Meet eArchiving

Learn how to preserve, migrate, reuse and trust your data

4 December 2019

Connecting Europe Facility

DIGIT

DG Connect

Directorate-General for Informatics

Directorate-General for Communications Networks, Content and Technology





Agenda Day 2

4 December	Activity	Presenters
09:30-09:35	Welcome	Anne Bajart, European Commission
09:35-10:45	eArchiving domain introduction Presentations • Geographical data • Taxation and customs domain Panel Discussion	Fulgenio Sanmartín, European Commission Chair: Adelina Cornelia Dinu, European Commission Rink Kruk, National Geographic Institute, Belgium Ulrich Kröner, European Commission Anja Paulic, Archives of the Republic of Slovenia
10:45-11:00	Coffee break	
11:00-12:30	Break-out session: What is the potential of eArchiving in your domain?	Chair: Fulgencio Sanmartín, European Commission
12:30-12:35	Conclusion and wrap-up	European Commission



Live Streaming

Meet eArchiving

Start 09:30 (CET)

Information for the Live Streaming

Please note that this webinar is recorded.

Send your questions to : cef-building-blocks@ec.europa.eu



Welcome

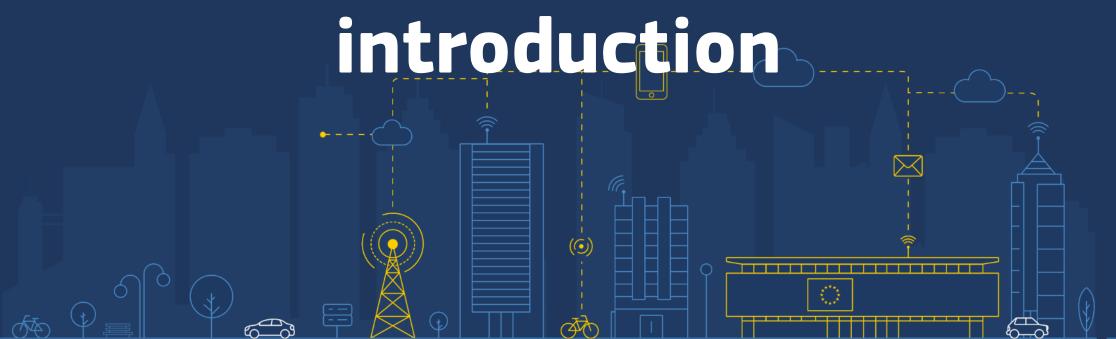


Anne Bajart

Deputy Head of Unit – Interactive Technologies, Digital for Culture and Education DG CNECT G2 – European Commission



CEF eArchiving domain



CEF eArchiving domain introduction



Fulgencio Sanmartín

Policy Officer –Interactive Technologies, Digital for Culture and Education

DG CNECT G2 - European Commission



Preserving your data in the long-term is challenging,

CEF eArchiving is here to **make it easier for you**.



Preserve

eArchiving allows you to preserve your data in the long-term.



Migrate

eArchiving's interoperable standards and open source software enable you to migrate your data safely, even across different technologies.



Reuse

Interoperable standards and specifications ensure data remains accessible and easy to reuse.



Trust

eArchiving ensures that your data remains authentic, secure and unaltered.



#eArchiving #ConnectingEurope



- Did you use a GPS when coming?
 Do you have it enabled now?
 Galileo satellites
- CEF eArchiving offers geopreservation specifications based on the EC Directive's INSPIRE
- Geo #eArchiving event last April 2019 in Ljubljana

#eArchiving #ConnectingEurope



- What is the biggest paper archive that you have? At home, at work?
- Why do you keep it and for how long?
- Proof of concept of CEF eArchiving by DG TAXUD

