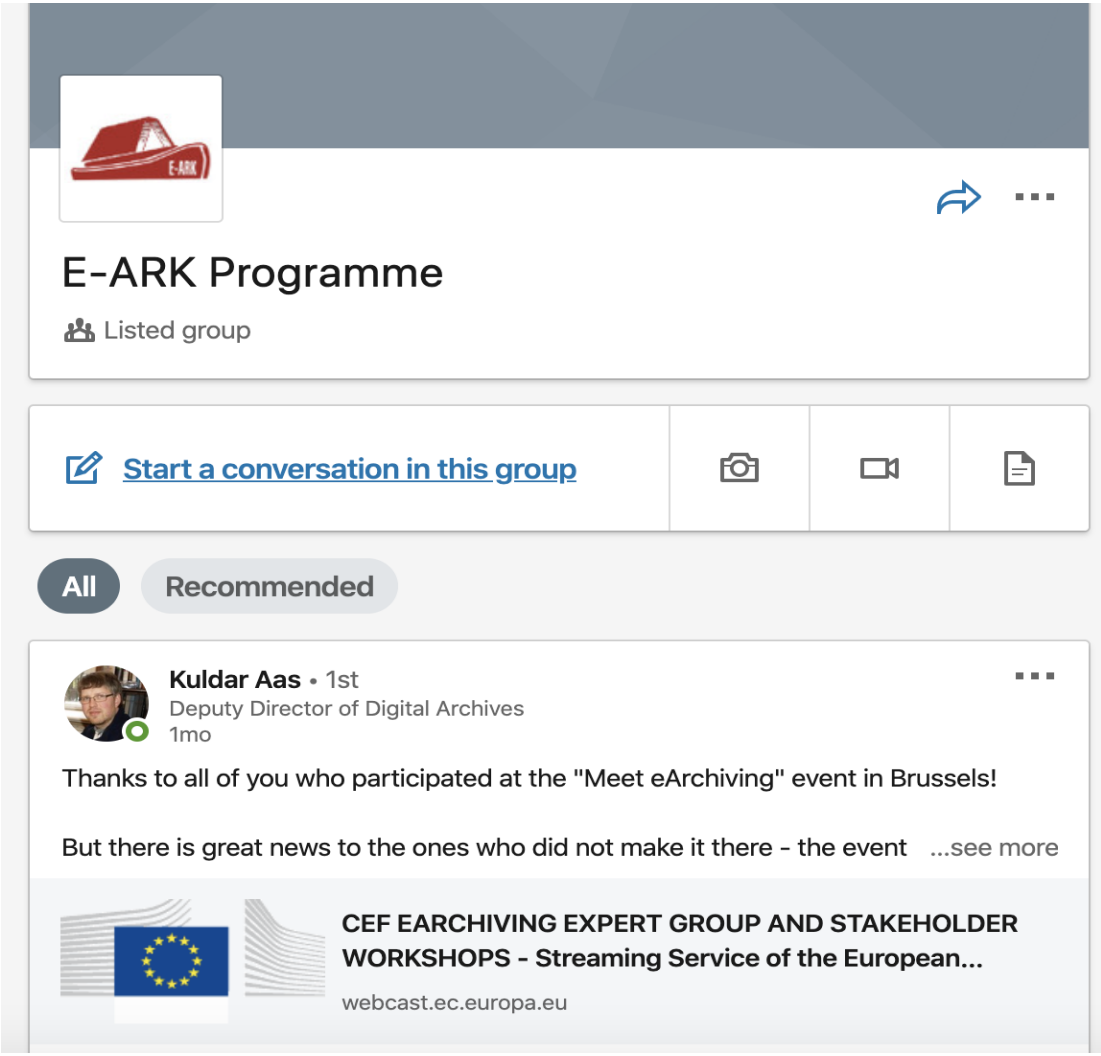
https://www.youtube.com/channel/UCaPOT_MBdE-kL5AJQzrCBDw/videos?view=0&sort=dd&flow=grid

