



Revamped INSPIRE Geoportal: Cooking the next generation of spatial data catalogues

Jeroen Ticheler¹, José Garcia¹, David Blasby¹, Davide Artasensi²,
Jordi Escriu³, Marco Minghini³ and Alexander Kotsev³

¹ GeoCat B.V., Bennekom, The Netherlands

² Fincons Group, Lugano, Switzerland

³ European Commission, Joint Research Centre (JRC), Ispra, Italia

Florence, 25 August 2022

Agenda & speakers



Photo by [Glenn Carstens-Peters](#) on [Unsplash](#)

- INSPIRE Geoportal introduction
Jordi Escriu (JRC)



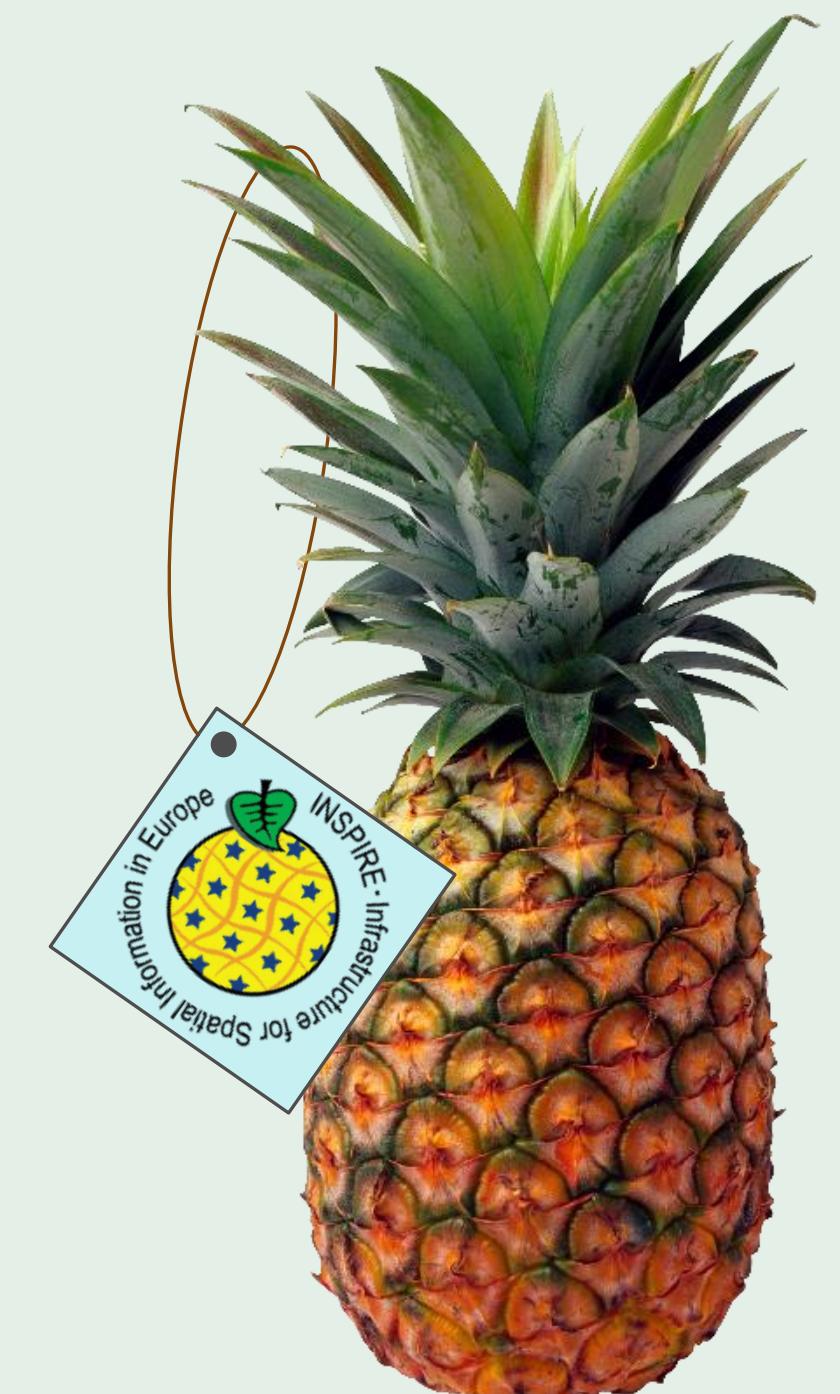
- Revamp based on GeoNetwork OS
Jeroen Ticheler (GeoCat B.V.)



