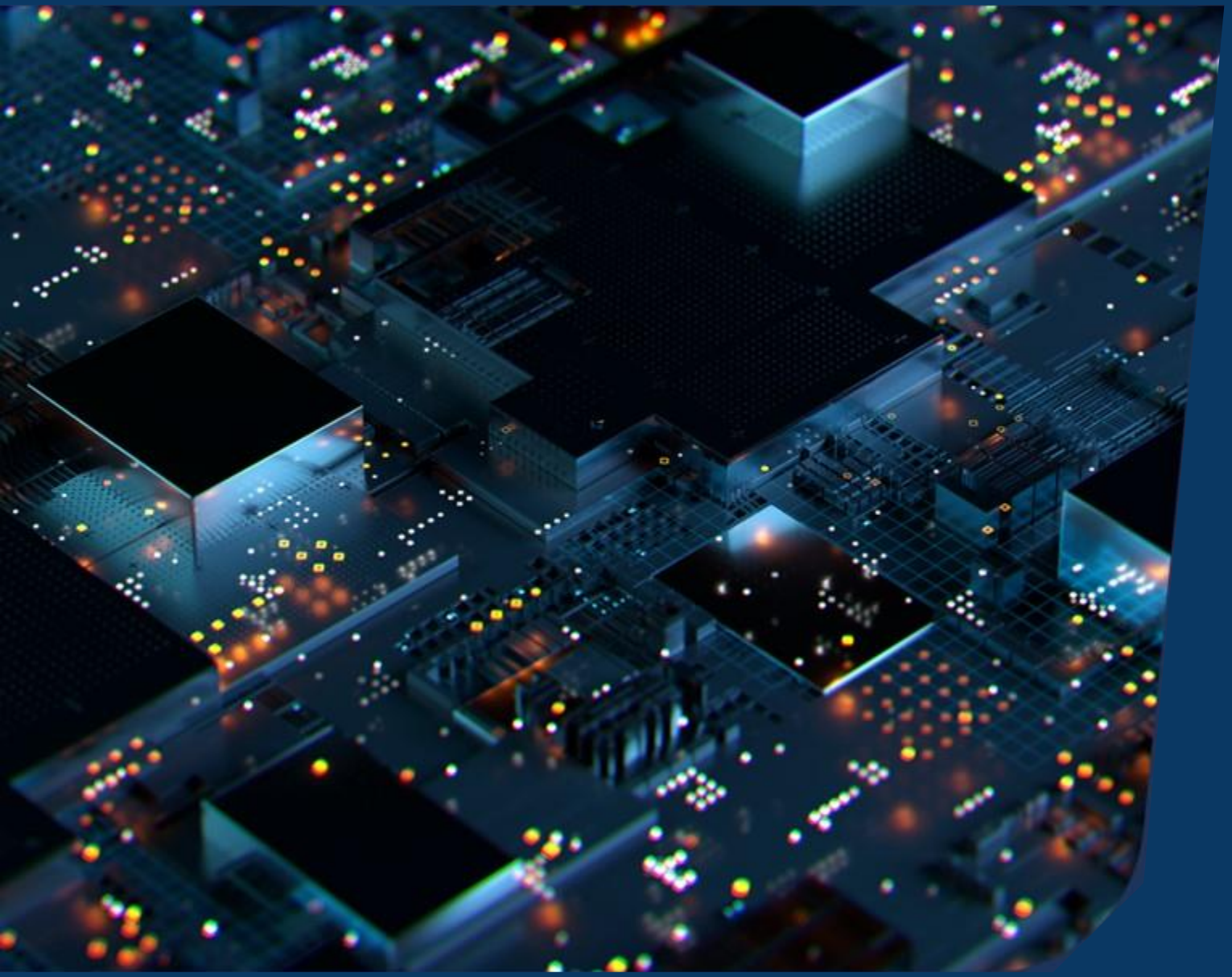


eArchiving Webinar #16

E-ARK Validation: What's inside the package?

Start 10:00 (CET)

