The eArchiving Initiative

Ensuring interoperable and sustainable access
to geospatial records

Gregor Završnik, Geoarh

Introduction

- Motivations for the eArchiving initiative

- What is the eArchiving initiative

- How do you use eArchiving for geospatial records?

- What is next?

European Commission

Motivations for the eArchiving initiative

Why do we need specifications for digital preservation ?

And why backup is not archiving.

European Commission

Long term preservation challenges

- **Missing context, documentation**

Parcels Feature class

Shape	ID	PIN	Area	Addr	Code
	1	334-1626-001	7,342	341 Cherry Ct.	SFR
	2	334-1626-002	8,020	343 Cherry Ct.	UND
	3	334-1626-003	10,031	345 Cherry Ct.	SFR
	4	334-1626-004	9,254	347 Cherry Ct.	SFR
	5	334-1626-005	8,856	348 Cherry Ct.	UND
	6	334-1626-006	9,975	346 Cherry Ct.	SFR
	7	334-1626-007	8,230	344 Cherry Ct.	SFR
	8	334-1626-008	8,645	342 Cherry Ct.	SFR

Related ownership table

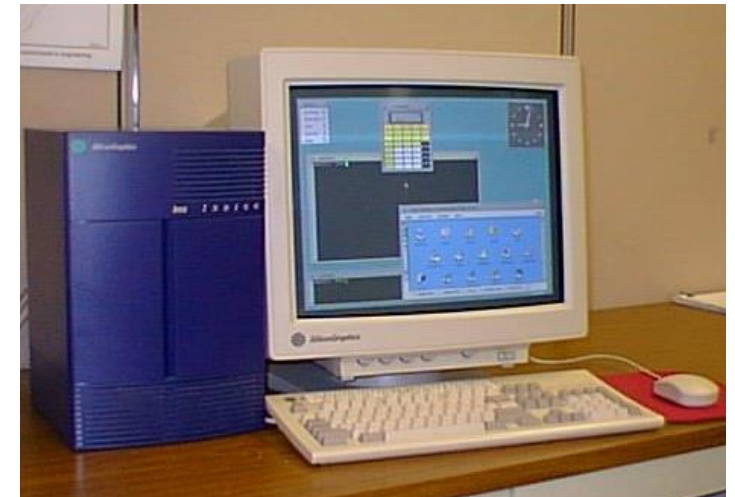| PIN	Owner	Acq.Date	Assessed	TaxStat
334-1626-001	G. Hall	1995/10/20	$115,500.00	02
334-1626-002	H. L Holmes	1993/10/06	$24,375.00	01
334-1626-003	W. Rodgers	1980/09/24	$175,500.00	02
334-1626-004	J. Williamson	1974/09/20	$135,750.00	02
334-1626-005	P. Goodman	1966/06/06	$30,350.00	02
334-1626-006	K. Staley	1942/10/24	$120,750.00	02
334-1626-007	J. Dormandy	1996/01/27	$110,650.00	01
334-1626-008	S. Gooley	2000/05/31	$145,750.00	02

https://desktop.arcgis.com/en/arcmap/latest/manage-data/geodatabases/table-basics.htm

- **Long term sustainable access**

By Photograph: Robert Jacek Tomczak - Own work, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=94360

- **Unreadable formats?**
- **Reusing old software?**

https://wiki.preterhuman.net/index.php?curid=1725

Long term preservation challenges
Use Case 1 - *Legal safety*

- Authenticity

  - How do we know data is authentic

  - How do we know we understand the data properly

- Provenance

  - What changes were made to the data

Long-term preservation challenges
Use case 2: Loss of Inter-agency references

- Water permit was issued in 2005

- The permit is issued to a Parcel number (as its location)

- Today this parcel doesn't exist

- How can we find it?

Long term preservation challenges
Use Case 3 – Information product recreation

GIS Tools

Symbology, geoprocessing, exports

Organization, Queries, Cartographic projections, transformations…

Vector data

Raster data

Database
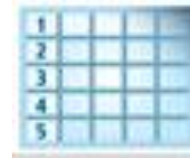
Lists, codepages

Models

European Commission
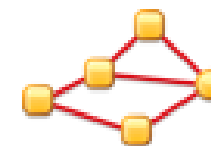
Digital economy is the future

- Digital Europe

Digital Economy

Digital skills
Supporting industry
Connectivity
Online platforms and e-commerce

Shaping Europe's digital future

Reuse of publicly funded data — Powering the economy and innovation

Ho

BF

E

Public sector bodies in the European Union, such as government agencies, local authorities, or statistical offices, produce and collect huge quantities of data. Examples of public sector data include:

- geospatial information;
- statistics;
- weather data;
- digitised books from libraries.

In
th

A

innovation

Allowing public sector data to be easily re-used for other purposes, including commercial ones, can stimulate economic growth and help address societal challenges. Public sector is a valuable resource for SMEs and startups to create new products and services.

https://digital-strategy.ec.europa.eu/en/library/building-data-economy-brochure

European Commission

What is eArchiving Initiative?

eArchiving key information

- *Digital Europe Programme*

- *DG CNECT and E-ARK Consortium*

- *Start date 1st October 2022*

- *Two years + two possible annual extensions*

https://digital-strategy.ec.europa.eu/en/activities/earchiving

eArchiving
helps people preserve and reuse information over the long-term

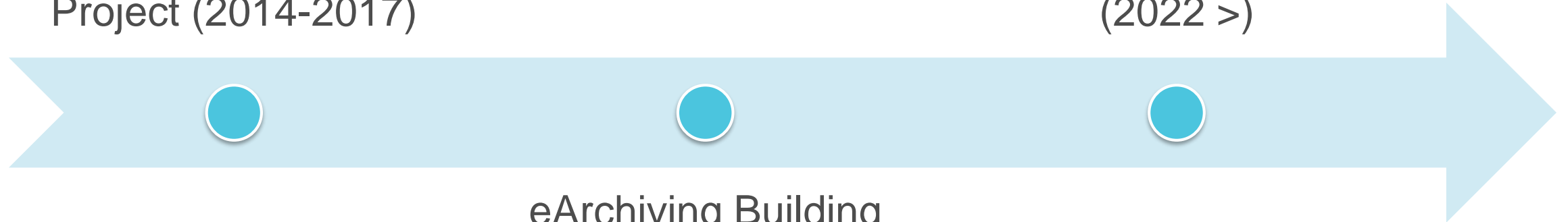- **Interoperability by default**

- **Openness and transparency**

- **Sustainability and legal compliance**

European Commission

Origins of eArchiving

E-ARK Research Project (2014-2017)

eArchiving Initiative (2022 >)

eArchiving Building block (2018-2019; 2019-2021)
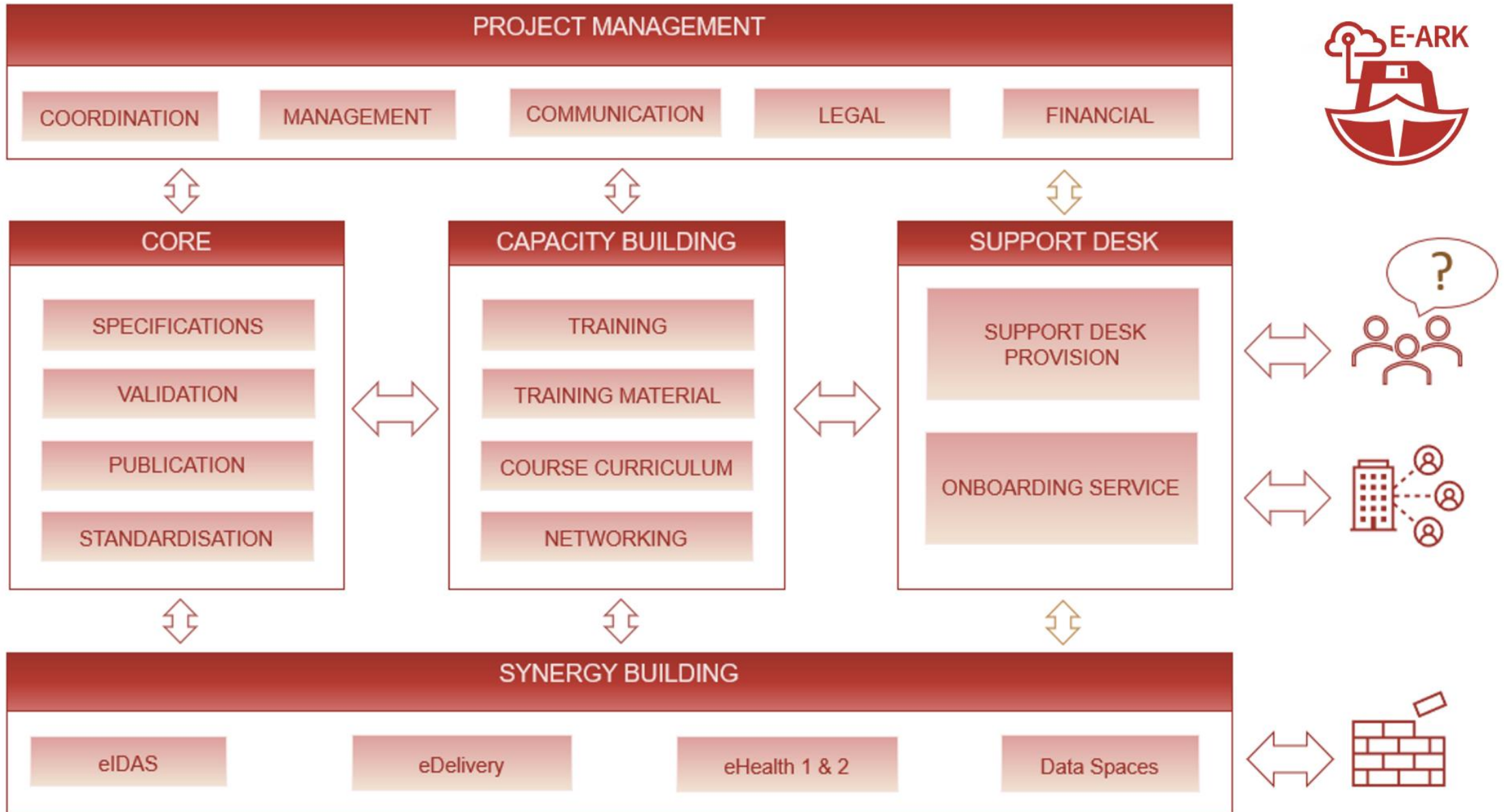
European Commission