INSPIRE

The geospatial pineapple

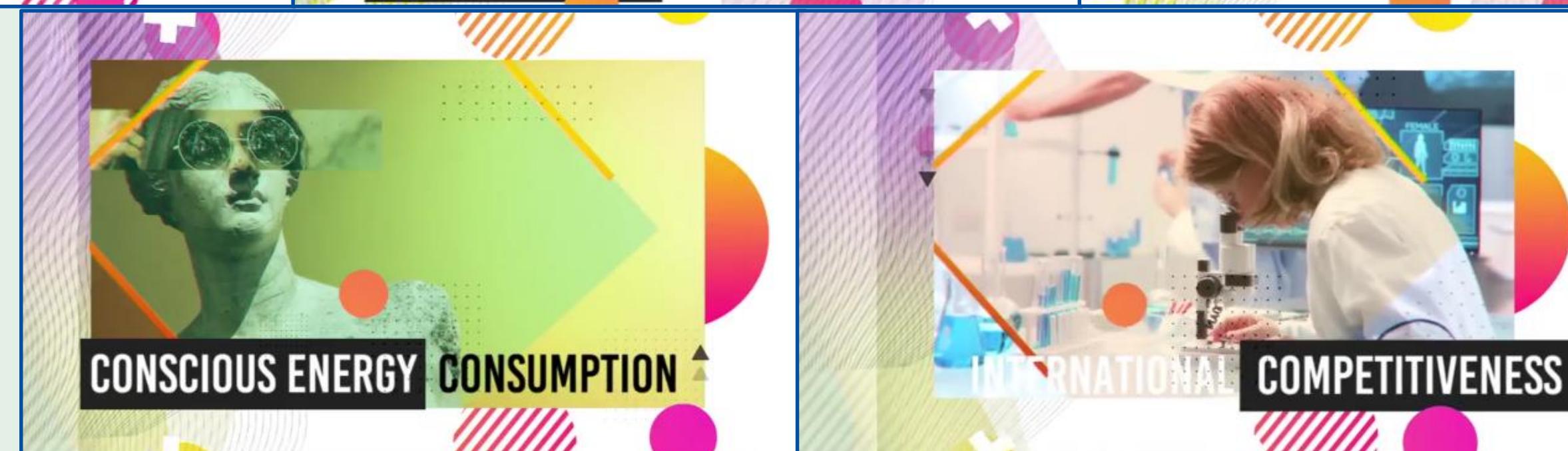
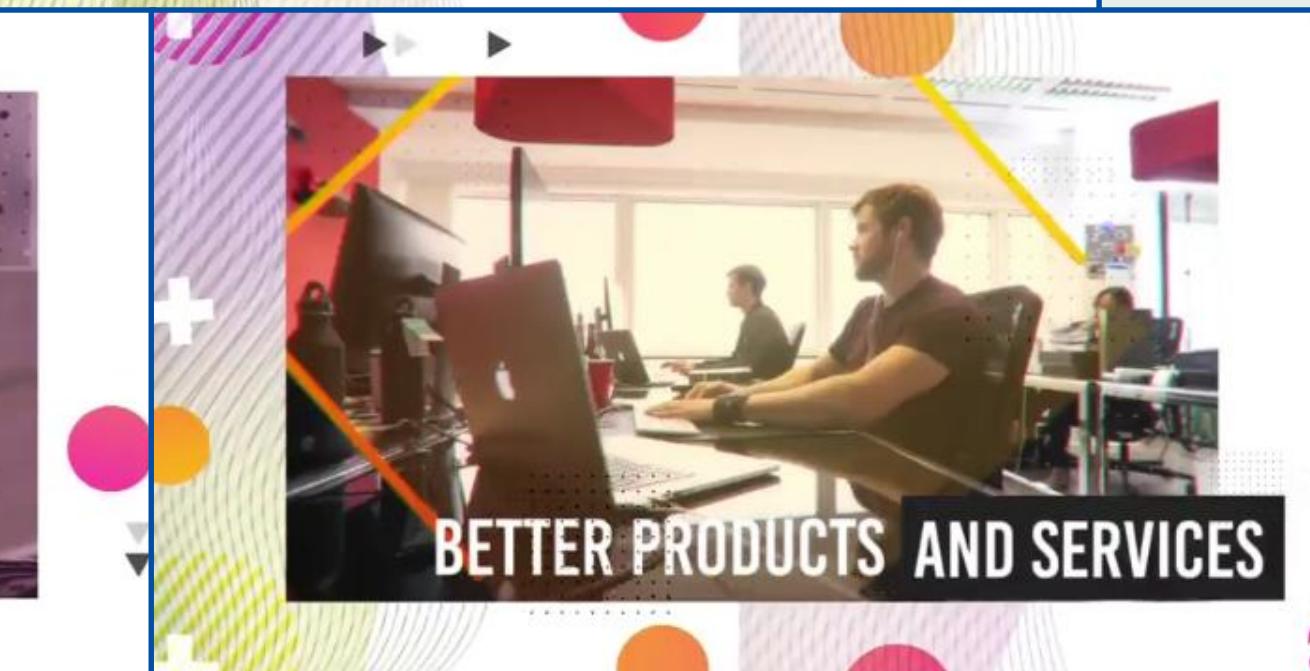
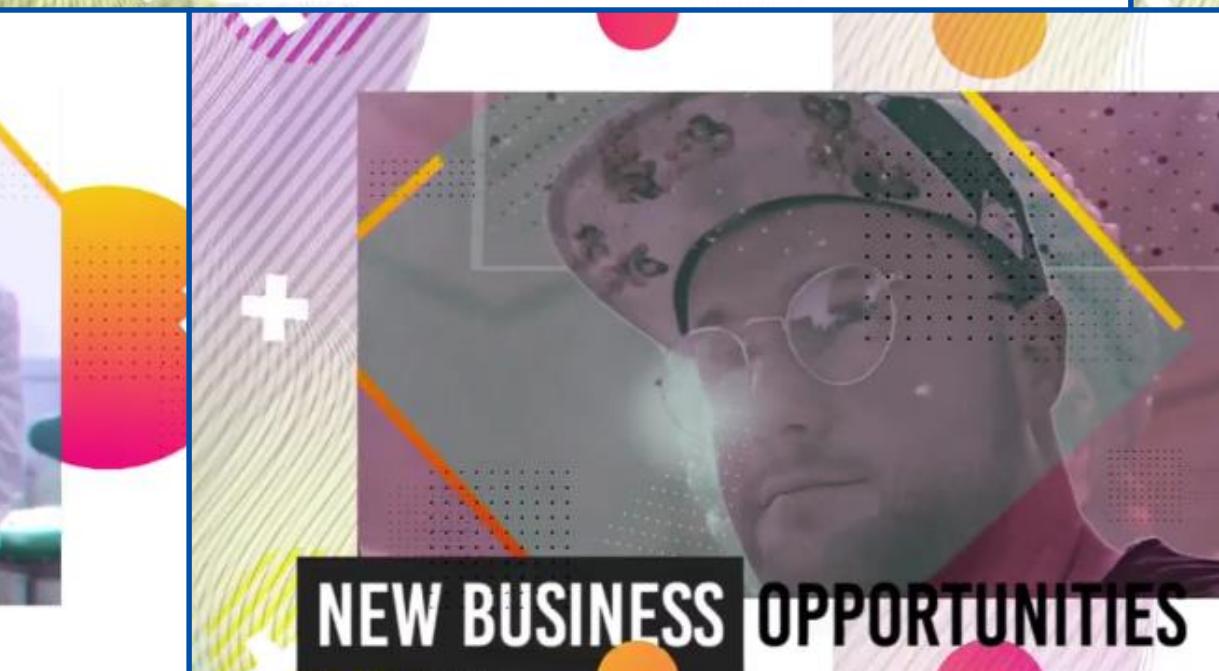
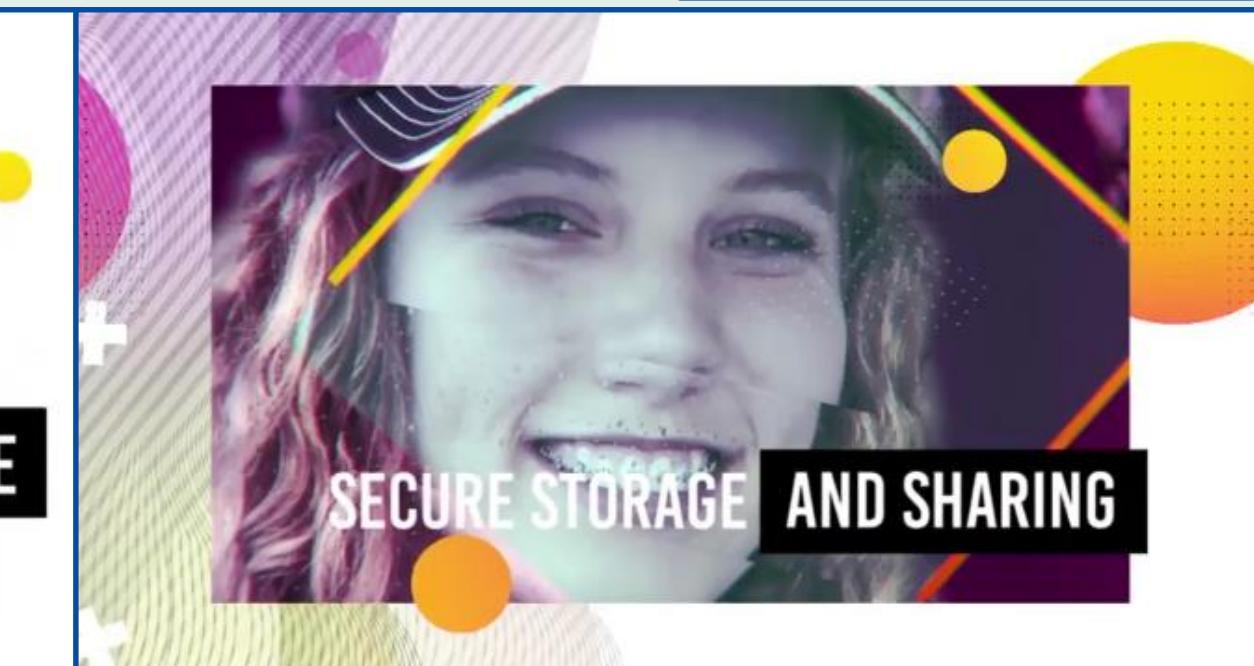
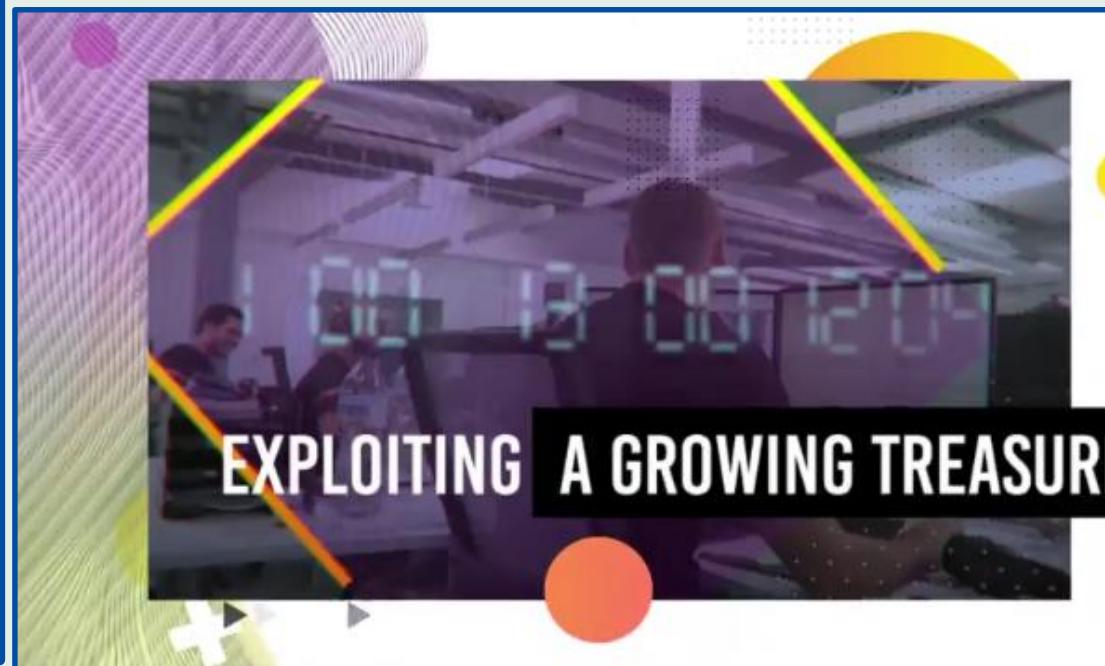
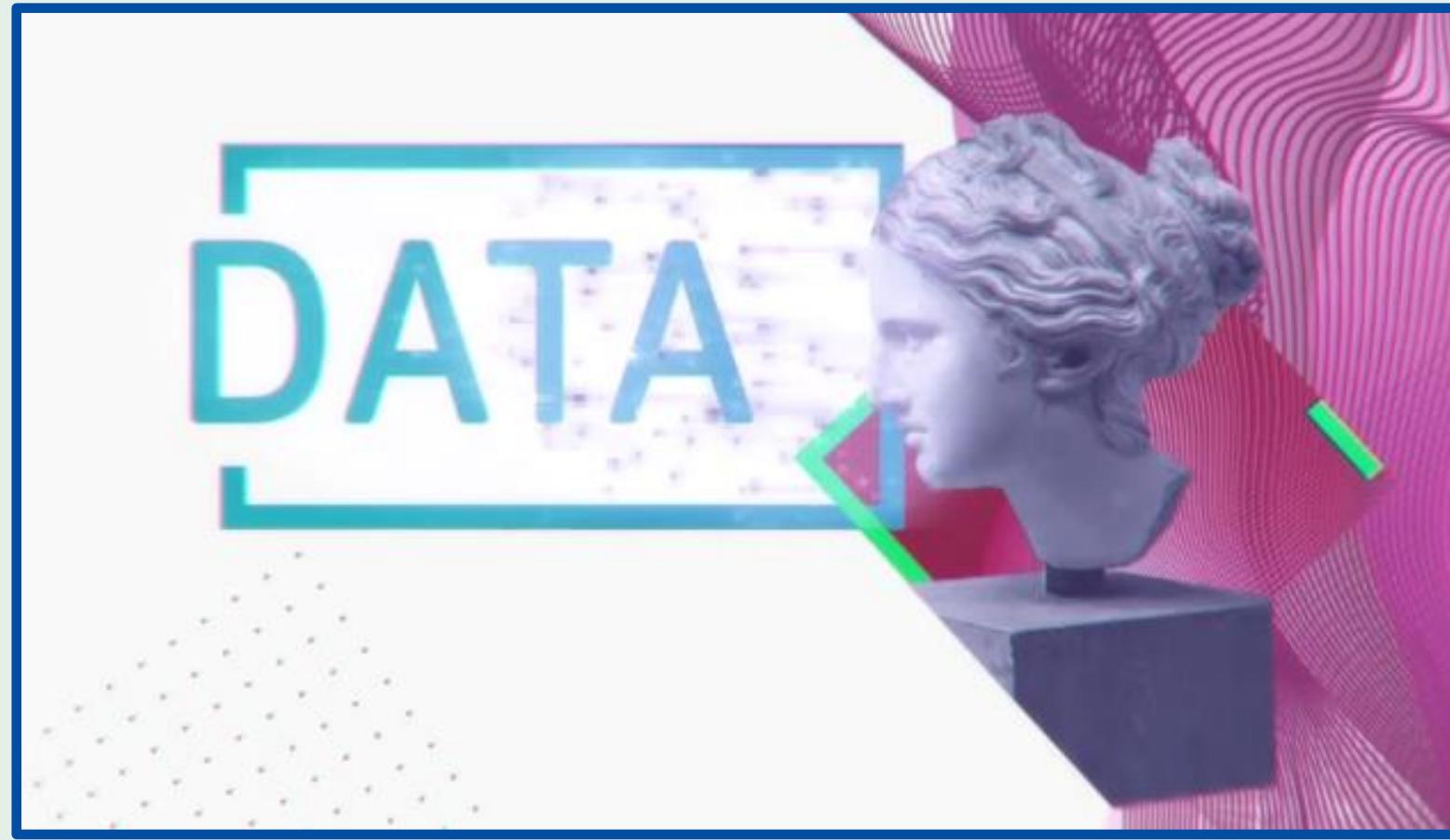
- One of the biggest geospatial data sharing initiatives in the world (**7000+ providers**).
- **Multi-faceted Spatial Data Infrastructure** framework
 - **Legal:** Directive, Implementing provisions, transposition in MS.
 - **Organisational:**
 - Governance structure with National contact points / structures, Multiple Commission Services.
 - Maintenance and Implementation Work programme;
 - **Technical:** Reusing building blocks from standardisation bodies (OGC, ISO, etc.)
 - Full stack of guidelines for discoverability, metadata, data encoding and data sharing.
- **Status of implementation**
 - Directive entered into force in 2007 / Roadmap finished by December 2021.
 - **Lights and shadows.** Objectives partially achieved. Heterogeneity of implementations across EU. Pan European coverage yet to be achieved.