30 September 2021

Agenda

10:00 – 10:05
Welcome
Pawel Stech, CEF Stakeholder Management Office – DIGIT

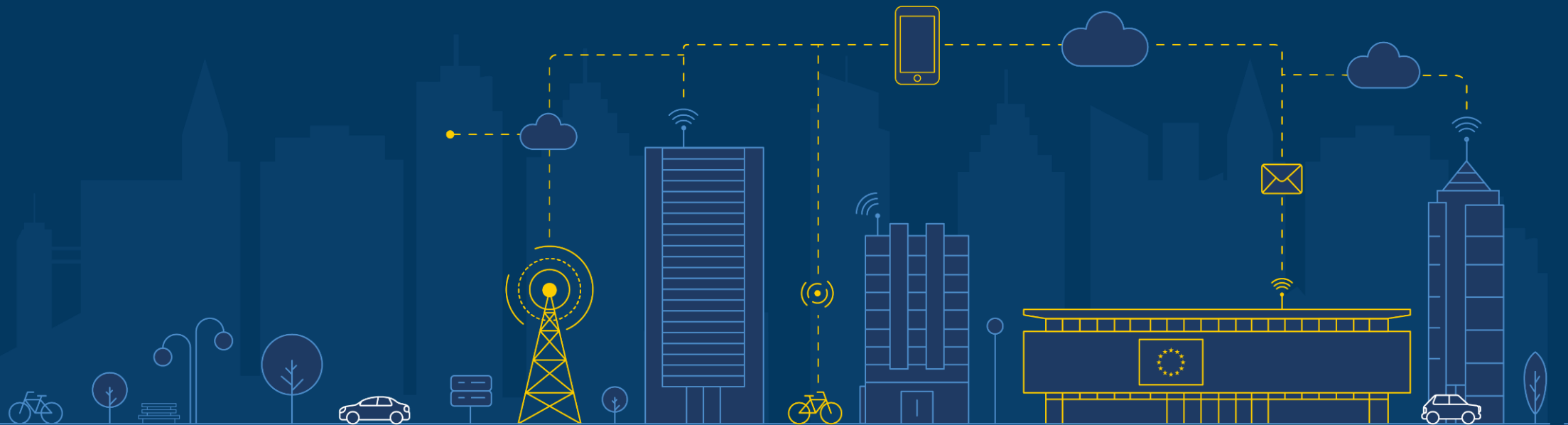
10:05 – 10:15
eArchiving Building Block welcome
Jaime Kaminski, eArchiving Building Block training lead

10:15 – 11:00
E-ARK Validation
Carl Wilson, Open Preservation Foundation
Costas Simatos, DIGIT

11:00
Q&A

Welcome

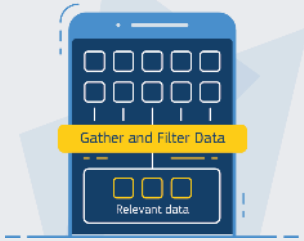
Pawel Stech
CEF Stakeholder Management Office – DIGIT



Building Blocks supported under the CEF Digital programme 2014–2021

Big Data Test Infrastructure

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

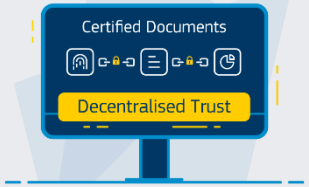


Context Broker

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

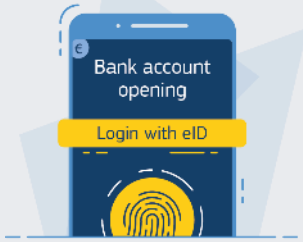


European Blockchain Services Infrastructure

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

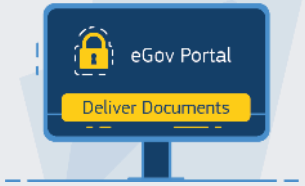
eInvoicing

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



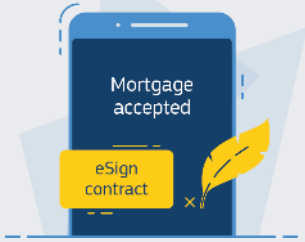
eID

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



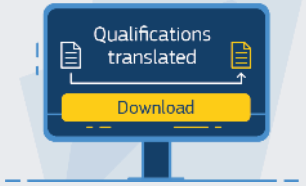
eDelivery

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

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

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

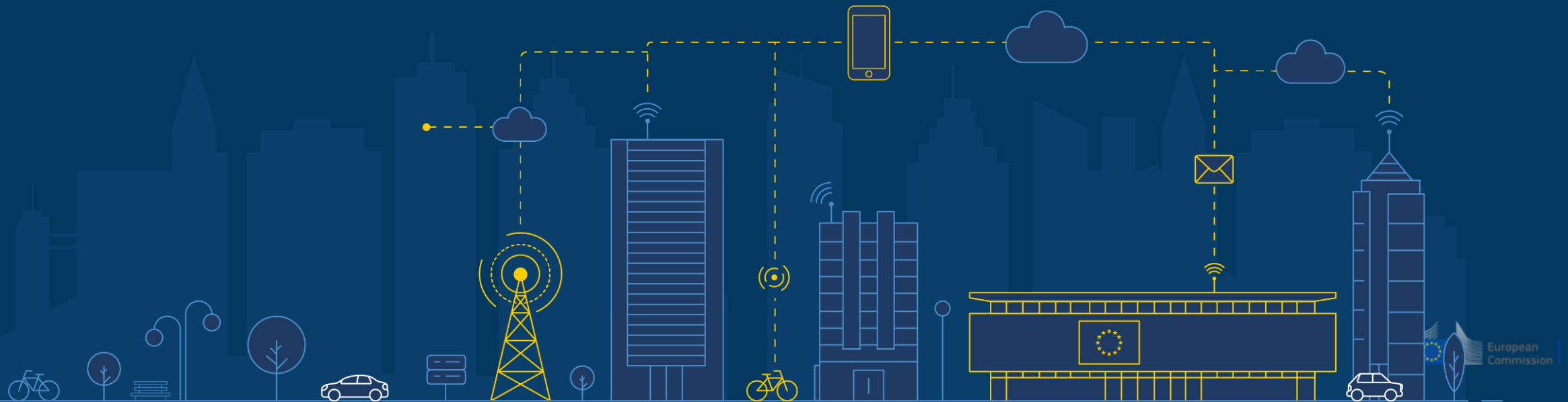
Reuse
Sample software available on CEF website