European Commission

eArchiving outreach

- Webinars
- Workshops
- LinkedIn group
- Twitter #EARKProject

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



European Commission

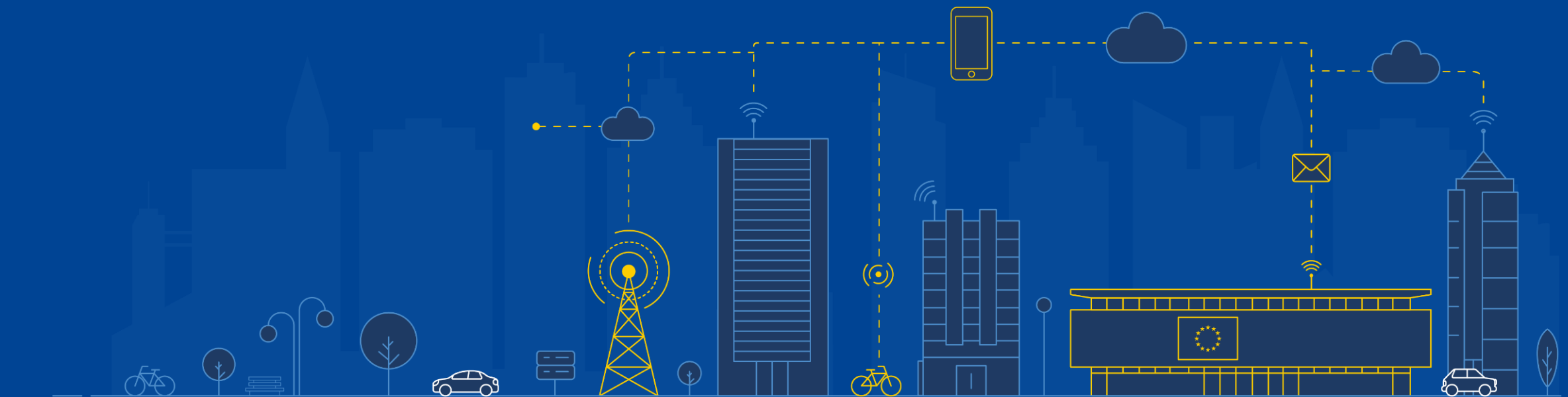
We want to hear about your requirements:

contact us at cef-building-blocks@ec.europa.eu

E-ARK validation

Carl Wilson
Open Preservation Foundation

Costas Simatos
DIGIT



What is E-ARK validation software and how might it help me?

Your presenters

Carl Wilson
Technical Lead at Open Preservation Foundation
E-ARK Validation Activity Lead

Costas Simatos
Solution architect at DIGIT
Technical Lead of the Interoperability Test Bed Action

Validation, conformance and compliance

The E-ARK validator provides instant package validation issues. The Test Bed instance is intended for use in conformance testing. The difference?

- **Validation**: The logical yes/no result after a test of an IP file against the structural tests, XML schema validation, Schematron rules and integrity tests

- **Conformance**: If an IP of each kind (collection, type, etc.) from a digital archive is valid (see above), then the Archive can claim #eArchiving Conformance