eArchiving mission

- provide non-hosted archival services to keep digital records **authentic** and **usable** based on current **best practices**.

- The services address the three main endeavours of an archive:
  - **acquiring**,
  - **preserving** and
  - **enabling re-use** of information.

- Archival processes at a pan-European level are **harmonized**, supported by guidelines and recommended practices

European Commission

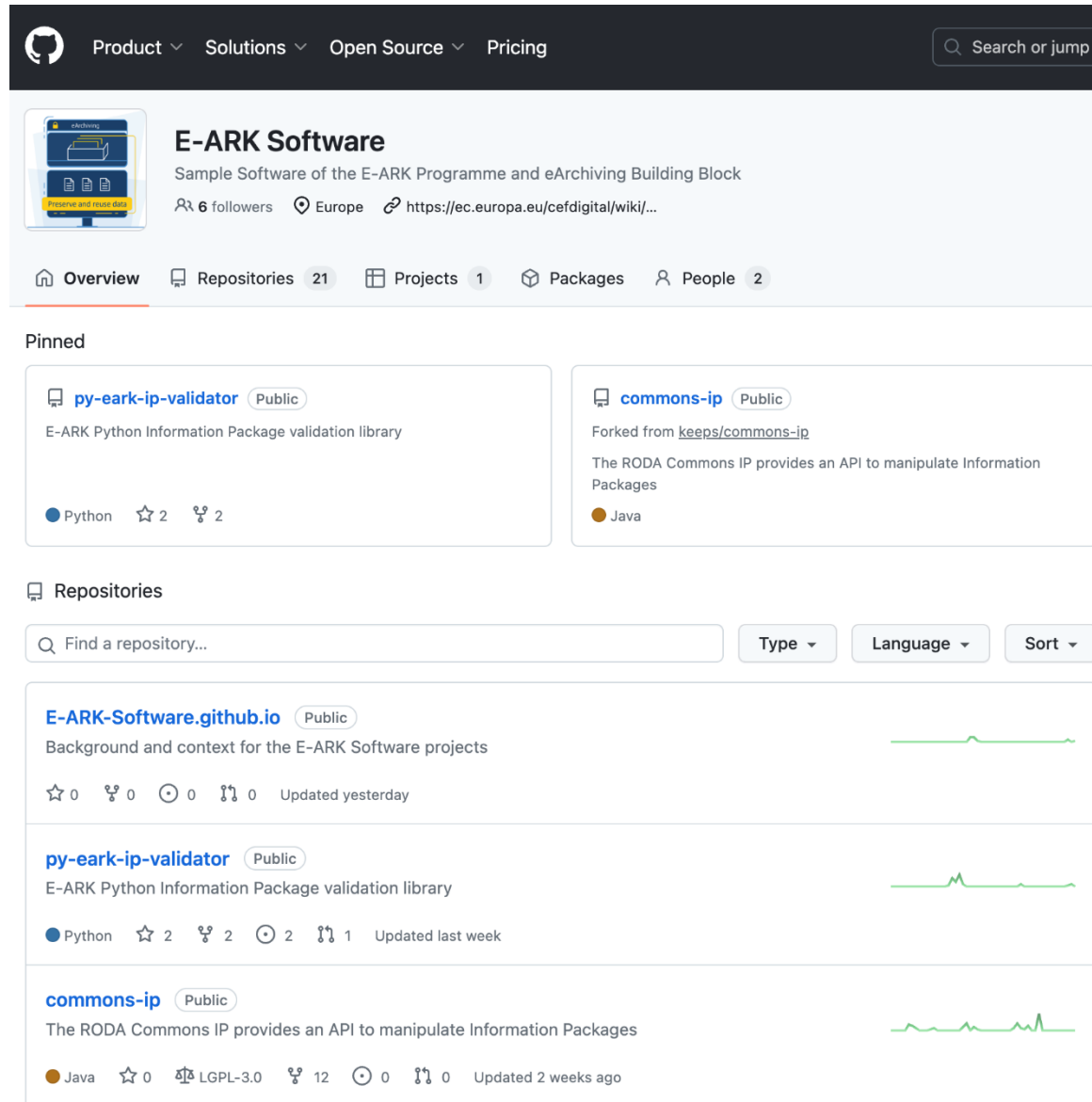# What you can use from eArchiving?

- Specification and validation for packaging and documenting

- Open-source tools for digital preservation management

- Resources for training, support and development

European Commission

The eArchiving Initiative: the five Activities

PROJECT MANAGEMENT

COORDINATION · MANAGEMENT · COMMUNICATION · LEGAL · FINANCIAL

CORE
- SPECIFICATIONS
- VALIDATION
- PUBLICATION
- STANDARDISATION

CAPACITY BUILDING
- TRAINING
- TRAINING MATERIAL
- COURSE CURRICULUM
- NETWORKING

SUPPORT DESK
- SUPPORT DESK PROVISION
- ONBOARDING SERVICE

SYNERGY BUILDING

eIDAS · eDelivery · eHealth 1 & 2 · Data Spaces

E-ARK Software (not a hosting Service)

https://github.com/E-ARK-Software

Who is using eArchiving so far

USERS:

- Denmark (DNA)

- Sweden (Package structure)

- Czechia

- Slovenia (included in legislation)

- **Croatia (eKultura Project)**

- Switzerland (SIARD)

- Nederland

- …

ADOPTING:

- Hungary

- Spain

- Portugal

- …

European Commission

How do you use eArchiving for geospatial records?

Archival use cases for preserving Geodata ?

- Backup for the Data producer

- Technical documentation of objects and events

- Legal disputes

- Research

- Future unknown uses

What do users want?

- Geodata for further analysis

- Outputs from GIS Systems (Information Products)

- Maps

- Analysis results with possibility of repeating them

- Source data

DATA

SORTED

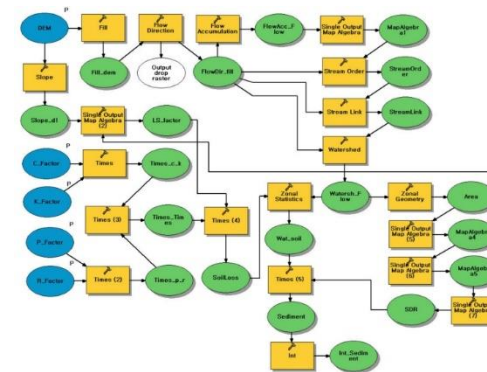ARRANGED

PRESENTED VISUALLY
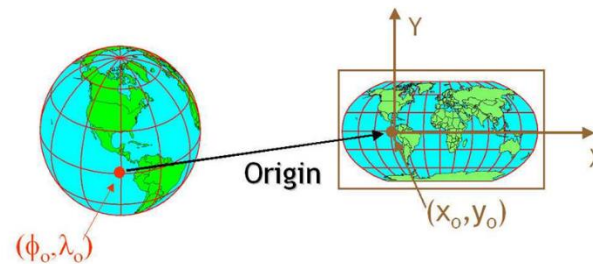
EXPLAINED WITH A STORY

@clearwaylaw

European Commission

What should we preserve?
The Significant Properties Model

- **Content:** conveys information, not necessarily human readable

- **Context:** background information on technical and business environments to which the digital objects relate