CEF eArchiving domain introduction **Geographical data**



Rink Kruk
National Geographic Institute, Belgium



Geodata as Instrument, Connector, and Archival Matter

Experiences from Belgian National Mapping Agency and State Archives

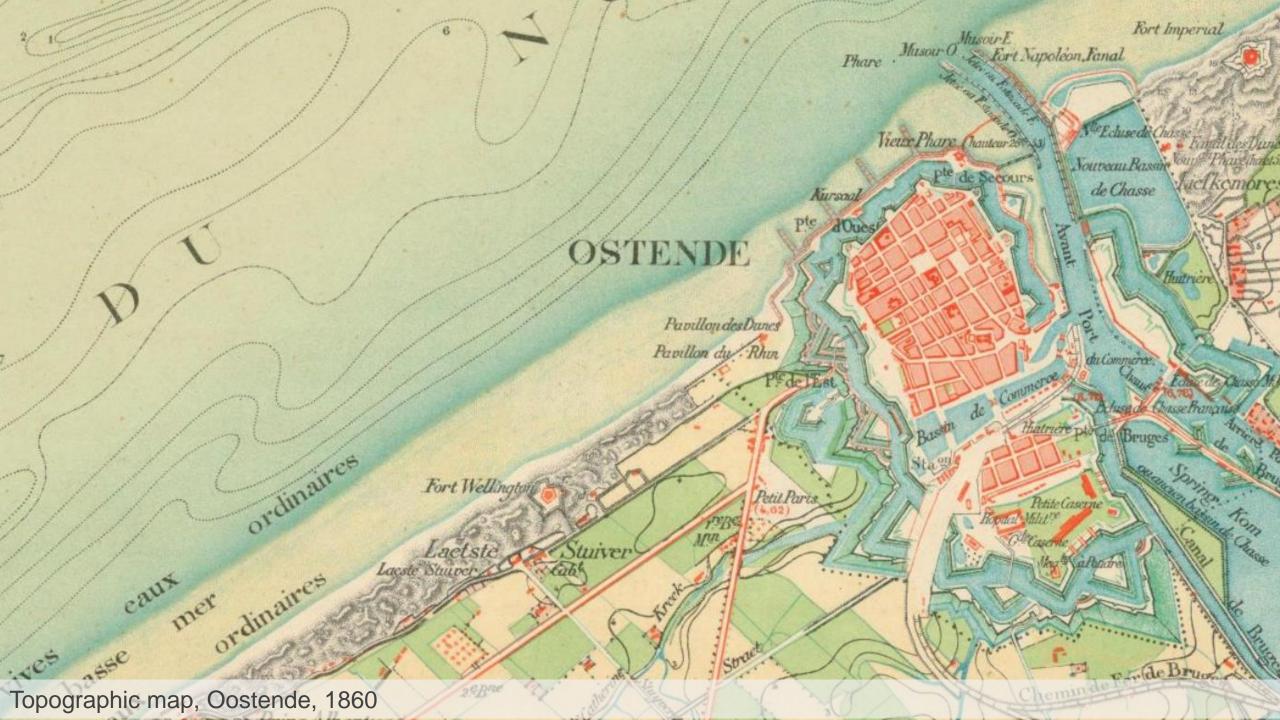
Rink W. Kruk

National Geographic Institute Belgium













Heizel (Brussels), 1947







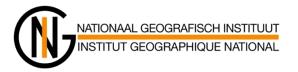






"Classical search" leads to confusion

Avoiding work, damage and confusion



Variations in the same language

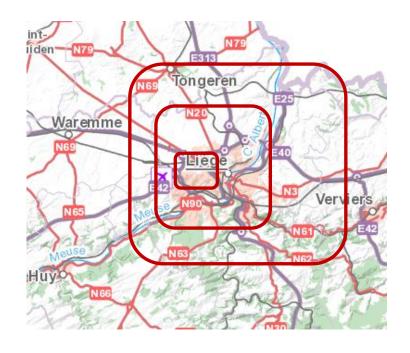
Brussel, Bruocsella, Brusella, Bruxelle,...

Name in multiple languages

Liège (FR), Luik (NL), Lüttich (DE), Lieja (ES), Liegi (ITA), Leodium (Latin), ...