European Standards

Welcome to the eArchiving Building Block

Dr Jaime Kaminski

CEF eArchiving activity lead training



eArchiving Building Block website

CEF Digital Connecting Europe

About us | Building Blocks | DSIs | CONTACT US

eArchiving • Get Started | Services | Documentation | Grants | Support

CEF Digital

eArchiving

Facilitate the preservation, migration, reuse and trust of your information

WATCH THE VIDEO

GET STARTED

eArchiving in use

38 projects

23 Projects reusing eArchiving

15 Projects committed to analyse or reusing eArchiving

VIEW FULL STATISTICS

eArchiving helps Swedish Customs to preserve its records

YESTERDAY AT 1:30 PM

CEF eArchiving: Review of eArchiving procedures open

SEP 10, 2020

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

eArchiving Building Block

eArchiving services:

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

eArchiving Webinar #17

"

eArchiving geospatial records: Approaches and benefits

Thursday, 21 October 2021

Ann-Kristin Egeland – The Danish National Archives
Gregor Završnik – Geoarh

An introduction to the eArchiving Content Information Type Specification (CITS) for geospatial data. Questions from adopters of the CITS Geospatial specification will be used to explain and illustrate selected central requirements of the specification in more detail.

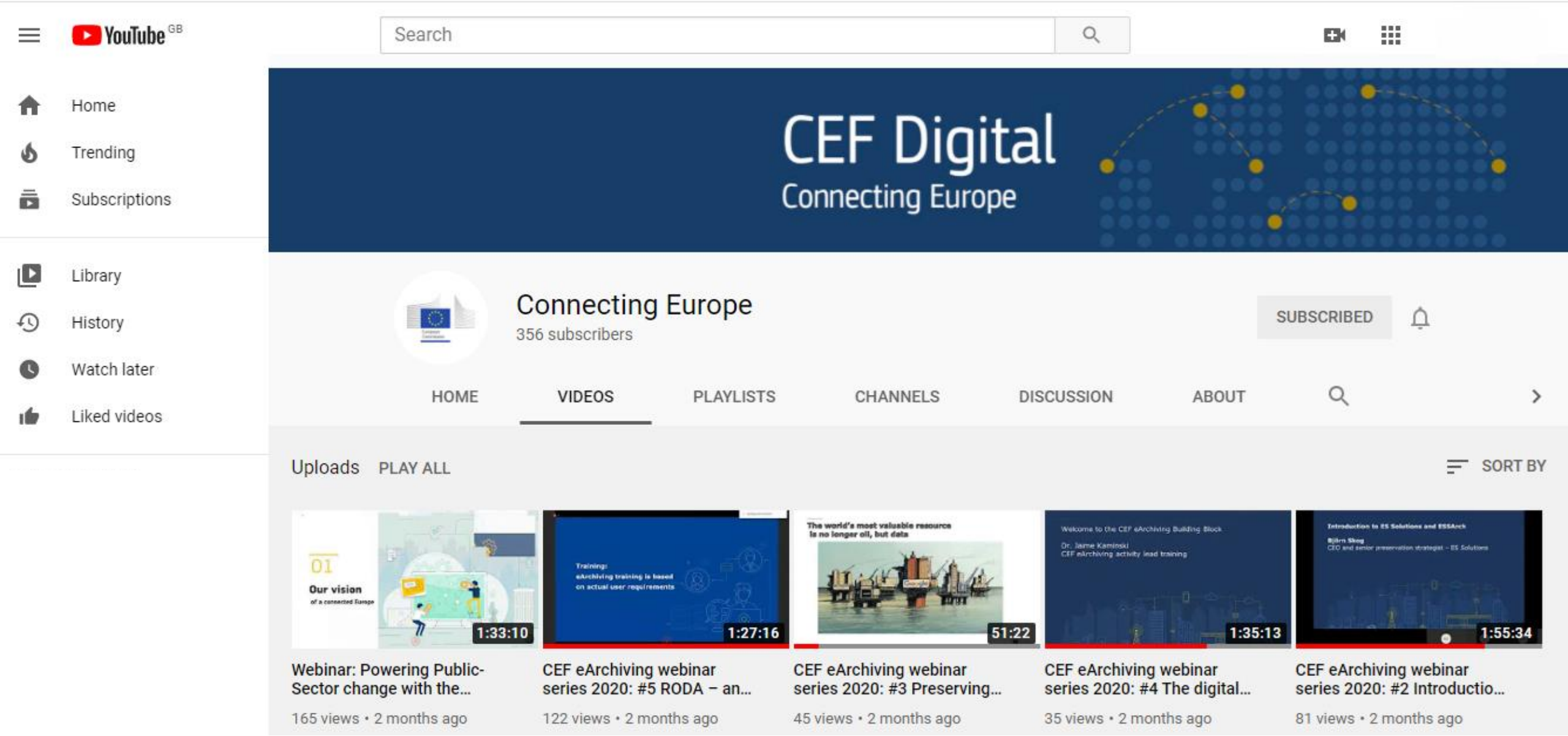
This is an opportunity for attendees to learn about:

- The eArchiving Content Information Type Specification (CITS) for geospatial data
- How your organisation can start using the eArchiving specifications for long term-preservation of geospatial records
- Challenges that organisations face when adopting the specifications and how they can be approached