- **Rendering:** how the content of the object appears or is recreate

- **Structure:** component parts of the object and how they relate to each other
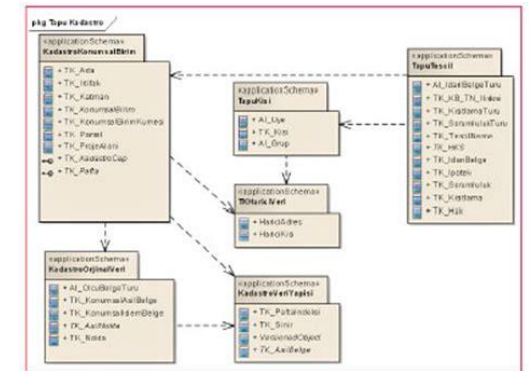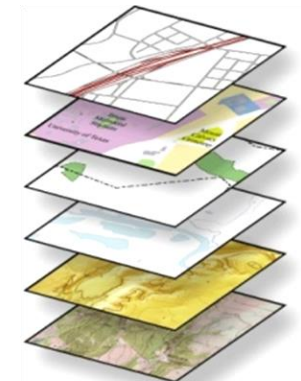
- **Behaviour:** functionality that is intrinsic to an object

Figure 3: Data groups (packages) of the application schema
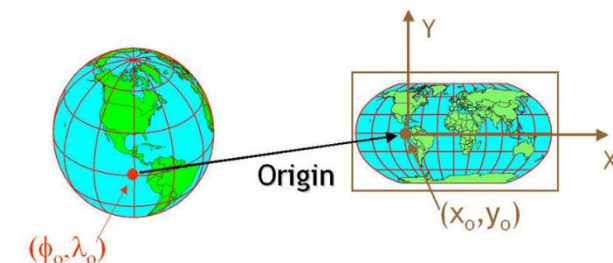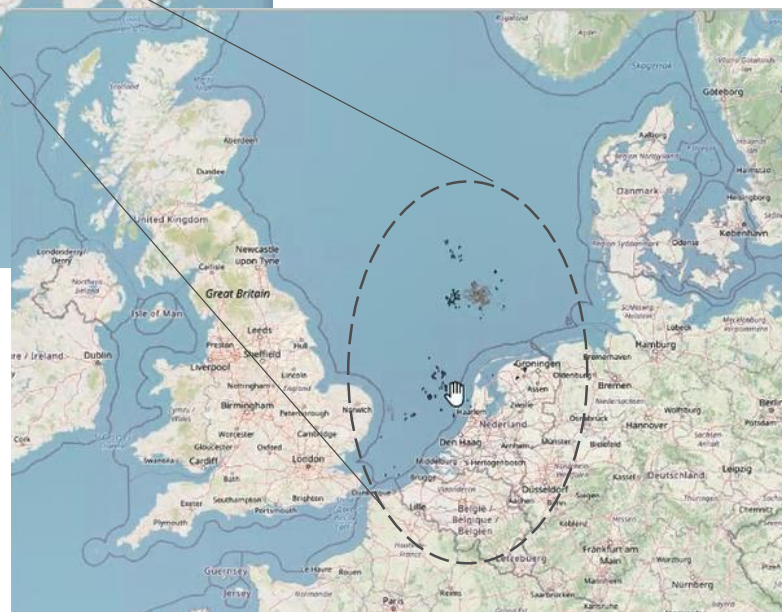
Context example

- Positional accuracy of cadaster parcels:
  20m

- Positional accuracy of orthophoto image (2006) in this area:
  6m

Knowing data limitations helps us interpret it

European Commission

Example: Rendering – Coordinate systems
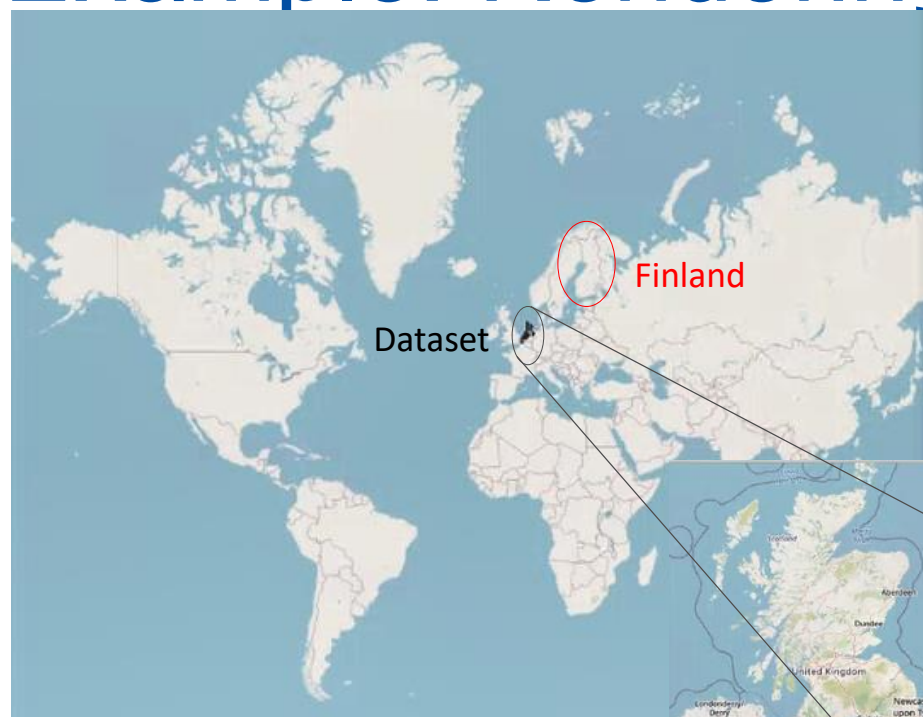
Dataset:
EPSG:3067
ETRS89 / TM35FIN(E,N)
Finland