Difference in meaning of the same place name

Liège: ville, agglomération, province, évêché, principauté



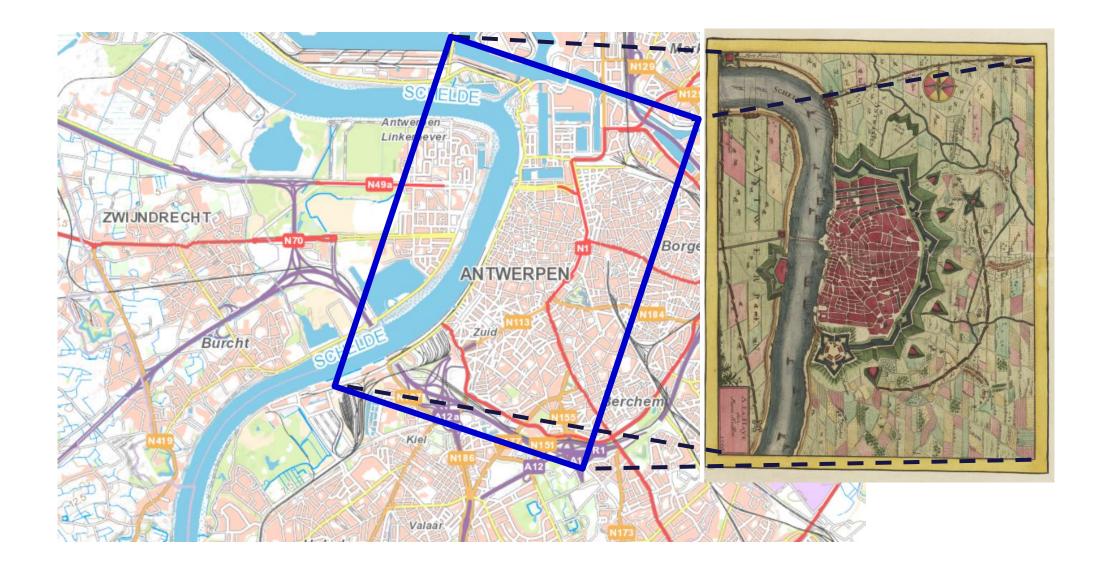
Different places with the same name

Bergen, Nazareth, Châteauneuf

Footprint is the answer to (almost) everything

Polygon, line or point in modern coordinates = area covered by an old map

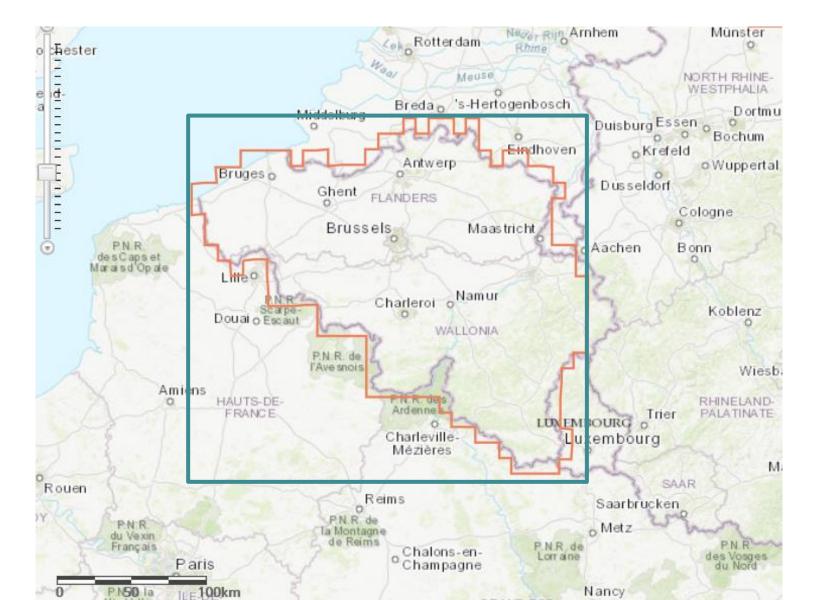




Polygon vs. bounding box

Precise vs. false matches



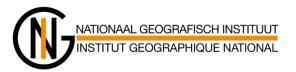


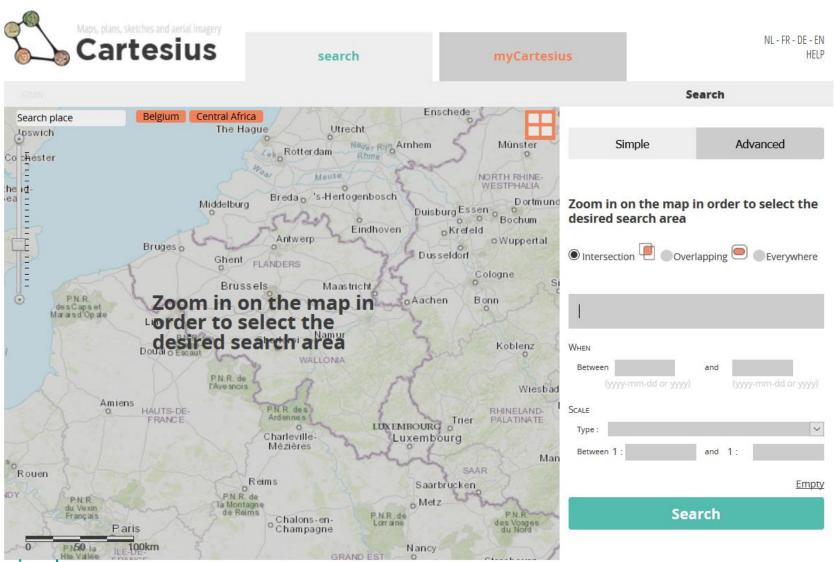
Polygon – Precise matches footprints often irregular ISO-19115-standard

Bounding box – Many false matches INSPIRE-standard

Searching by a modern map

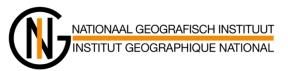
Polygon vs. textual search

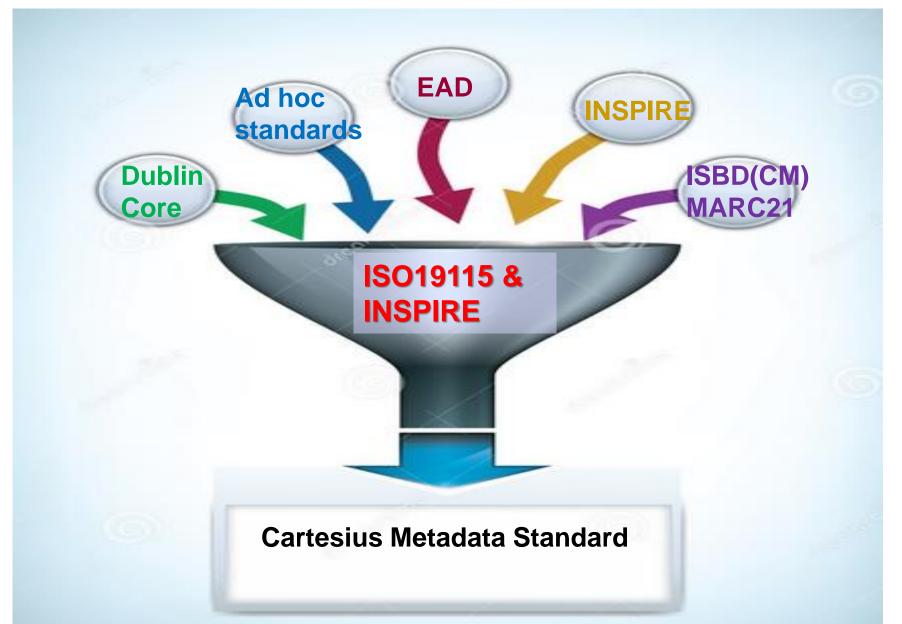




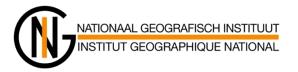
Cross breeding Metadata Standards

How to fix the mess





Evolution



Started in 2009 (in fact already before)