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

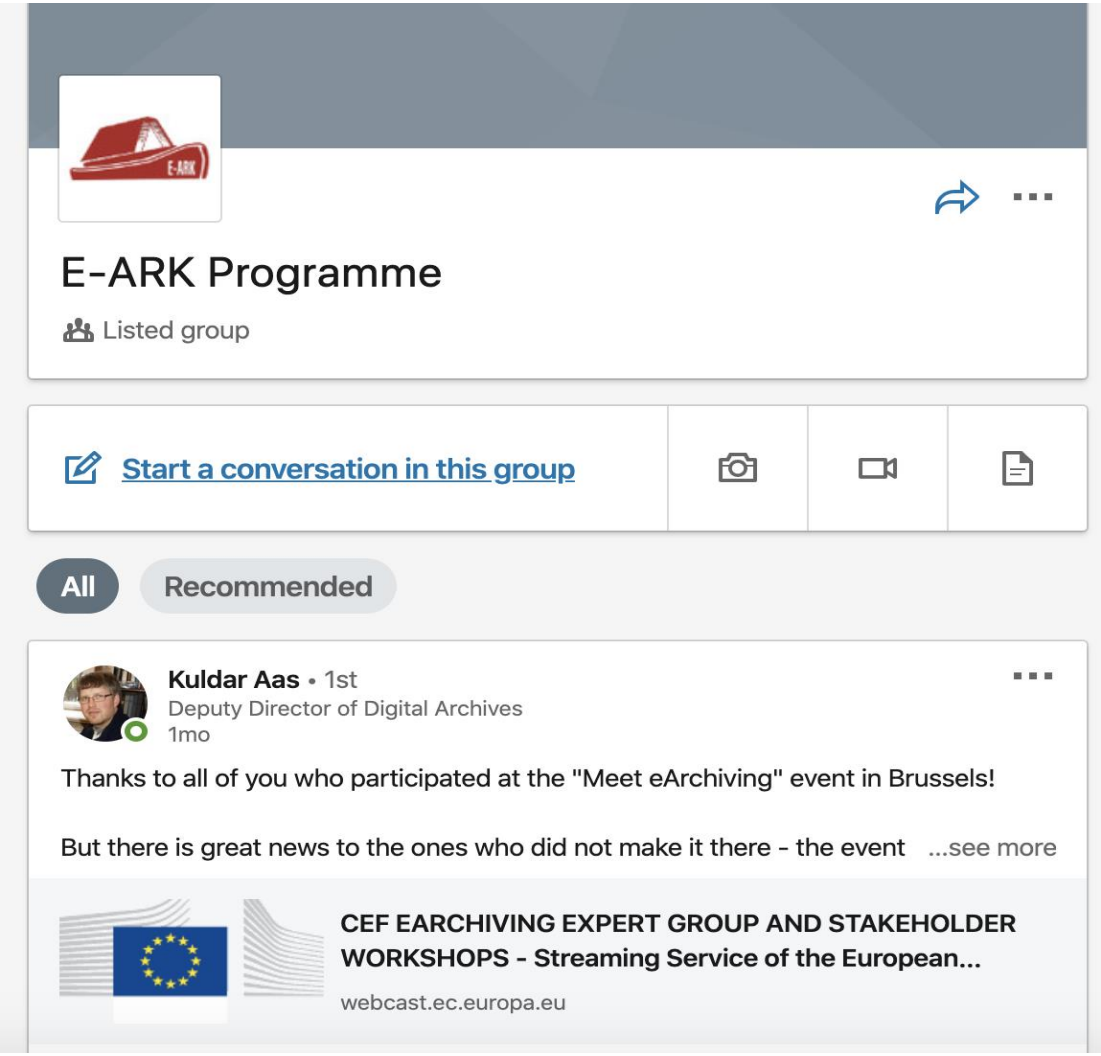


https://www.youtube.com/channel/UCaPOT_MBdE-kL5AJQzrCBDw/videos?view=0&sort=dd&flow=grid