Background map:
EPSG:3857
WGS84
OpenStreetMap
OSM Standard

European
Commission

Example: Structure
data coming from a complex system

Ministry of culture
Cultural heritage

Ministry of
defense

Environmental
agency
Environmental
restrictions

**Building
permit**

Ministry of
environment
Urban plans

Surveying &
Mapping
Authority
Cadaster maps

Ministry of
Agriculture
Land use

Example: Behavior

European Commission
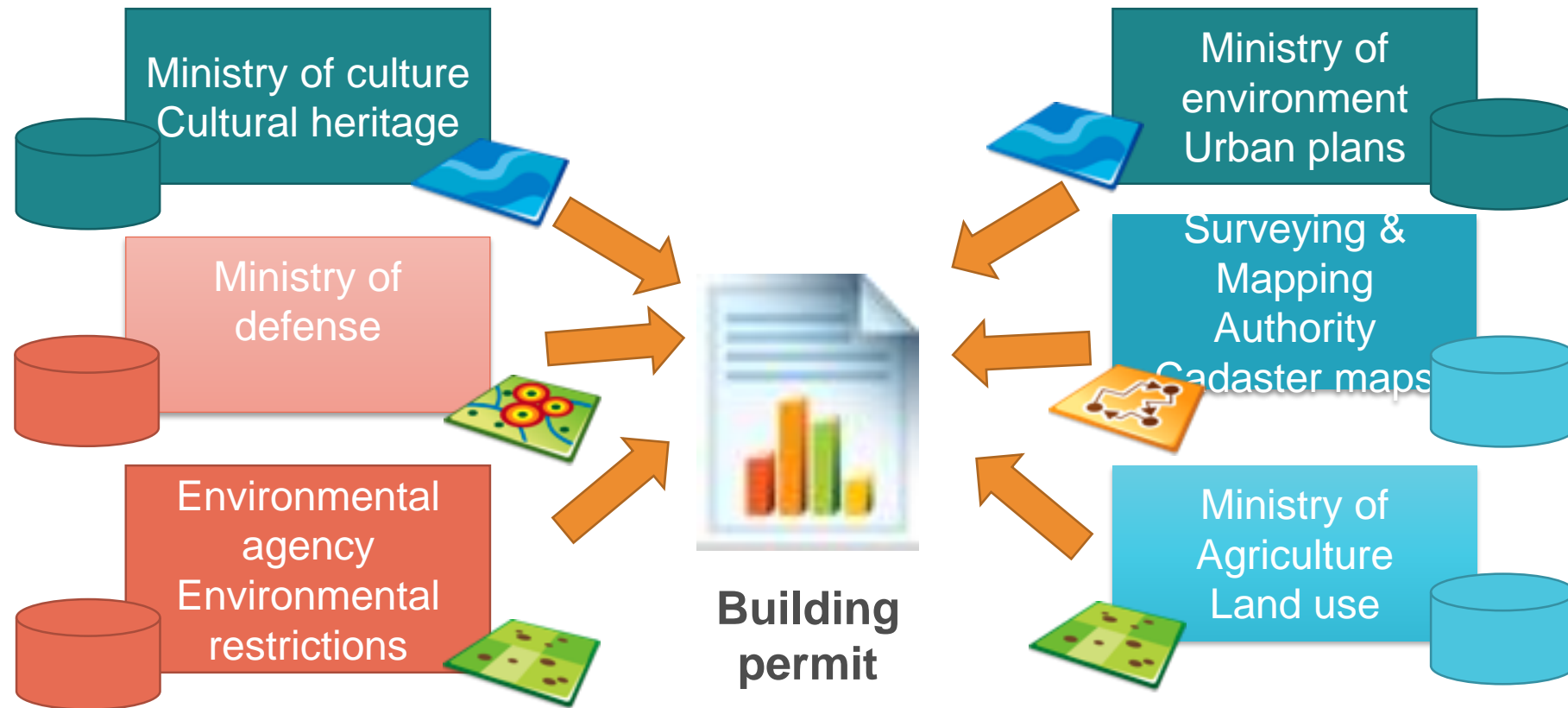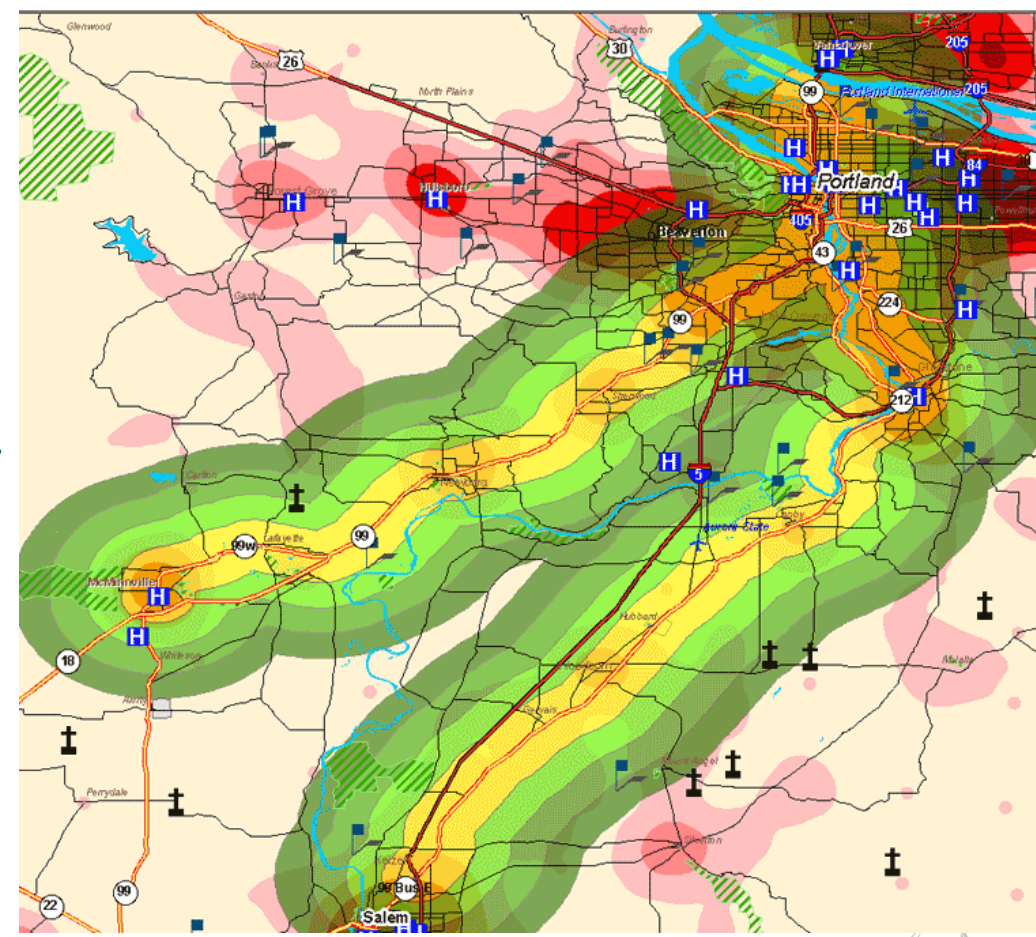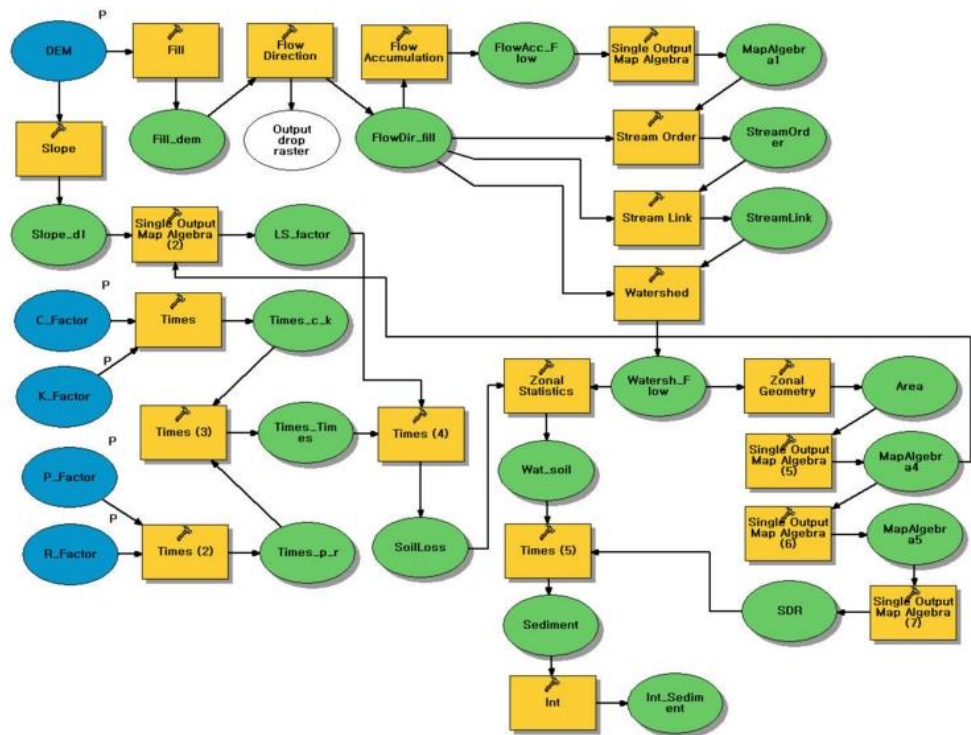
Solving the digital preservation issues

- Self descriptive packages

- Based on archival standards and standardized metadata

- Documented to preserve knowledge

CITS Geospatial

Generic

Common IP specification

General SIP, AIP, DIP specification

Content Information Type Specification

Business-specific specification

Local implementation

Detailed and specific

PROJECT MANAGEMENT				
COORDINATION	MANAGEMENT	COMMUNICATION	LEGAL	FINANCIAL

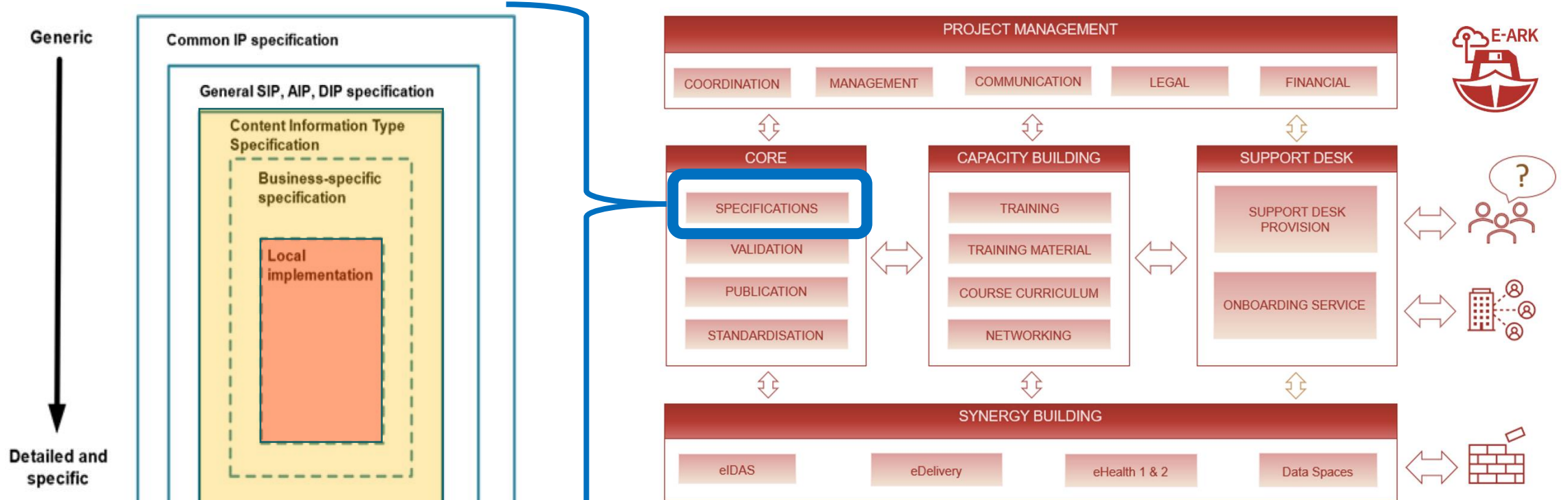| CORE	CAPACITY BUILDING	SUPPORT DESK
SPECIFICATIONS	TRAINING	SUPPORT DESK PROVISION
VALIDATION	TRAINING MATERIAL	
PUBLICATION	COURSE CURRICULUM	ONBOARDING SERVICE
STANDARDISATION	NETWORKING	