- **Compliance**: When a Digital Archive demonstrates evidence

Today’s webinar

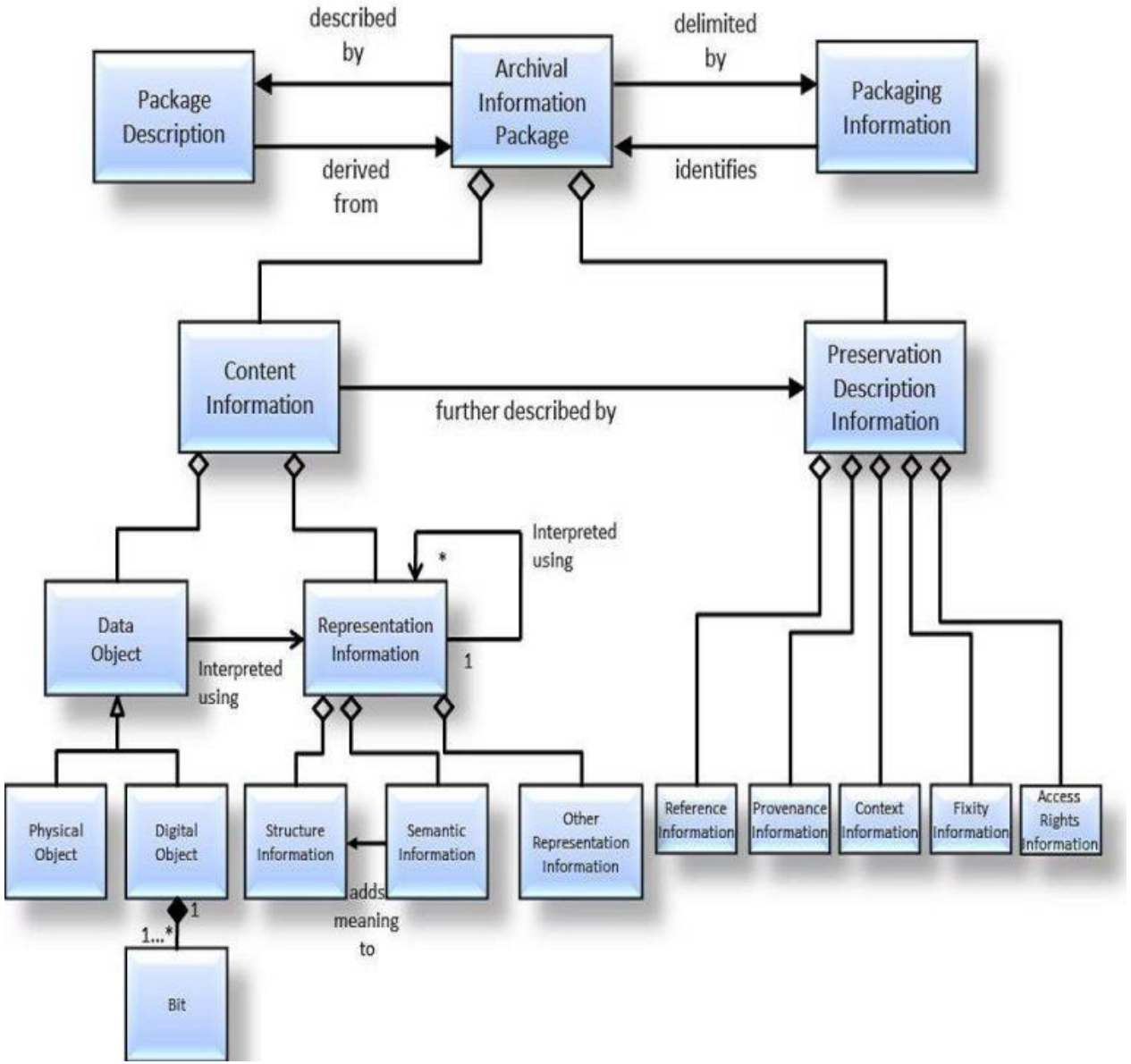
The purpose of today’s webinar is to help attendees

- Place information package validation in the digital archiving life-cycle context
- Understand E-ARK’s three-stage validation process for information packages
- Discover the capabilities of E-ARK validation software
- Learn how they can use the online validators to discover and fix issues in information packages
- Learn how they might use E-ARK validation tools in their organisation
- Understand eArchiving aims for conformance in the next year of the project

Information Packages: OAIS

An OAIS Information Package binds:

- **Content Information**: The information object that is the target of preservation
- **Preservation Description Information:** The information necessary for the adequate preservation of the Content Information