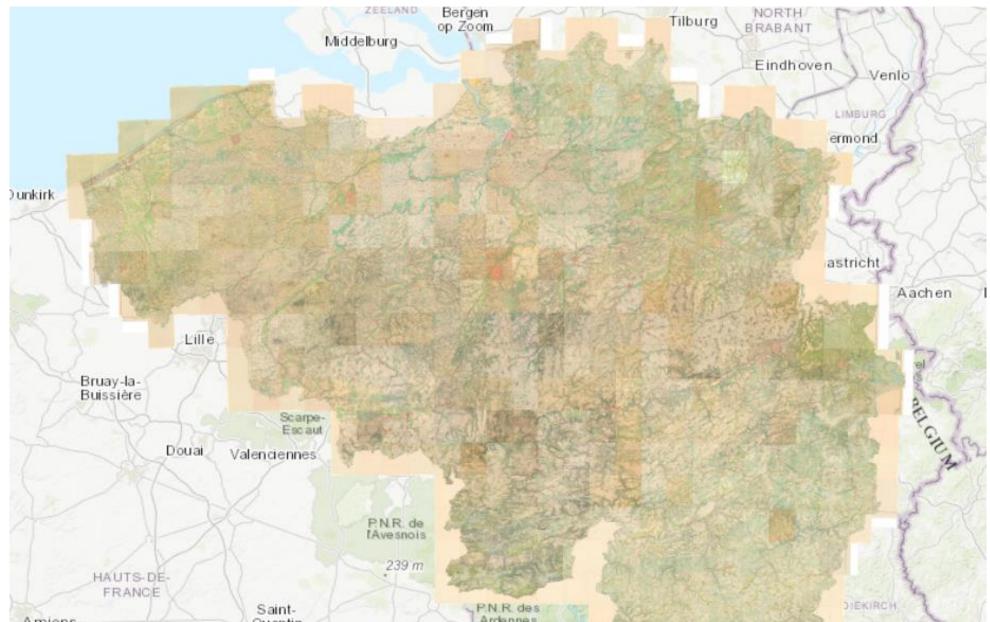
- Lessons learned (vendor lock-in, open source and proprietary have costs)
- Very popular, impressive quantity and quality
 - Most important map collections become available already by now (70.000 items)
 - ~10.000 visitors each month
 - Good collaboration between different types of institutions
 - A lot initiatives started, Cartesius.be keeps on going strong
 - Cartesius 3.0 analysis with public & scientific community
- First step: Open-up access!
- Second step: What can we do with it?



Evolution



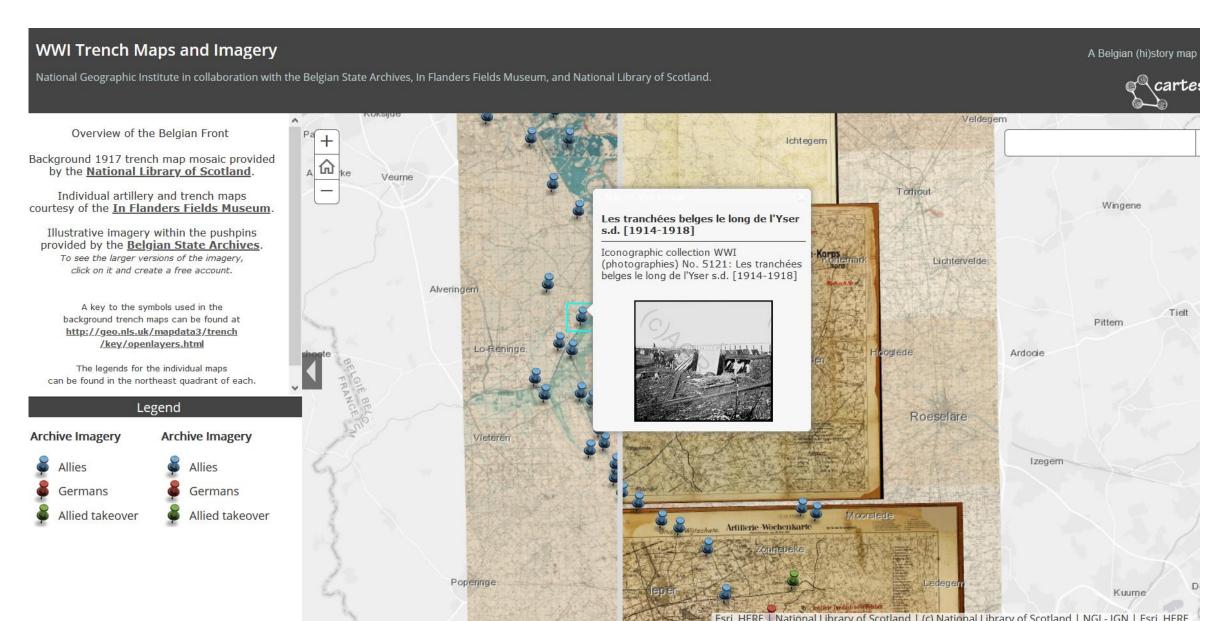
Stitching mapsheets together (of different years – into a country-wide coverage)



Evolution

Combining sources from all over Europe into one of many stories...