eArchiving outreach

- LinkedIn group
- Twitter #EARKProject

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



Our speakers

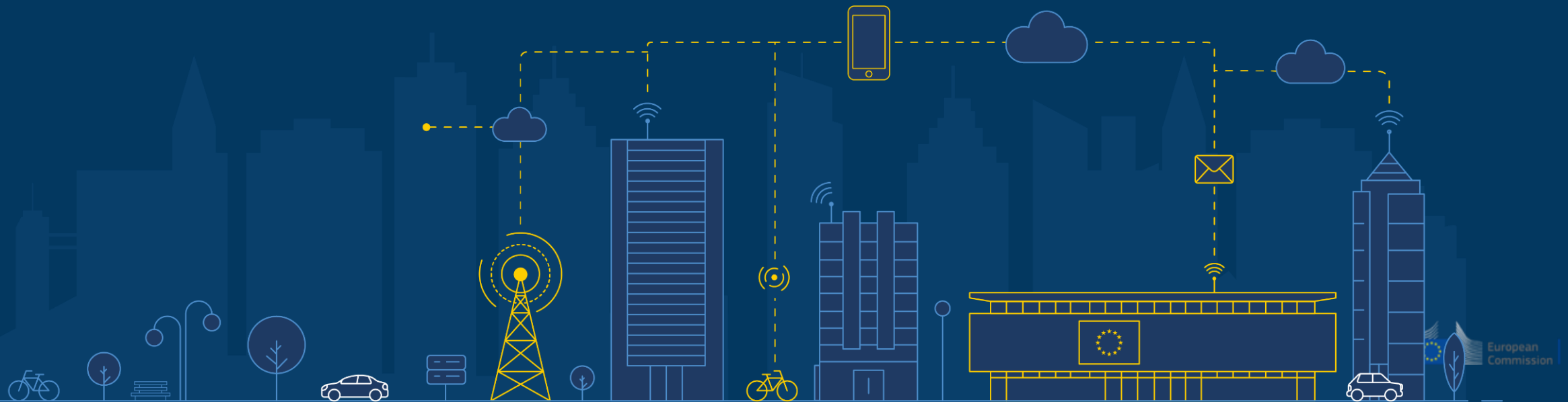
Carl Wilson
OPF

Costas Simatos
DIGIT

E-ARK Validation

Carl Wilson
Open Preservation Foundation

Costas Simatos
DIGIT



Webinar Programme

- **Specifications, Standards, Compatibility & Interoperability**
  - Compatibility through standards, advantages and limitations
  - Why interoperability is better than compatibility
  - Interoperability, it’s all in the details
  - The hidden advantages of standards
- **E-ARK specifications, from compatibility to interoperability**
  - It all starts with a CSIP, the base archival package
  - Special considerations for SIPs and DIPs
  - What’s in the AIP?
  - The role of Content Information Type Specifications, there’s always more
- **Validation and Validators**
  - E-ARK validation and compatibility
  - CITS validation, to interoperability and beyond
  - Conformance and compliance
  - What’s improved with the new validator, fly past demonstration of validation tools

Defining Terms

We'll start by defining some terms we use in this presentation and their relationship to each other.

- **Compatibility**
  Software and data that get along
- **Interoperability**
  Software and data that work together
- **Specification**
  A detailed description to make the world a better place
- **Standard**
  A good specification that gets used

Compatibility and interoperability

Compatibility: the art of getting along

- Software and data that are compatible when they can be used together without breaking
- Often achieved through permissiveness and ignoring what’s not understood

Interoperability: working together

- When software and data can be used together to add value
- The greater the specified level of detail the more interoperable

Specifications and Standards

Specifications

- Provide a detailed description of a design, arrangement of data, etc.
- Specifications become standards through authority and use
- Good specifications should be:
  - Generally applicable enough for meaningful adoption in the domain
  - Specific enough to allow for interoperability between adopters
  - Modular and extensible to allow implementers to start small and progress

Standards

- A standard is really just a good specification that gets used

Interoperability, it’s all in the details

Compatibility can be achieved without having to provide extensive detail. A plug that fits a socket is compatible, but is it interoperable?

- What if the voltage is different?
- Or the amperage?
- Which pin is the earth?

Sometimes superficial “compatibility” can even be dangerous.

We need detail and completeness for real interoperability.

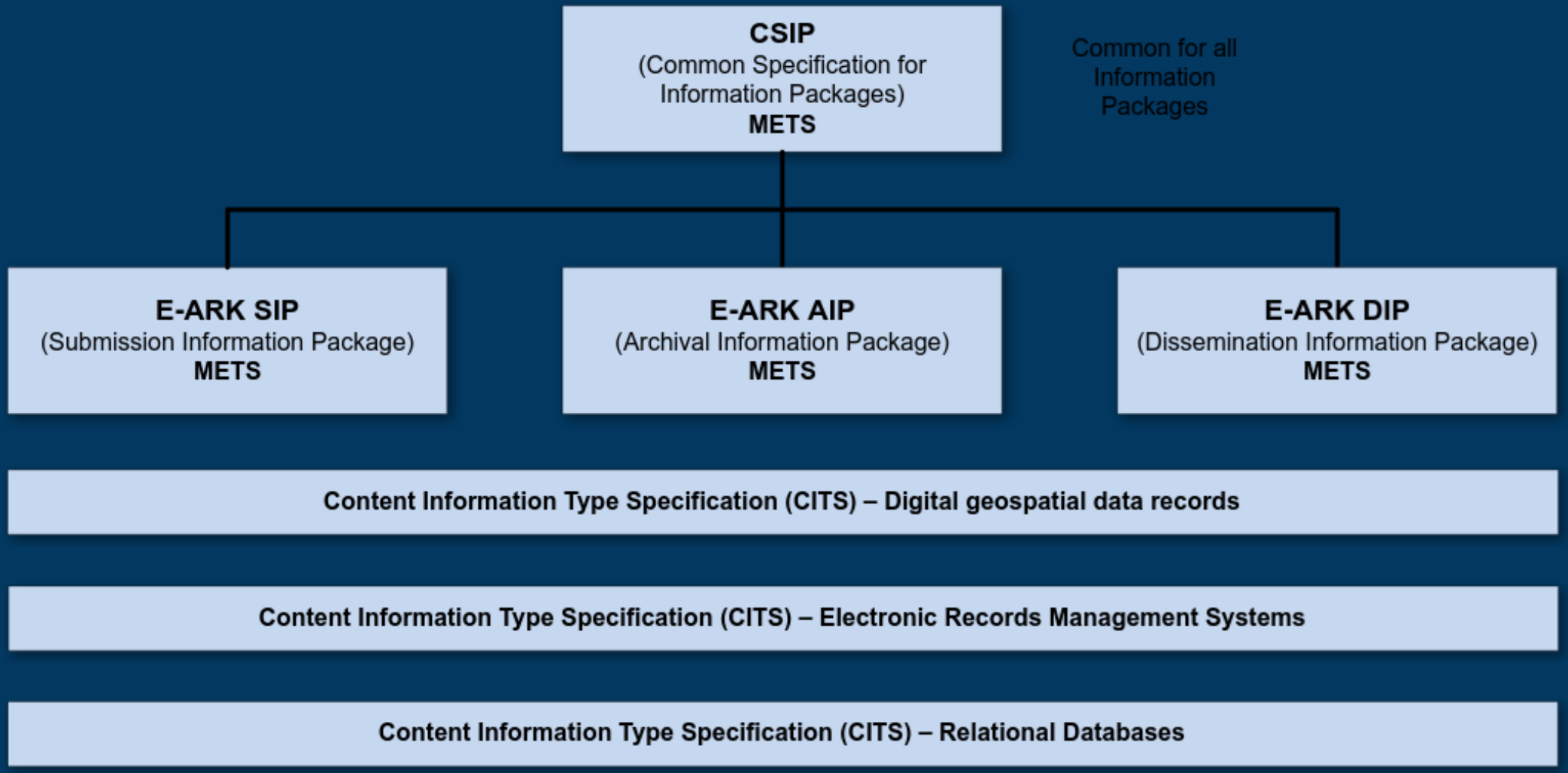
The hidden advantages of public specifications