SYNERGY BUILDING			
eIDAS	eDelivery	eHealth 1 & 2	Data Spaces

E-ARK

Content
Information
Type
Specification

European Commission

CITS Geospatial Archival package
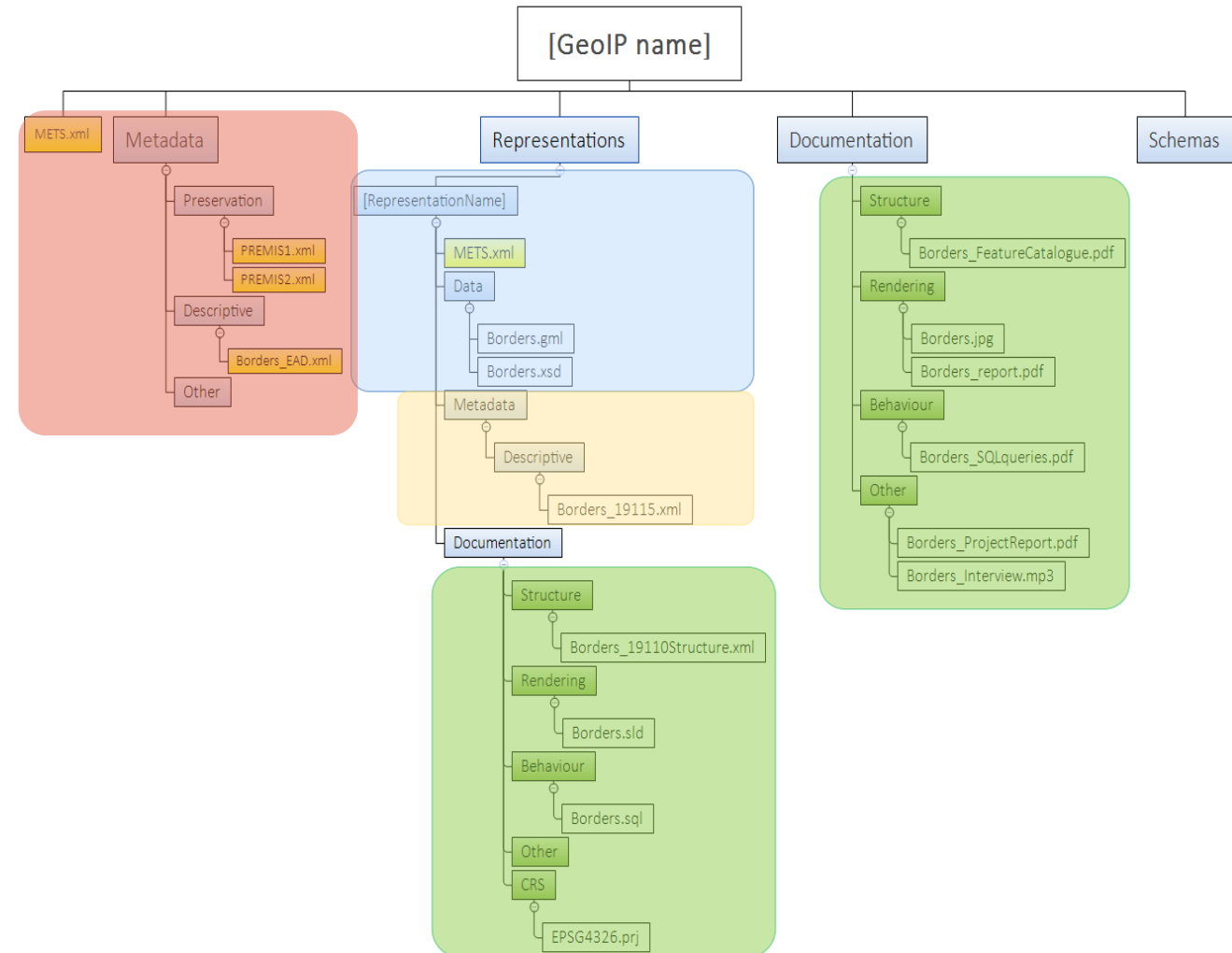
- Folder structure

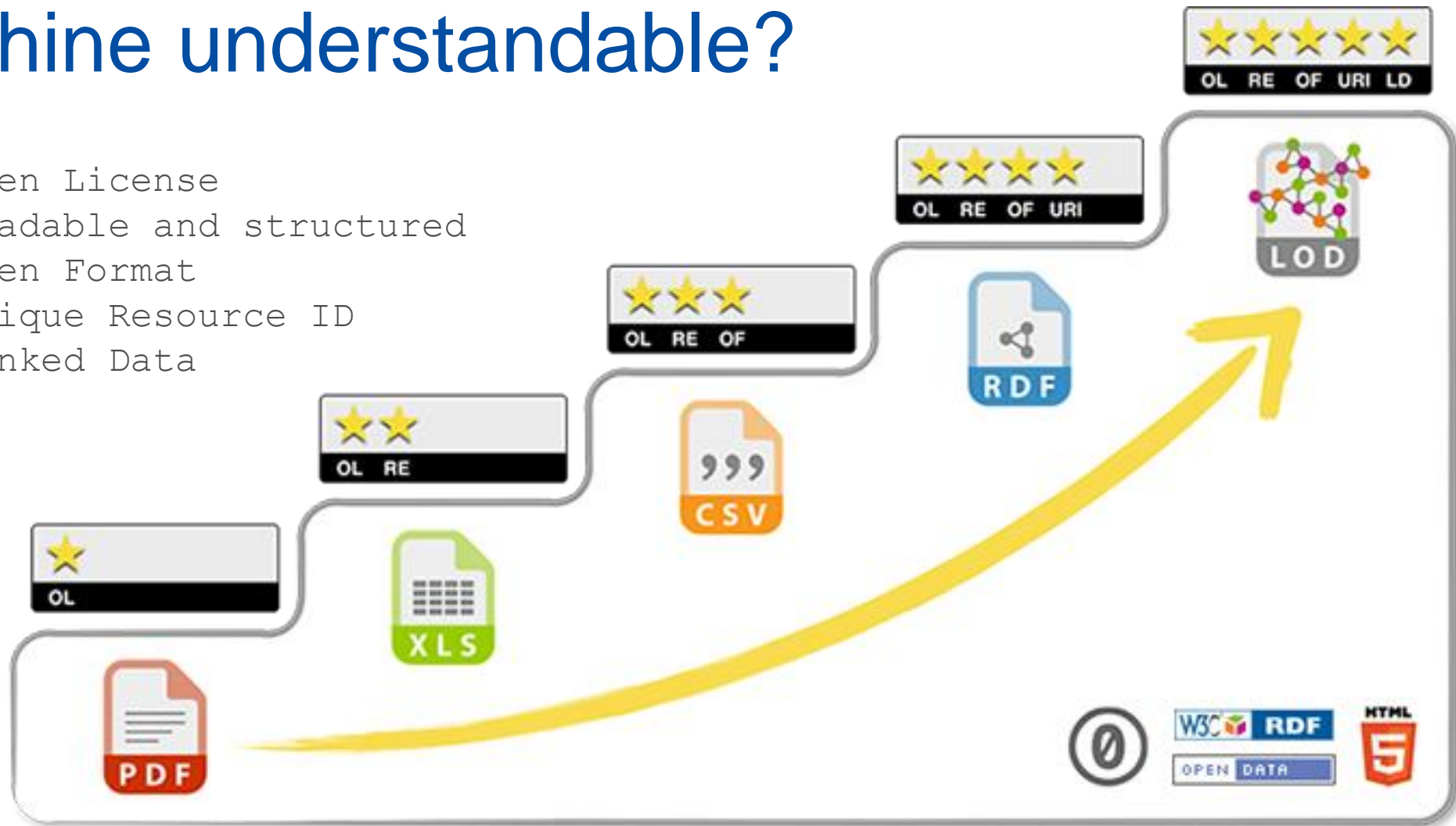 - Administrative metadata

- Geospatial Data requirements

- Documentation requirements

- Descriptive metadata (ISO 19115, INSPIRE…)