Vectorisation

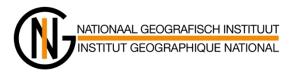






Historic gazetteers

e.g. from the time of the Austrian-Hungarian Empire



For Belgium ~ 33.000 entries

Gazetteer

Parser

• • •



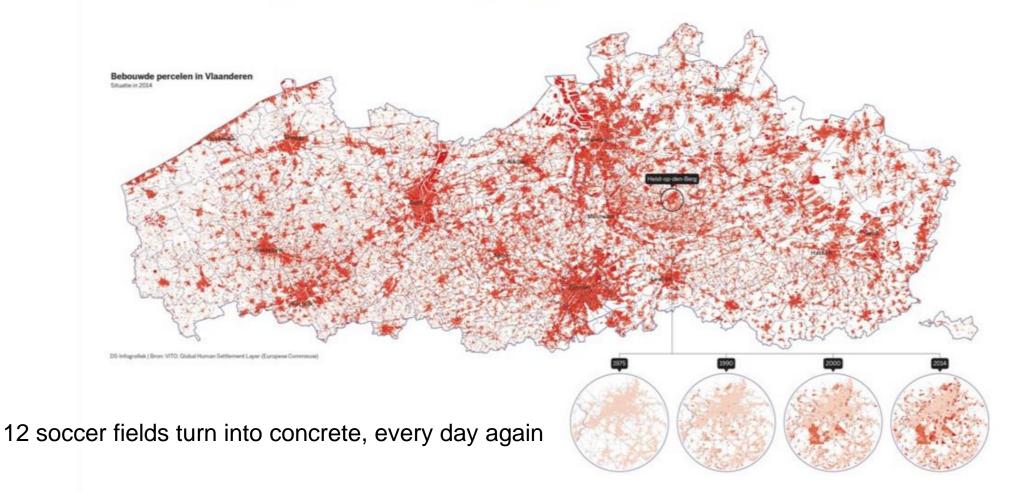
Research by journalists

NATIONAAL GEOGRAFISCH INSTITUUT
INSTITUT GEOGRAPHIQUE NATIONAL

Using AI to make the urbanization visible over the past decades – to impose against today's

Spatial Planning

12 voetbalvelden open ruimte weg, elke dag opnieuw



>100y Preservation of digital Geographic Information



16 fundamental principles agreed by National Mapping Agencies and State Archives

- 1. Archiving of digital Geographic Information begins at the point of data creation, rather than at the point of withdrawal from active systems
- 2. All organisations must have a maintained Archiving Policy
- 3. Be selective and decide what to archive and what to dispose of
- 4. Consider preservation timeframes of 1, 10, 100 years
- 5. Migration or emulation is inevitable in the medium and long term. Be prepared and choose which properties to preserve in advance
- 6. The output of the archival planning process should also be preserved over the long-term to accommodate future preservation requirements
- 7. Archiving is not back-up. You must also back-up your archive
- 8. Geographical data should be preserved in a way that non geo-specialists can handle
- 9. Information objects should be self-contained and independently understandable
- 10. Keep the gold copy version of the 100 year data archive in open, file based repositories, not in databases, nor other complex environments
- 11. Consider keeping a graphical representation alongside the logical representation of the data
- 12. Restrict the number of formats and encodings to a widely agreed set of open, simple and well documented file formats
- 13. Prefer simple data models and schemas over complex ones
- 14. Keep the access mechanism for archived data simple. Focus on basic current user requirements an archival viewing system does not need to be a fully functioning GIS
- 15. Ensure effective management and quality assurance of the metadata associated with your data
- 16. Make some assumptions about future use, but don't be too restrictive



http://www.eurosdr.net/research/project/eurosdr-archiving-working-group

Preservation of Geographical Production Process

European Spatial Data Research Working Group



EuroSDR-Working Group 2013-2017



Essential to know what the content on the map in fact means



















aquaterra















Preservation of the Geographical Production Process

Sexy brass instruments – Part of the Geographical Production Process





Archiving of Aerial Imagery

EuroSDR study 2019







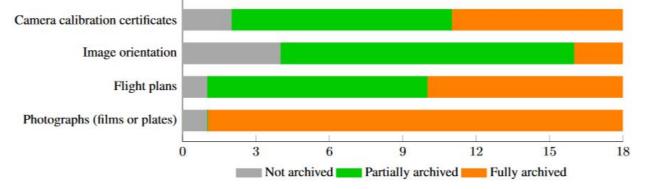
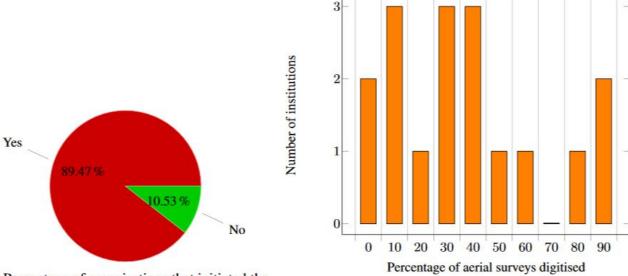


Figure 12: Data archived with the analogue surveys.