- It allows you to build on the knowledge of others
- There’s bound to be things that you hadn’t thought about
- Everyone who uses them helps to improve them
- You don’t have to write your own documentation

E-ARK specifications: compatibility to interoperability

- It started with a CSIP, the base archival package
- Special considerations for SIPs and DIPs
- What’s in the AIP?
- The role of Content Information Type Specifications

It started with a CSIP, the base archival package



Special considerations for SIPs and DIPs

SIPs have their own rule set

- Extension set of 35 requirements which override or further restrict the 150 CSIP requirements.
- Unsurprisingly these concern themselves with submission concerns, e.g. submission agreements.

DIPs also get their own rules

- Minimal extension set of 5 requirements
- These are packages derived from an AIP.

What’s in the AIP?

The AIP is the submission package with whatever archival life-cycle metadata and data additions the package accumulates:

- Package transformations
- Content / metadata updates
- Export format for dark archive / system upgrade

An updated AIP specification will be published in October 2015

The role of Content Information Type Specifications

The IP specifications provide compatibility with a thin layer of interoperability

- Interoperability is at a basic archival metadata level, key METS fields
- Metadata and content is only compatible
  - The specifications describe where to put content, metadata and supporting documentation
  - The specifications place no restrictions on the format of metadata and content in a package
- For this E-ARK provides Content Information Type Specifications (CITS)
  - These take us from compatibility to interoperability
  - They provide detailed descriptions of metadata and/or content for a particular domain and purpose
  - Intended for modular extensibility for the IP specifications.

Validation and Validators

- E-ARK validation and compatibility
- CITS validation, to compatibility and beyond
- Conformance and compliance
- What’s improved with the new validator, fly past demonstration of validation tools

E-ARK Validation and Compatibility

What does E-ARK IP validation look-like?

- Ensures that a package is structured properly
- Ensures that the master package metadata is well-formed and valid
- Will checksum files to test for corruption if desired (this can take a long time)
- Ensures that a package is self-contained and complete(ish)