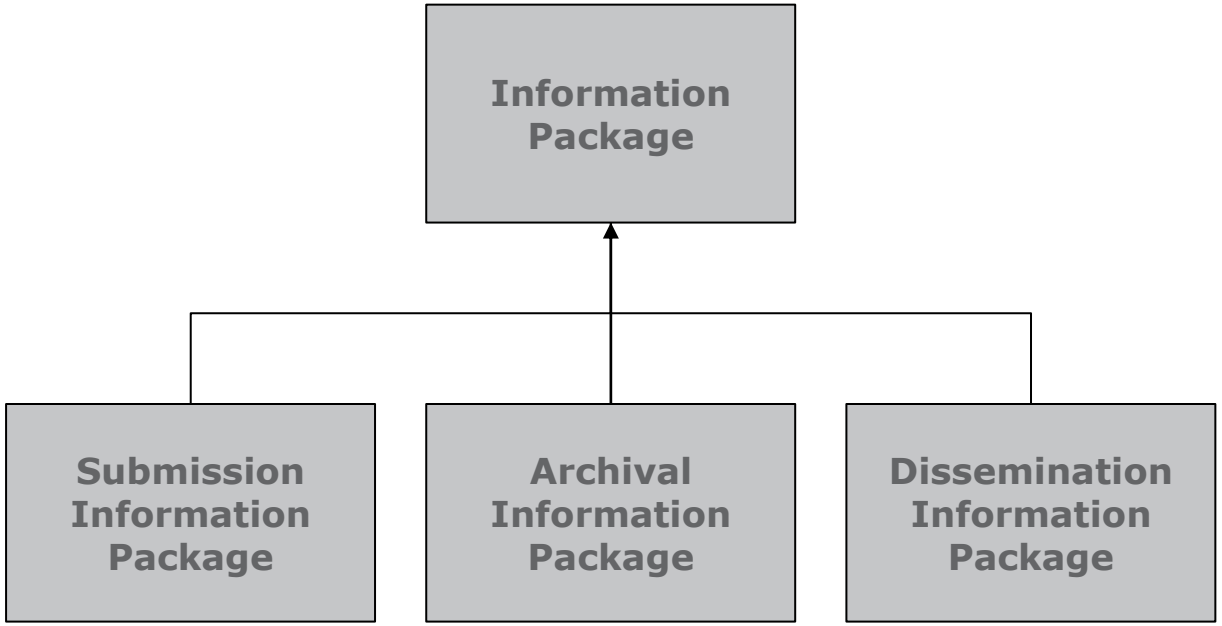


European Commission

Information Packages: OAIS

OAIS Defines three information package types:

- **Submission Information Package (SIP):** The package sent to an archive by a producer
- **Archival Information Package (AIP):** The package preserved by the archive
- **Dissemination Information Package (DIP):** A package derived from the AIP and sent from the archive to a consumer



Information Packages: E-ARK

OAIS defines **WHAT COULD** be in an Information Package, but says little about **WHAT SHOULD** be in there or **HOW** the package is arranged

- There is little guidance for implementers
- There is no basis for interoperability, information packages from different systems may have very little common ground

E-ARK sets out to address these issues by defining a formal logical and physical structure for information packages