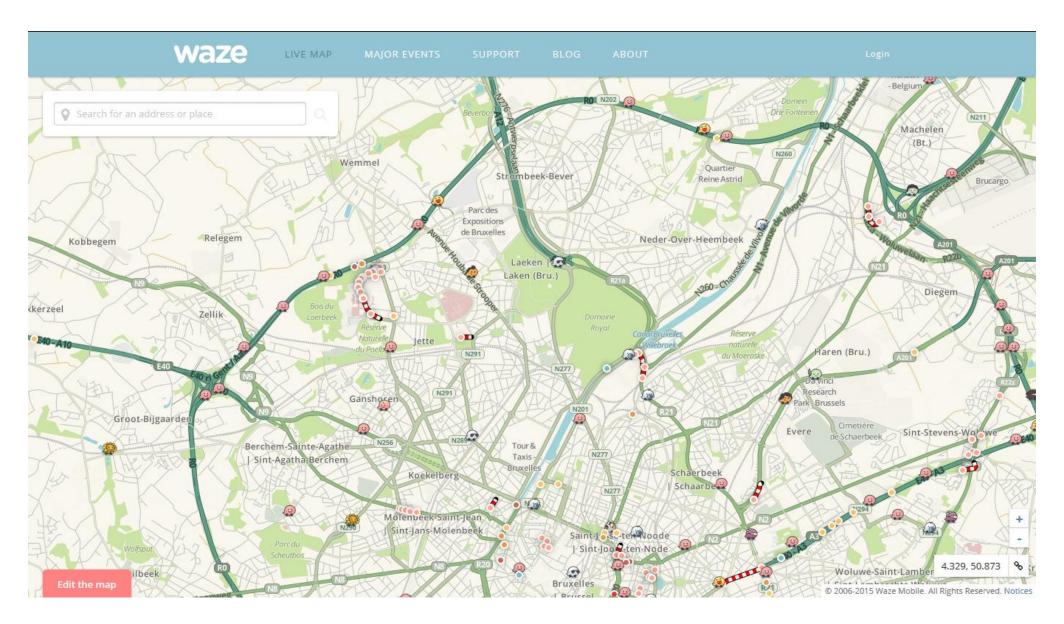


(a) Percentage of organisations that initiated the digitisation process.

(b) Progress of the process.

What should we archive?





State of Affairs...



- Although there are
- Good examples of implementing e-Archiving building blockss
- Intelligent initiatives to preserve the past and make and keep the past accessible
- Many willing employees at NMCAs and State Archives
- Archiving of (geo)data is not always happening in the way it should, because
- NMCAs have other priorities
- State Archives lack resources (in number and skills)
- NMCAs & State Archives do not know where to start as they lack training/awareness
- We see often
- Copies of digital maps (raster) and databases stored on a file server à la improvisio
- Archiving roles, responsibilities and archiving policies have not been defined

Nevertheless

There is a community of the willing that needs leverage



- How are we going to make our data accessible?
- How are we going to leverage our communities?