Demo of E-ARK Validation: Issues and Detection

The E-ARK Validation service is available at

https://pyip.openpreservation.org/

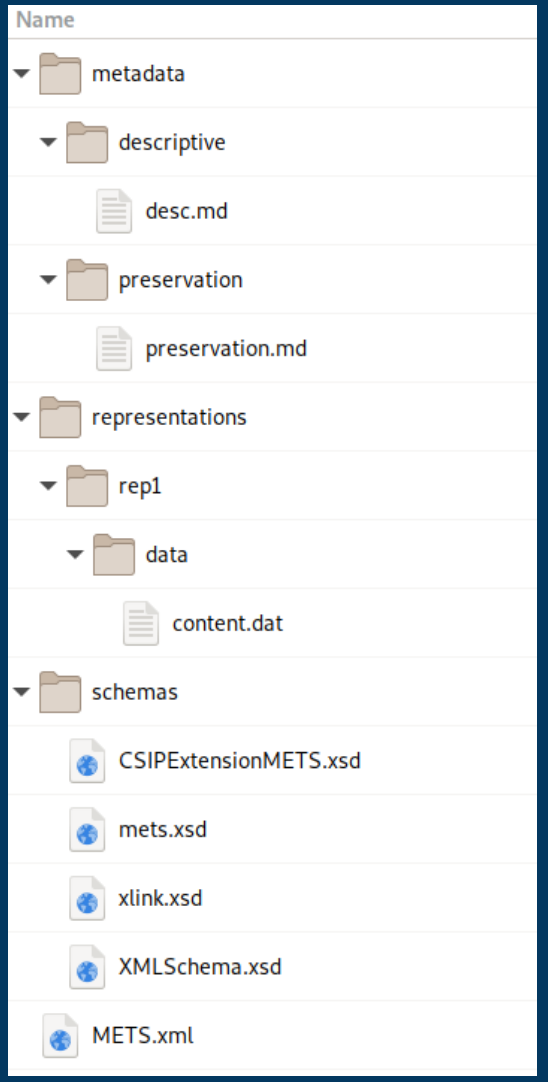
A Minimal E-ARK Package

A valid package doesn’t require much

- A package level METS file
- Some key METS fields filled in

Bonus points for

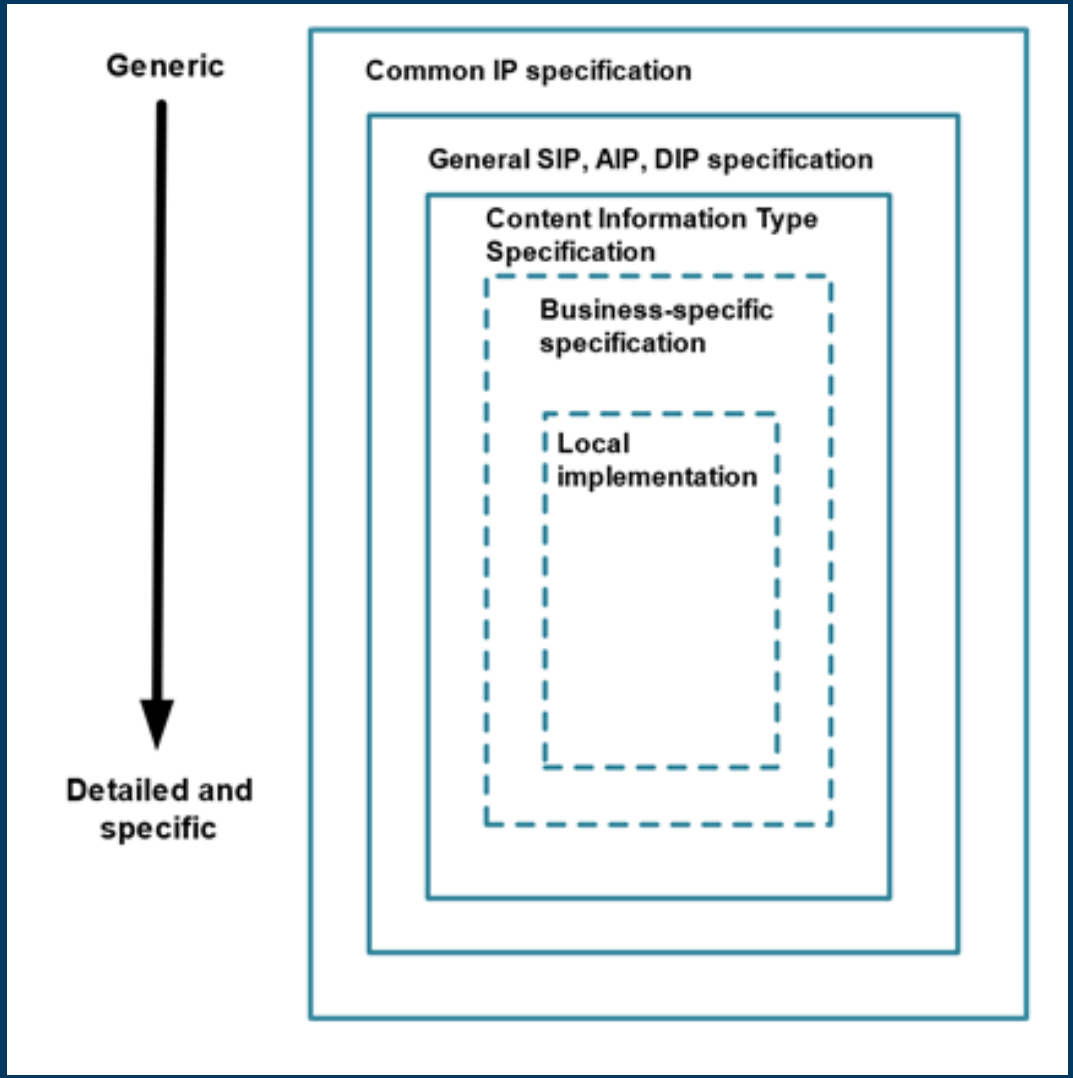
- Representations in dedicated discrete folders
- Metadata in dedicated folders
- Supporting schema in dedicated folder



CITS Validation, to Compatibility and Beyond

Content Information Types Specifications add meat to the IP specification bones.