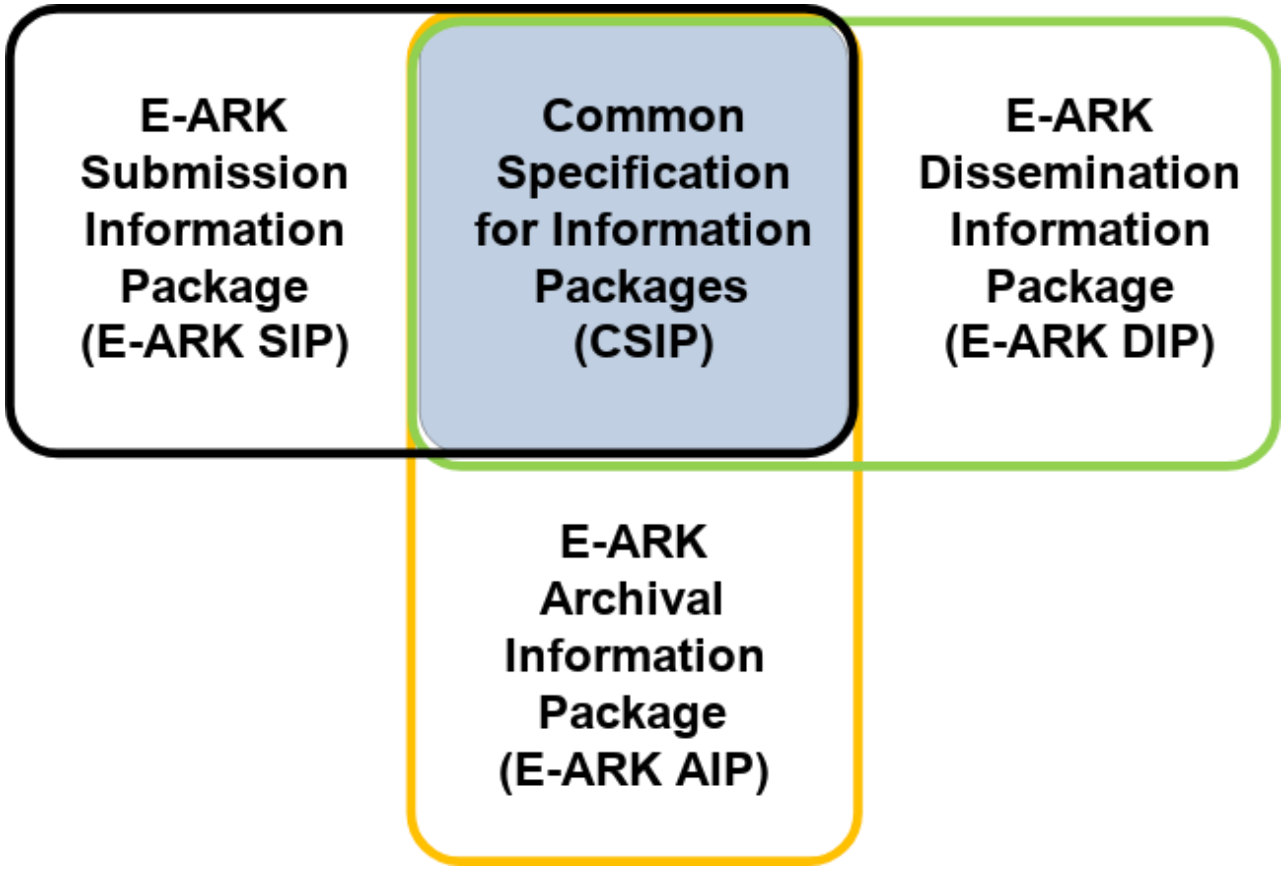
E-ARK IP specifications

E-ARK have produced machine validatable specifications for:

- Submission Information Packages
- Dissemination Information Packages

This is published as three documents:

- E-ARK CSIP (Common Specification for Information Packages)
- E-ARK SIP
- E-ARK DIP



IP validation: Lifecycle

IP Validation has utility across the entire life-cycle of an IP:

- SIP Creation by the producer
- SIP Submission by the archive
- AIP Creation by the archive
- AIP Transformation by the archive
- DIP Creation by the archive