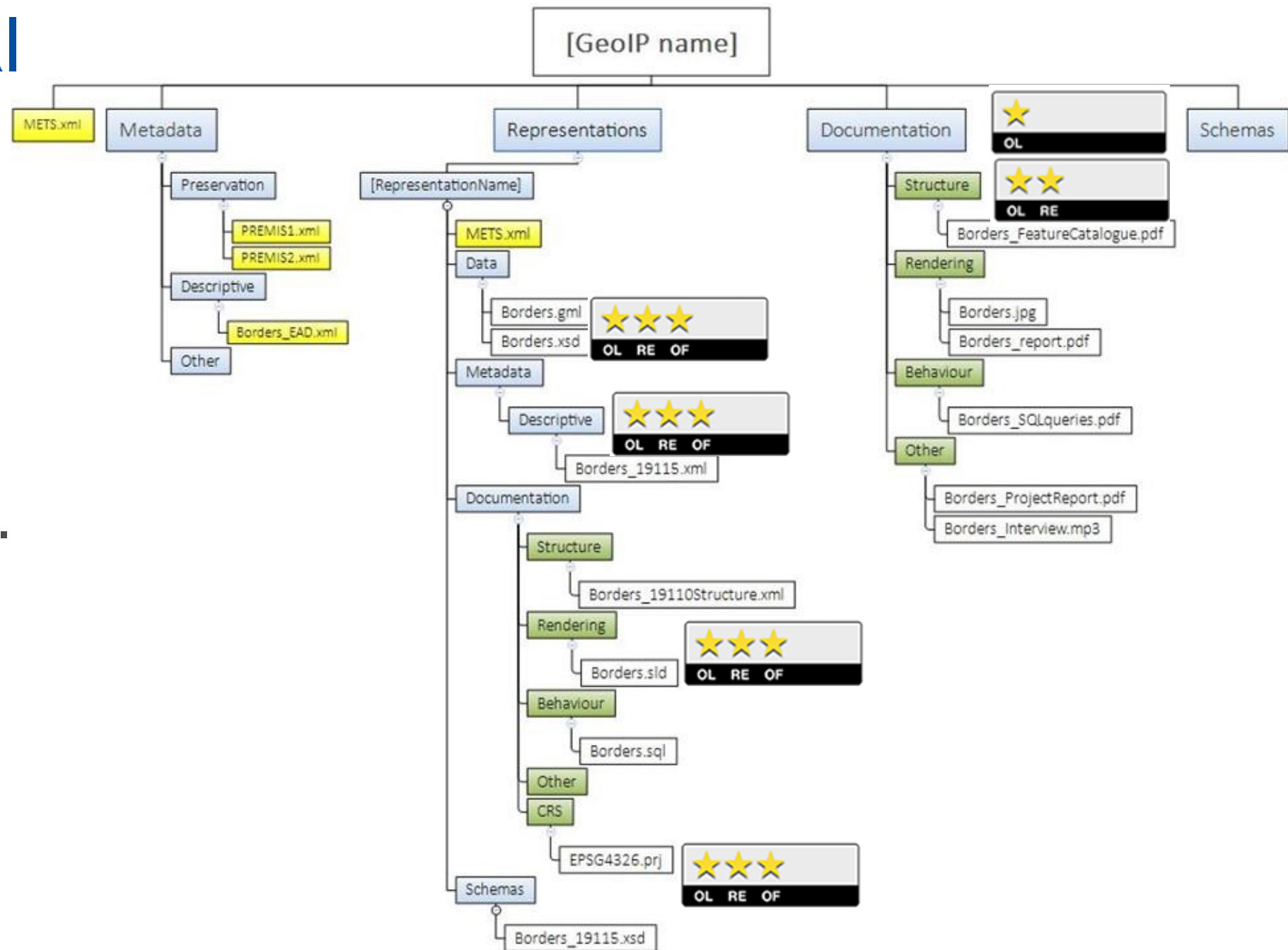# How do we make data more machine understandable?

OL - Open License
RE - Readable and structured
OF - Open Format
URI - Unique Resource ID
LD - Linked Data
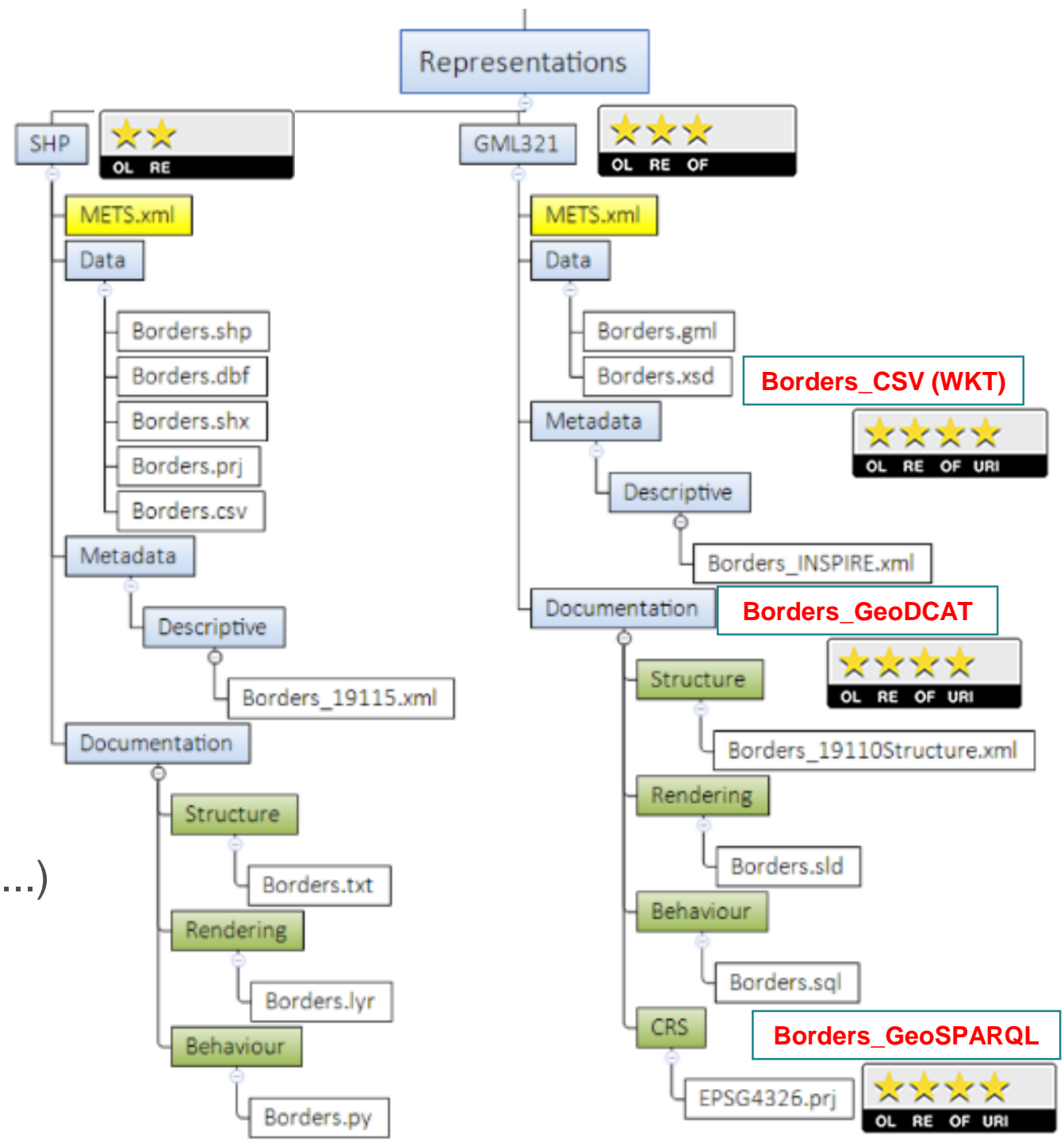
European Commission

Geospatial Archival Package structure

- Human readable vs.

- Machine readable

Storing multiple types of data representations

- Original representation vs open format representation

- Possible multiple types of metadata
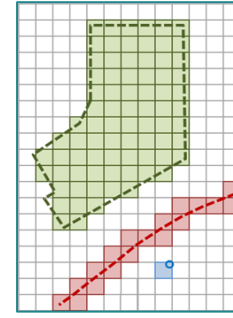 - Standardised (INSPIRE, ISO 19115...)

- Storing data for AI, linked data….

Data Validation Requirements

- **General data requirements**
  - **Vector requirements**
  - **Raster requirements**

- **Long term Preservation format profiles**
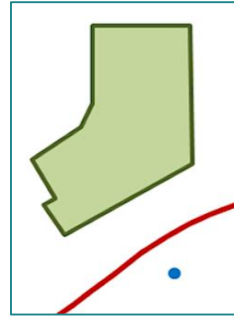
Vector

Raster

| ID	Name, Location & Description	Card & Level
GEO_11	Minimum one file in a geospatial format If the value in mets/@csip: CONTENTINFORMATIONTYPE is "citsgeospatial_v3_0 ", then there **SHOULD** exist at least one file in a geospatial format in representations/[RepresentationName]/data	0..n SHOULD
GEO_12	Subfolders in data representations/[RepresentationName]/data If there are more geospatial records in a representation, each geospatial file **MAY** be placed or grouped in subfolders in representations/[RepresentationName]/data	0..n MAY
GEO_13	Long term preservation format representation The Information Package **SHOULD** contain at least one representation of geospatial record in a long-term preservation format, as defined by the Archive or in the Long-term Preservation Format Profile (See chapter 3.3.5.)	0..n SHOULD

???? GML, SHP, KML, TIFF, JPEG2000, .gpkg, csv .... ????