A shared responsability

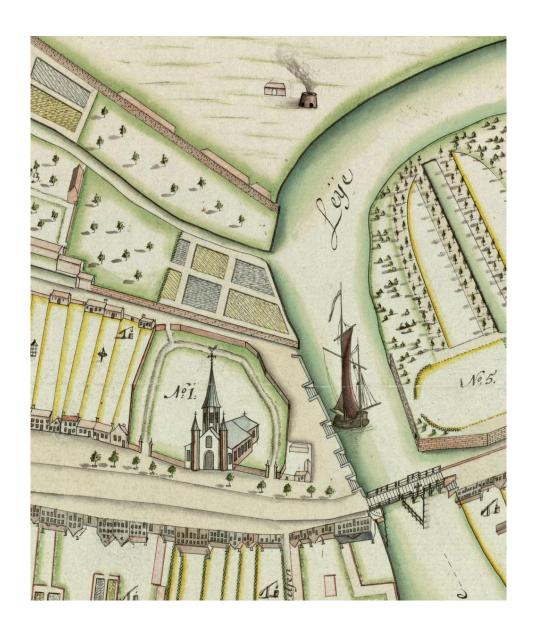




Archival Community





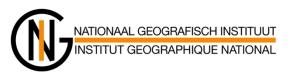


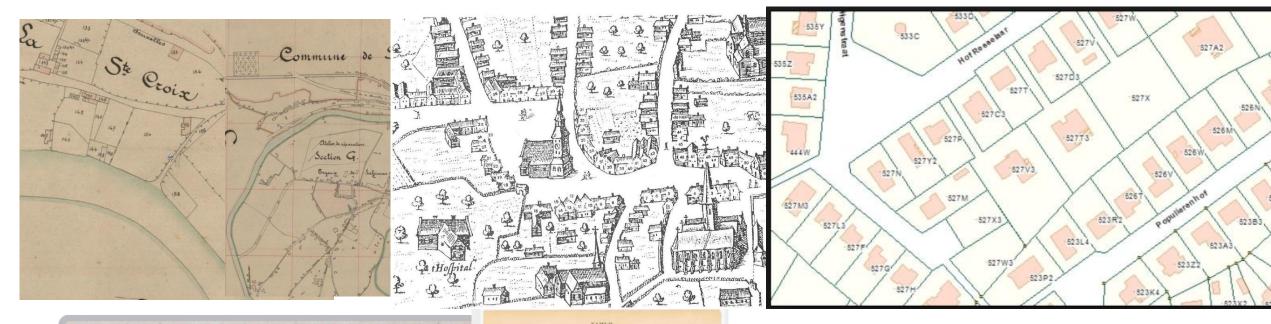
Questions?

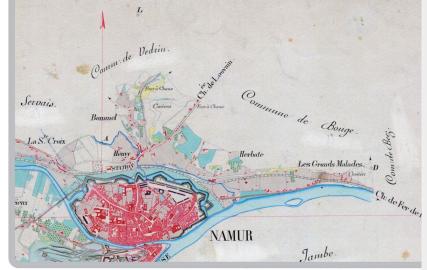
Rink.Kruk@ngi.be

Nevertheless

There is a community of the willing, but we need some leverage









CEF eArchiving domain introduction **Taxation and customs domain DG TAXUD project**



Ulrich Kröner

Head of Sector – Entreprise IT Architecture and Strategy (EAS)

DG TAXUD B2 – European Commission





Archiving system at TAXUD

- Specific business need
- Challenges, how addressed by eArchiving
- Project execution and planning



Archiving system at TAXUD – Business Need

- New ICS2 system (and later others)
- Exchange with Economic Operators: CEF eDelivery AS4 Profile
 - TAPAS (TAXUD Access Point for AS4) implemented using DIGIT's Domibus
 - Message signature for non-repudiation, certificates in message, TLS
- High volume of small messages, daily archival, keep for 10 years
- Archival of audit and logging records, keep for 7 years



Archiving system at TAXUD - Objectives

- Storage of defined data sets together with an associated set of metadata (to, from, date, ID, ...)
- Retrieval of archived records sporadically on demand
- Archiving and retrieval secure and business-agnostic

System access must ensure data protection (GDPR)



Archiving at TAXUD – Technical Challenges

- Archival of 8 million transactions per day (small size messages)
- Remaining business agnostic
- Key questions:
 - How? Methodology for identifying the data sets to be extracted. Format? Compress? Encrypt?
 - When? Timeframe and Frequency
- Validation, deletion and confirmation





Archiving solution



- KEEP Solutions services RODA/SOLR advantages
 - Long-term digital repository solution, capable of ingesting, managing, and providing access to various types of digital content
 - Ability to ingest 800k records in 16 hours on standard infrastructure
 - Scalable



Archiving solution decision

 Requirements communicated to DIGIT CEF/ KEEP Solutions

March '19

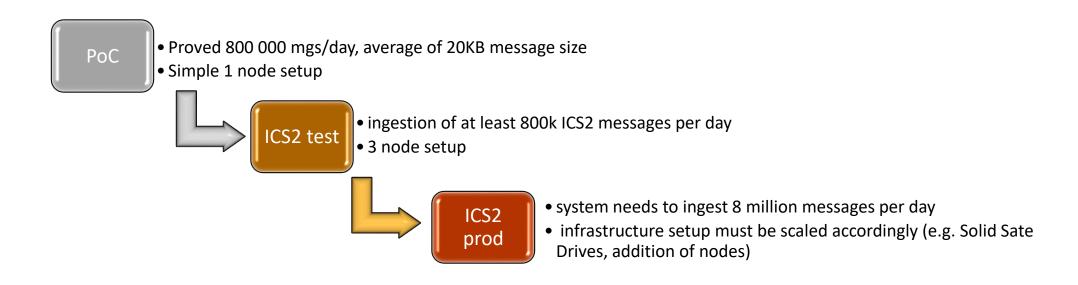
• eArchiving PoC confirmed 800K msg/day (1-node config., slow disks)