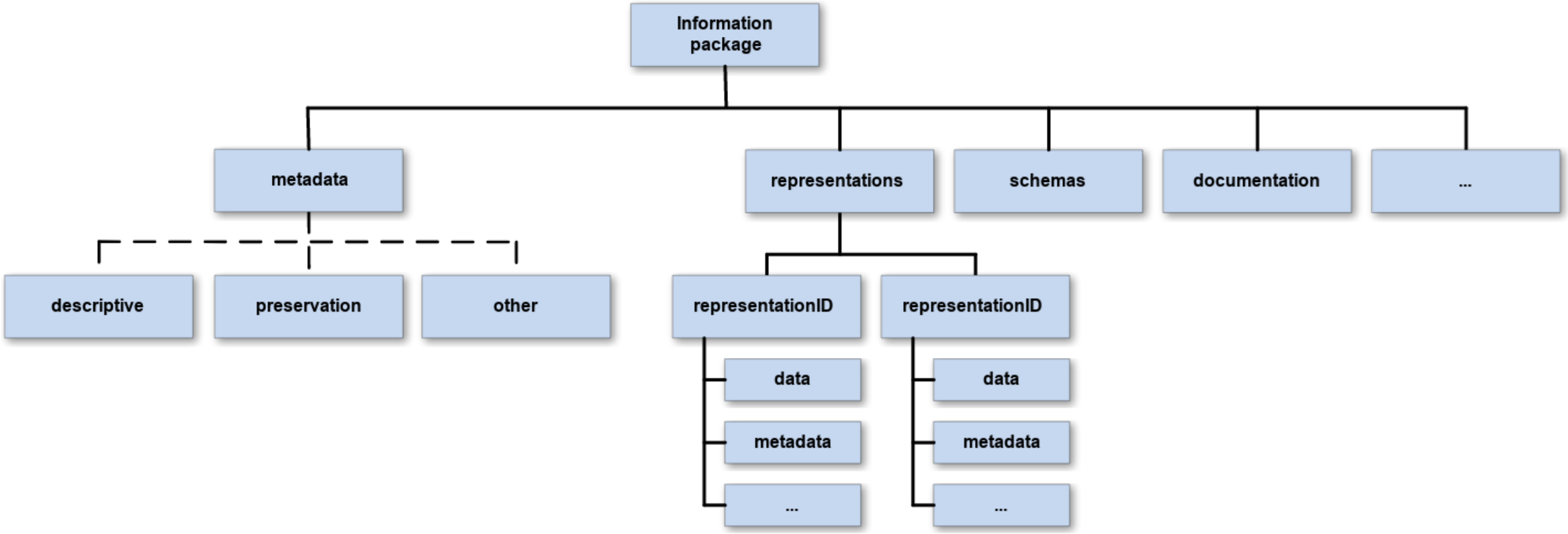
E-ARK IP validation: Three phase model

E-ARK separates IP validation into three logical phases:

- **Structure**
  Sixteen rules that ensure that the package structure is valid, this covers the presence of specific directories and a main metadata file
- **Metadata**
  Schema validation of metadata files followed by an extended set of metadata checks, currently about 150 rules
- **Integrity**
  Checks the package manifest, ensures that all files are present, verifies any checksums and ensures that no “orphaned” files are present

E-ARK IP validation: Structure

The E-ARK IP specifications define a standard physical structure:



E-ARK IP validation: Metadata

Metadata validation covers:

- Validation of XML METS files against the METS schema
- Additional validation against the E-ARK extension schema for additional attributes defined by the project
- Validation against an extended rule set that is considerably more prescriptive than the METS schema
  - Published as a METS profile available from the specification sites
  - Enforced through the use of XML Schematron

E-ARK IP validation: Integrity

Integrity checking:

- Ensures all files listed in the metadata documents are present in the package
- Verifies the checksums of these files against those in the metadata, if present
  - This can be time consuming for large packages
  - Checksum validation is an optional step
- Ensures that there are no orphaned files in the package:
  - An orphan is any file that is not referenced in the metadata somewhere
  - Possible that the file does not belong in the package
  - Alternatively the metadata for the file has been omitted

E-ARK: Content Information Type Specifications

Information Package validation as defined is a valuable first step, but it is just a first step ensuing:

- A package conforms to a standard structure
- All package level metadata is present and valid
- All package content is present and is referenced in metadata

It does nothing to validate any of the content or lower level metadata in a package.

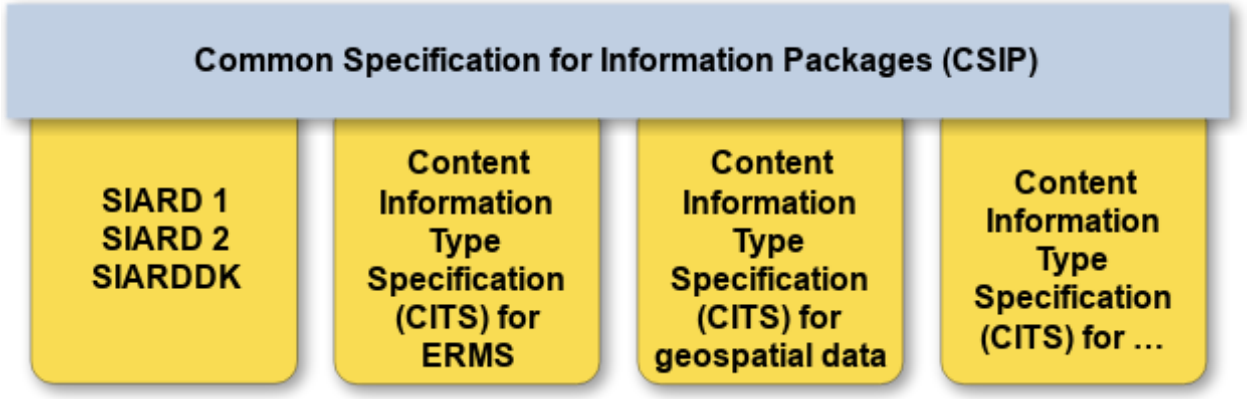
To address this E-ARK has also produced Content Information Type Specifications (CITS)