European Commission

Guidelines for CITS Geospatial

- Introduction to Geospatial records

- Introduction to "Significant Properties" concepts for Geodata

- Rationalizations of all CITS Requirements

- Examples for many requirements
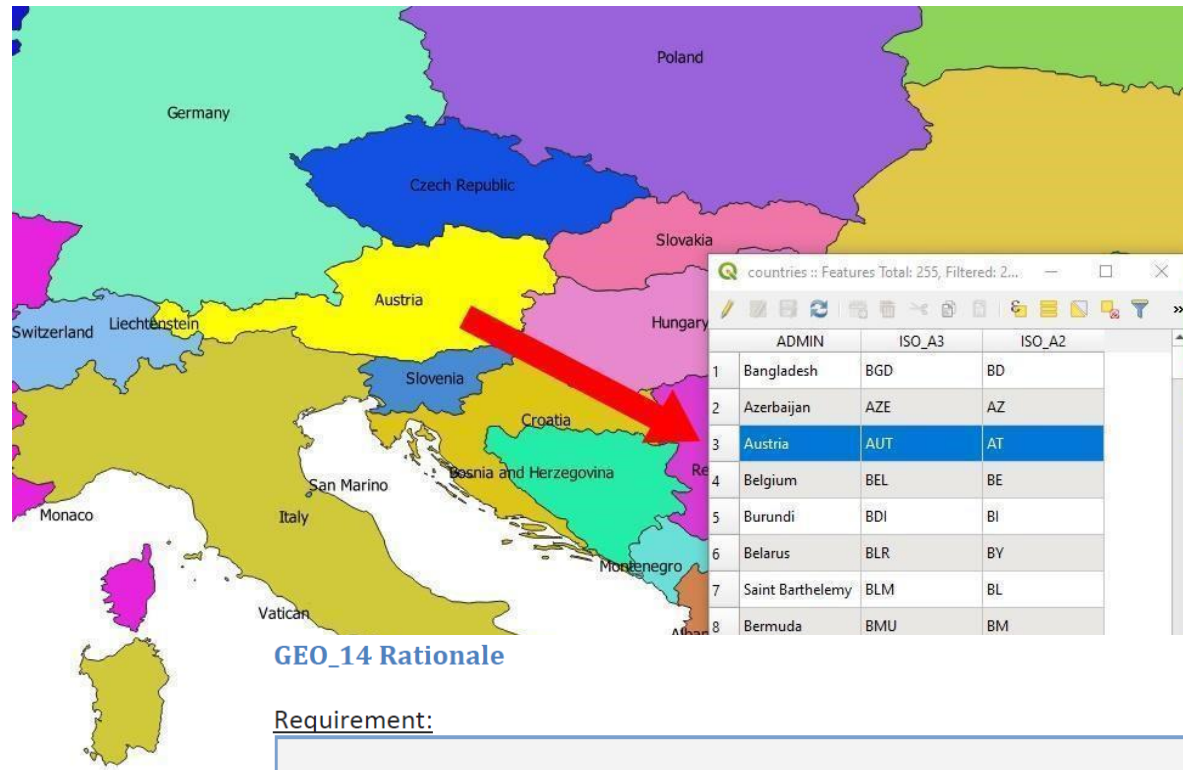
GEO_14 Rationale

Requirement:

GEO_14	Original format representation	The Information Package **MAY** contain a separate representation of the same data, containing geospatial data in its original format	0..1 MAY

Description:
This requirement allows an additional representation in the IP with the geospatial data in the original format.

Example:
Figure 2 in chapter 3.1.2 shows an IP with two representations. One representation contains a Long-Term Preservation vector data format (GML321), and the other contains a representation of the original format in an ESRI shapefile format (SHP).

Rationale:
Original formats are often richer and easier to use than the preservation format and suitable for dissemination in the short term. However, it does not ensure the long-term preservation of the data. Geospatial data in original format can also be used for validation on submission mitigating loss of data and significant properties during migration to preservation format. The idea is that the users could use this representation until the original formats becomes obsolete.

Guidelines for CITS Geospatial with GIS

- **Introduction to GIS Systems**

- **GIS Preservation strategies**

- **Rationalizations of all CITS Requirements**

- **With Examples based on standards**
  - OWS Context
  - WKT2 for CRS
  - SLD
  - Etc.

GEO_33 Rationale

Requirement:

GEO_33	Rendering configuration	A standardised machine-readable rendering configuration for one or more geospatial datasets **MAY** be provided in the Information Package	0..n MAY
GEO_33a Ref GEO_33	Placement of rendering configuration	If a standardised machine-readable rendering configuration for one or more geospatial datasets exists, it **SHOULD** be provided in representations/[RepresentationName]/documentation/rendering	0..n SHOULD