• GO decision

June '19

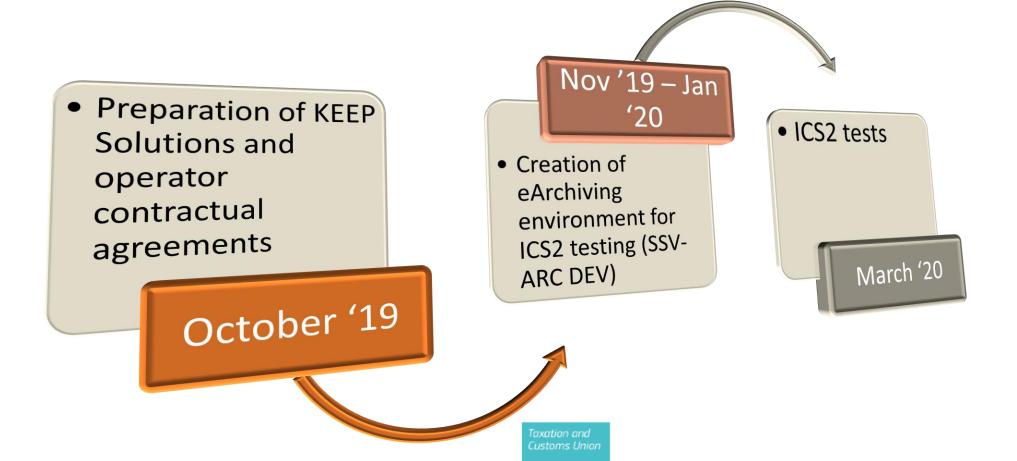


eArchiving phased requirements



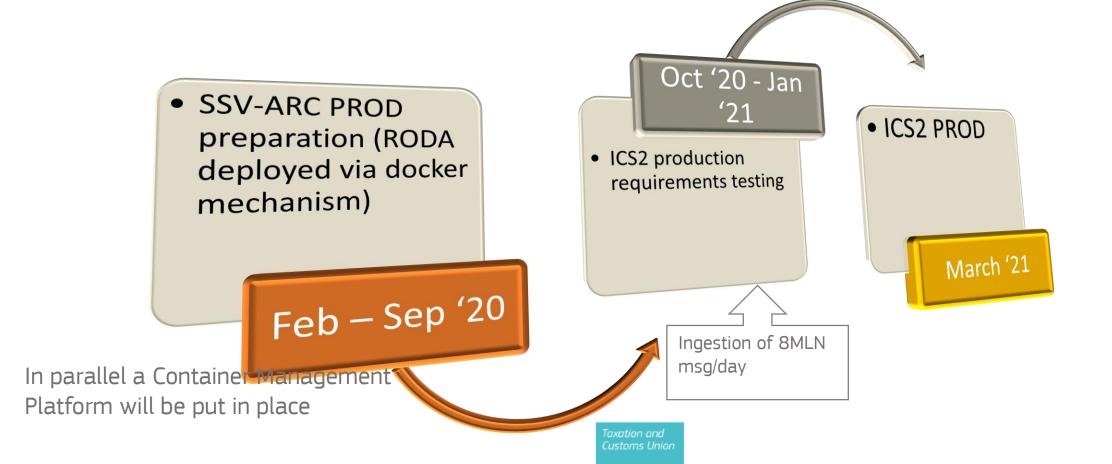


Implementing eArchiving as a Shared Service





eArchiving as a Shared Service – next steps





Archiving system at TAXUD

Take-away: Scaling and optimisations needed in order to support high message volumes.

CEF eArchiving domain

Programme Assistant- EU Policies - Interactive Technologies, Digital for Culture and Education,

DG CNECT G2 - European Commission

Panel/Q&A

introduction

Rink Kruk National Geographic Institute, Belgium

Adelina Cornelia Dinu



Ulrich Kröner

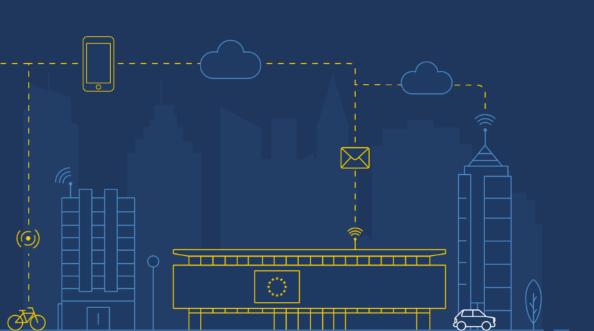
Head of Sector – Entreprise IT Architecture and Strategy (EAS)

DG TAXUD B2 – European Commission



Anja Paulic Archives of the Republic of Slovenia





Coffee break



Break-out session



Break-out session

What is the potential of CEF eArchiving in your domain?



Fulgencio Sanmartín

Policy Officer – Digital Cultural Heritage and Interactive Technologies,

DG CNECT G2 – European Commission



Questions



1 How do you deal with the preservation of your data? What reuse do you do of it?

- Digitisation? Websites, emails?
- Data Preservation Plan? Metadata?
- Cloud? Other outsourcing?



2 What kind of data are you dealing with, and what do you do with these opportunities?

- Domain? File formats?
- Variety? Volumes? How often?
- Opportunities?



3 What are the potentials and barriers of eArchiving in your domain?

- Legal? GDPR?
- Costs? Migration? Lock-in software?
- Potentials?



Plenary Session





Conclusion & wrap up



Fulgencio Sanmartín

Policy Officer – Digital Cultural Heritage and Interactive Technologies, DG CNECT G2 – European Commission



CEF Digital

Let's connect Europe, together. cef-building-blocks@ec.europa.eu

Fulgencio Sanmartín
Fulgencio.SANMARTIN@ec.europa.eu
DG Connect

Directorate-General for Communications Networks, Content and Technology Adelina Cornelia Dinu

Adelina-Cornelia.DINU@ec.europa.eu

DG Connect

Directorate-General for Communications Networks, Content and Technology #ConnectingEurope #eArchiving



Thank you for joining us in Brussels.

Let's keep in touch!