E-ARK: Content Information Type Specifications

CITS provide extended rules for specific content types for example:

- SIARD for database preservation
- Geospatial data
- Electronic Records Management Systems data

These documents are currently out for public review



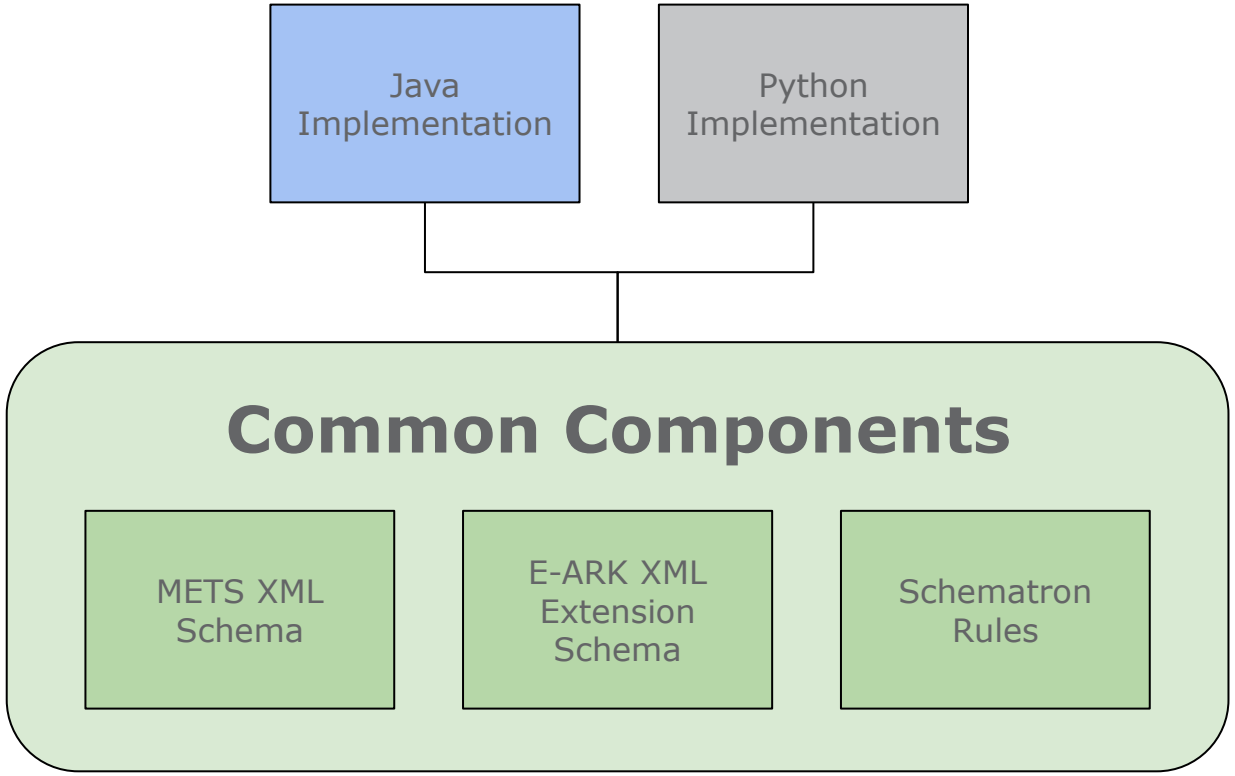
Validation software: Components

E-ARK provides standard reference software components for validation:

- A Python implementation
- A Java implementation

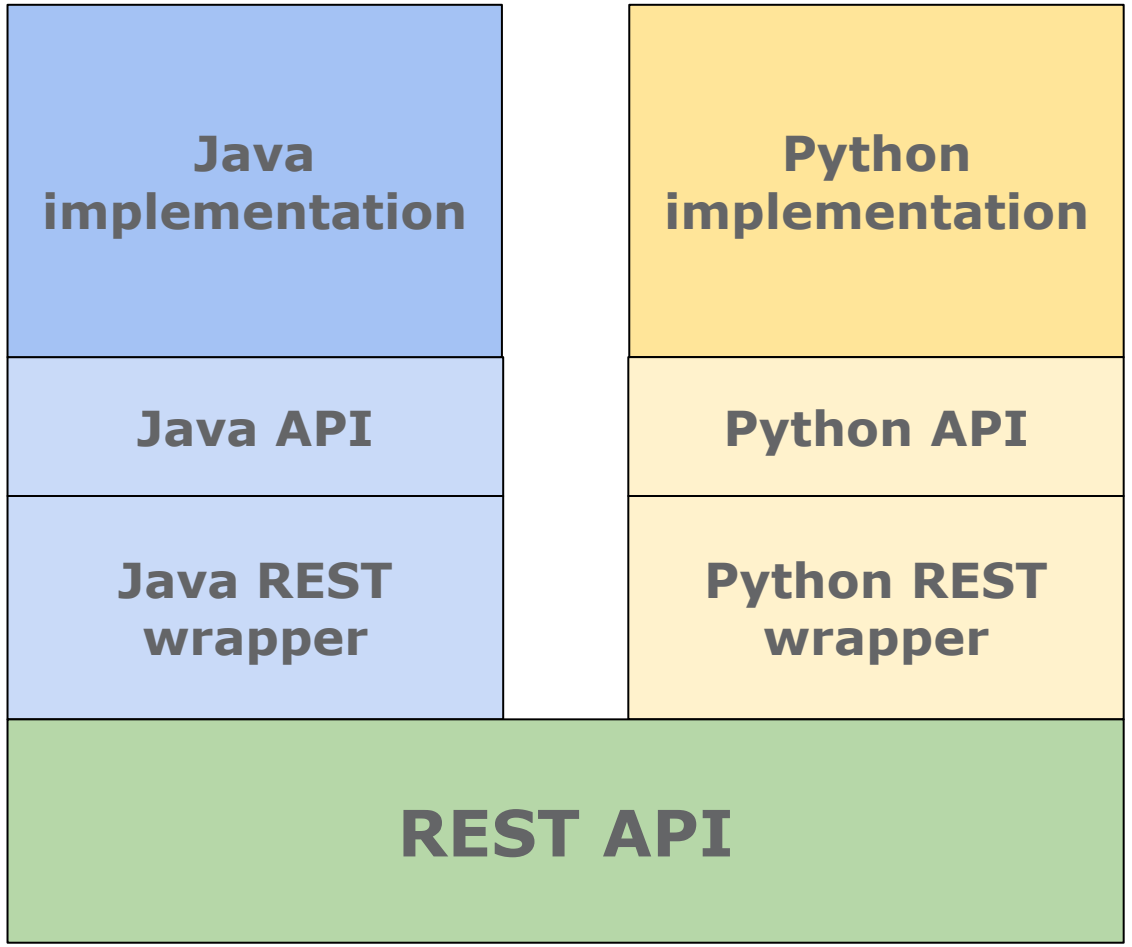
These share common sub-components:

- XML schema for METS
- E-ARK's own extension schema
- XML Schematron rules



Validation software: Common API

- Each implementation provides a set of language-specific methods and data types for use in native code
- Additionally there is a REST wrapper for each implementation that provides a common REST API
- The REST API is defined as Open API v3 (formerly known as Swagger)



Online services

There are two online validation services available:

- **E-ARK online validator**
  A stateless instance that runs synchronously
  (i.e. upload an information package and get the validation result)

- **CEF Test Bed instance**
  A testing platform to assess conformance based on test scenarios, with results being recorded to allow reporting and monitoring

E-ARK Validator Demonstration

CEF Test Bed Instance

For CEF e-Archiving:

- Conformance testing is organised in scenarios used to validate conformance statements
- The E-ARK validator is used as a building block
- The current setup is minimal but is planned for further extension

Uses the ISA² Interoperability Test Bed, an online and self-service platform offered by DIGIT for the conformance testing of software against technical and semantic specifications

CEF Test Bed Demonstration

eArchiving aims for conformance in the next year of the project

CITS timetables

Release plans for 2021

Questions?

Carl Wilson
Technical lead

Open Preservation Foundation
carl@openpreservation.org

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!