Description:
This requirement recommends that rendering configurations are documented in a standardised machine-readable format to support dissemination automatisation.
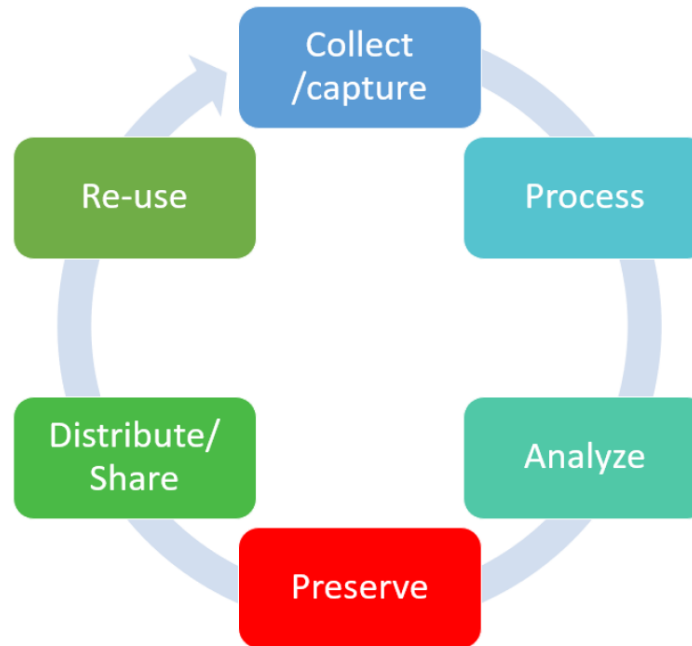
Example:
An example of Standardised machine-readable formats for the rendering of geospatial records are SLD[20] files. KML[21] files also have some of that capability:
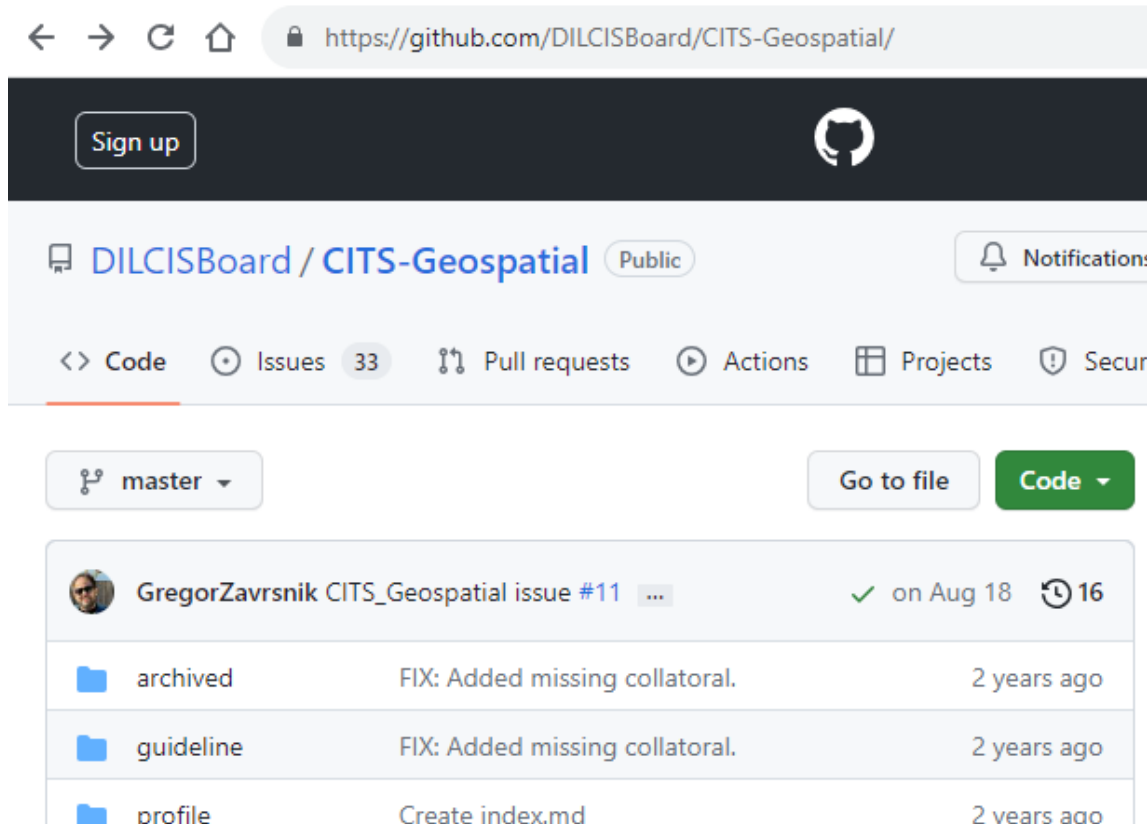
SLD files example
SLD is an OGC[22] (Open Geospatial Consortium) standard for symbology and is the OGC Styled Layer Description XML format (SLD files). If the producer cannot provide the archive with SLD files, these can be recreated from the description provided in the Documentation in an open-source GIS application like QGIS[23]. Raster files can have a colour map associated with the pixel value. The SLD standard is used for rendering geodata in OGC web services and, therefore, could be used as an appropriate input for an easier DIP creation in the future. An example of an SLD file is shown in figure 13.

```
<StyledLayerDescriptor xmlns="http://www.opengis.net/sld"
      xmlns:ogc="http://www.opengis.net/ogc"
      xmlns:xlink="http://www.w3.org/1999/xlink"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      version="1.0.0"
      xsi:schemaLocation="http://www.opengis.net/sld StyledLayerDescriptor.xsd">
  <NamedLayer>
      <Name>Simple Point</Name>
      <UserStyle>
          <Title>SLD Cook Book: Simple Point</Title>
          <FeatureTypeStyle>
```
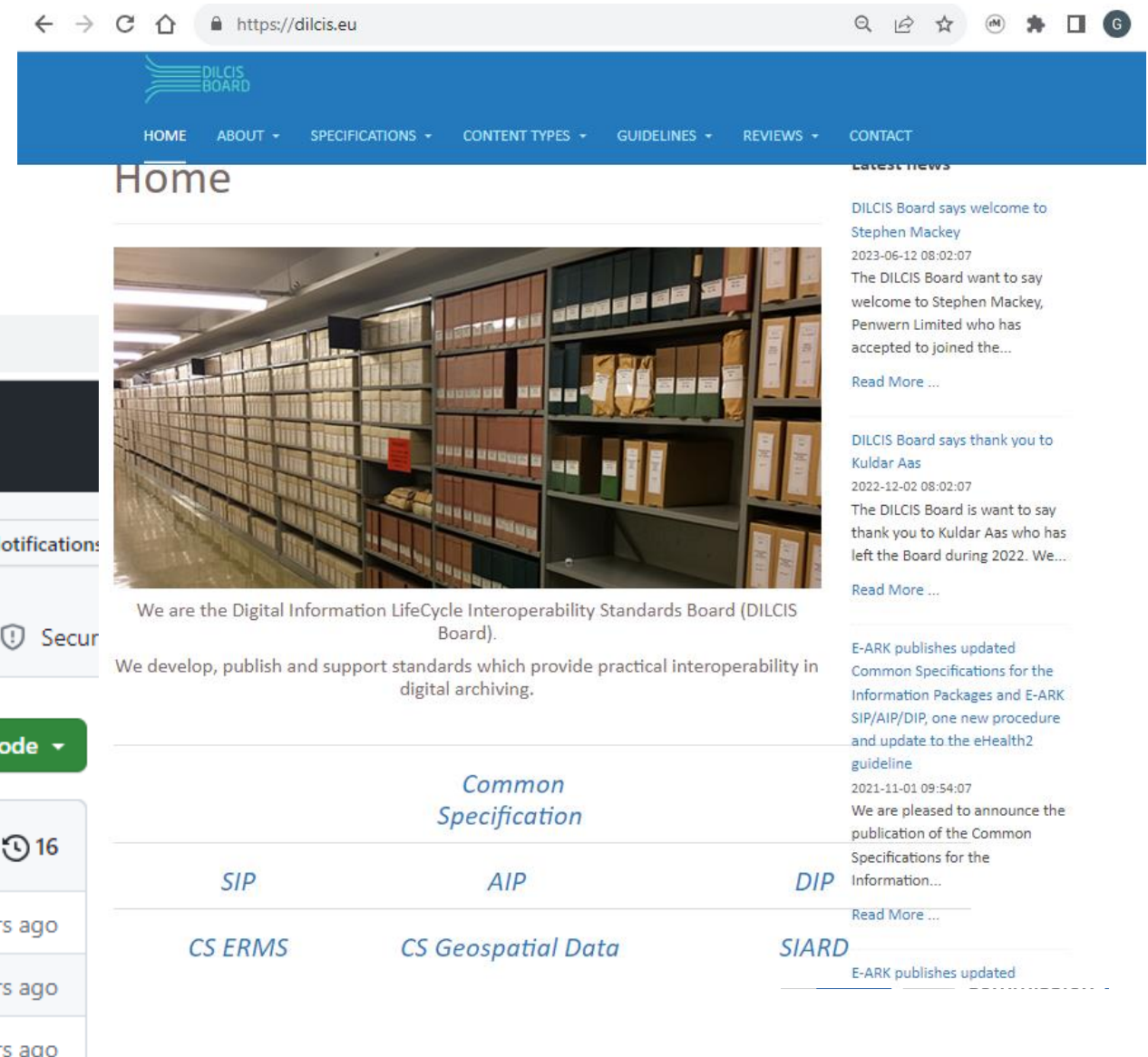
Where you can find CITS Geospatial?

- www.dilcis.eu

- https://github.com/DILCISBoard/CITS-Geospatial/

What is next?

How you can participate?

- Do you have your own archiving guidelines?

- Do you have solutions for archiving and reuse, that you can share?

- Do you have questions you need answering?

- Do you need partners for EU funded projects?

Core benefits of adopting eArchiving

Extended possibilities for collaboration in tool development

Shared development = less cost for each individual partner

Digital preservation is affordable for everyone!

European Commission

Go here first! support@e-ark-foundation.eu

The eArchiving Support desk

The [eArchiving support desk](#) is the primary contact point for institutions and organisations to whom we provide support regarding questions and requests related to theeArchiving specifications and components.

We offer the following services:

Support desk provision

The support desk is set up to address issues, questions, and feedback that organisations and institutions might have related with any eArchiving service.

Onboarding service for new organisations

This service aims to assist and give guidance to organisations and institutions that want to implement or receive training in the eArchiving specifications and components.

You can contact us via email (support@e-ark-foundation.eu) for your questions, comments about eArchiving and advice on implementing, integrating and testing in your solution!

Check the [privacy statement](#) before sending your questions and comments.

European Commission

Join us at the DLM Geoforum (May 28th)

Workshop: Preservation of Digital Geospatial Records, the Cornerstone of Europe's Digital Decade

📅 28 May 2024

📍 National Archives of Belgium

The eArchiving Initiative and DLM Geoforum announce the "Preservation of Digital Geospatial Records, the Cornerstone of Europe's Digital Decade" workshop on Tuesday May 28 2024 in Brussels.

The European Union's digital strategy views the reuse of public sector geospatial information as a driver of economic growth and a key resource to address societal changes. The public sector is a valuable resource for SMEs and startups to create new products and services.

There are many challenges with digital transformation, and more ahead, with the increasing adoption of AI. What, where and how do we manage and preserve, in this digital revolution? With geospatial records being so complex and ever-changing, management, reuse and especially preservation continue to be a challenge. That is why sharing knowledge, experiences and ideas with different stakeholders in this domain is crucial. And the DLM Geoforum is where it can happen, so we invite managers of geospatial records, archivists, solution providers, researchers and policymakers to meet us in Brussels in May.

The Geoforum aims to be as interactive, diverse, and inclusive as possible, so the program of the conference will include different types of contributions.

The hybrid Geoforum workshop will bring together managers of geospatial records with data producers, archivists, solution providers, researchers and policy makers, creating a transnational

created with DALL-E AI tool
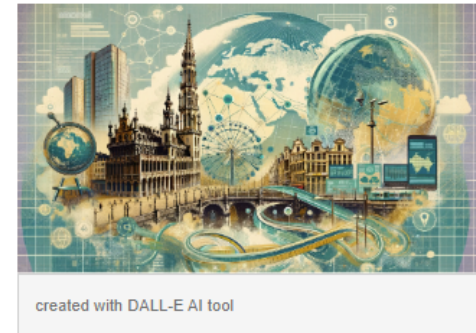
Organiser

European Commision, DG Connect - eArchiving Initiative and DLM Forum

Register here to attend in presence

See also

Register here to attend virtually

Contact

eArchiving Initiative support desk

Related topics

European Commission

Questions?

support@e-ark-foundation.eu

gregor@geoarh.si