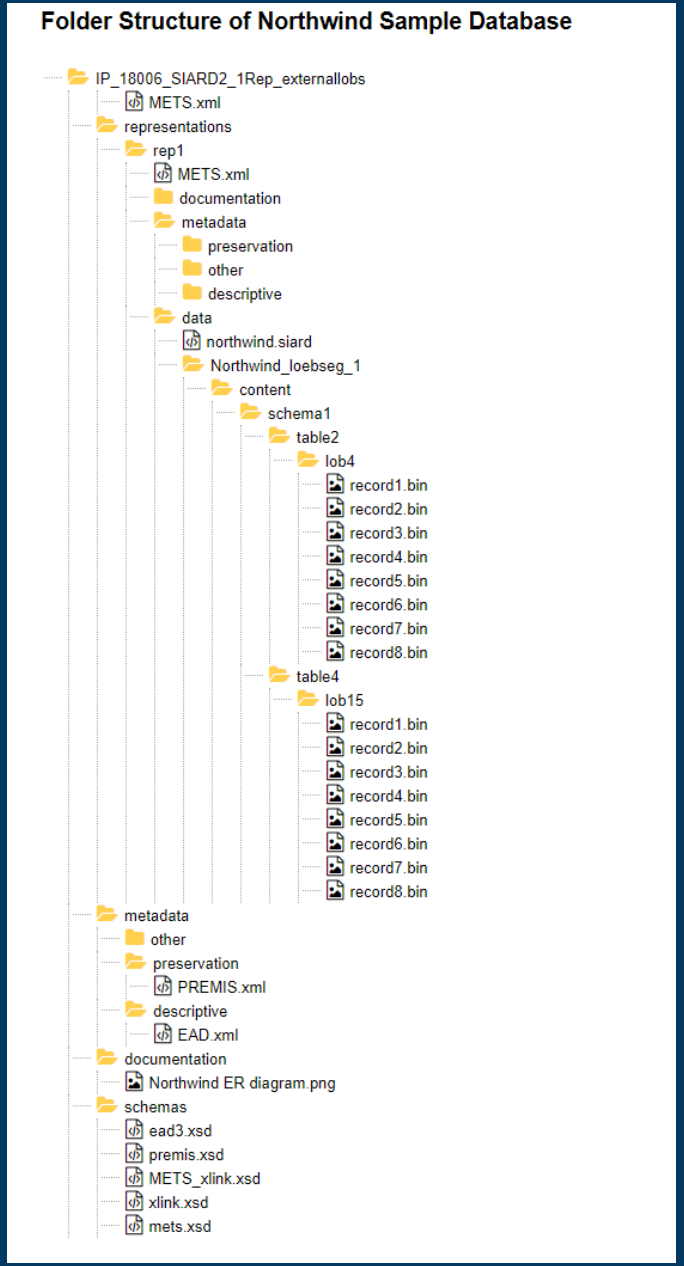
- IP validation tells you that you have content and metadata and which is which
- IP metadata relationships between metadata and content, i.e. this item is described by but says nothing about the form of either
- A CITS prescribes the form of metadata and/or content in the package



A SIARD CITS Information Package

The SIARD CITS provides another 33 rules covering:

- Package metadata
- Specific metadata for DB preservation
- Using SIARD as an archive format for DBs
- Splitting large databases



Validation, Conformance and Compliance

E-ARK's definitions of these terms

- **Validation**
  Refers to the result of a validation test of some entity, usually an information package, against a formal specification. This can be viewed as a logical yes/no result, although the generation of warning and informational messages is also possible.
- **Conformance**
  Describes the level to which an archive or software component conforms to E-ARK specifications. If an information package of each type stored in a Digital Archive, e.g. collection, type, etc., is validated (see above), then the Archive can claim a level of eArchiving Conformance.
- **Compliance**
  Is the term used when a Digital Archive, or software component demonstrates evidence of Conformance (see above) and further requirements.

Put Another Way

- **Validation**
  Validation gives “spot” answers as to the validity of an information package at a particular point in time.
- **Conformance**
  Can be viewed as “validity coverage” for a particular software component, that is package validity is established at different points in the life-cycle across different types of packages.
- **Compliance**
  What might “further requirements” to the above look like?

The CEF Testbed Service

The CEF Test Bed service provides a platform to test, verify and monitor conformance and compliance.

- It is complementary to the e-ARK validator.
- The Test Bed defines test scenarios that can provide such context, ensuring that archives are not only valid but that they also match the test scenarios expectations.
- The current setup in the Test Bed is more of a proof of concept as it simply passes archives to the validator and reports on their validation result. It can be extended however in the future to include the discussed scenarios, potentially with several test cases, each expressing a different scenario.
- Access to the Test Bed is account-based, with test sessions and reports stored to allow monitoring and reporting (by testers and administrators).
- The CEF Test Bed is available at https://www.itb.ec.europa.eu/cef/itb/.
- A party can currently register to try it by clicking on the “Register in a public community” link and selecting the “CEF eArchiving community” (an EU Login account is necessary to register).

Scenarios for Compliance

Compliance would be tested using established scenarios, for example:

- A SIP consists of a sequence of TIFF page scans making up a document
- The package contains structural metadata that places the TIFF masters in sequence
- Policy states that on ingest a repository creates a PDF of JPEG conversions of the master images as an access copy
- Technical metadata derived from the TIFF masters, the access copies and the conversion process should be added to the package

What’s Coming?

- Commons IP v2.0.0: Validation library for Java, beta out now: https://github.com/keeps/commons-ip/releases/tag/2.0.0-beta1
- Updated REST validator and Python Library 15/10/2021:
  - REST Endpoint: https://pyip.openpreservation.org/
  - Docker REST: https://hub.docker.com/repository/docker/eark4all/rest-ip-validator
  - Docker CLI: https://hub.docker.com/repository/docker/eark4all/cli-ip-validator

Questions?

Carl Wilson
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!

European Commission