A new context is ahead

New European digital society

#DigitalEU



INSPIRE Evaluation & Future JRC Science for Policy Report

- INSPIRE - A Public Sector Contribution to the European Green Deal Data Space

<https://publications.jrc.ec.europa.eu/repository/handle/JRC126319>



- Vision:
 - Evolution to a data ecosystem.
 - Broadening the scope to new sectors and communities.
 - Widening the range of applications and use cases.
 - Making INSPIRE more simple, flexible and agile.
 - Reusing well-adopted and working standards and technologies.



Defining now the future! Sectoral European data spaces

Rich pool of data
(varying degree of accessibility)

Free flow of data
across sectors and countries

Full respect of GDPR

Horizontal
framework for data
governance and data
access

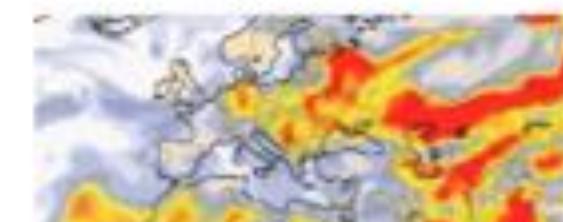


- Technical tools for data pooling and sharing
- Standards & interoperability (technical, semantic)
- Sectoral Data Governance (contracts, licenses, access rights, usage rights)
- IT capacity, including cloud storage, processing and services

Personal Data Spaces



High Value Data sets



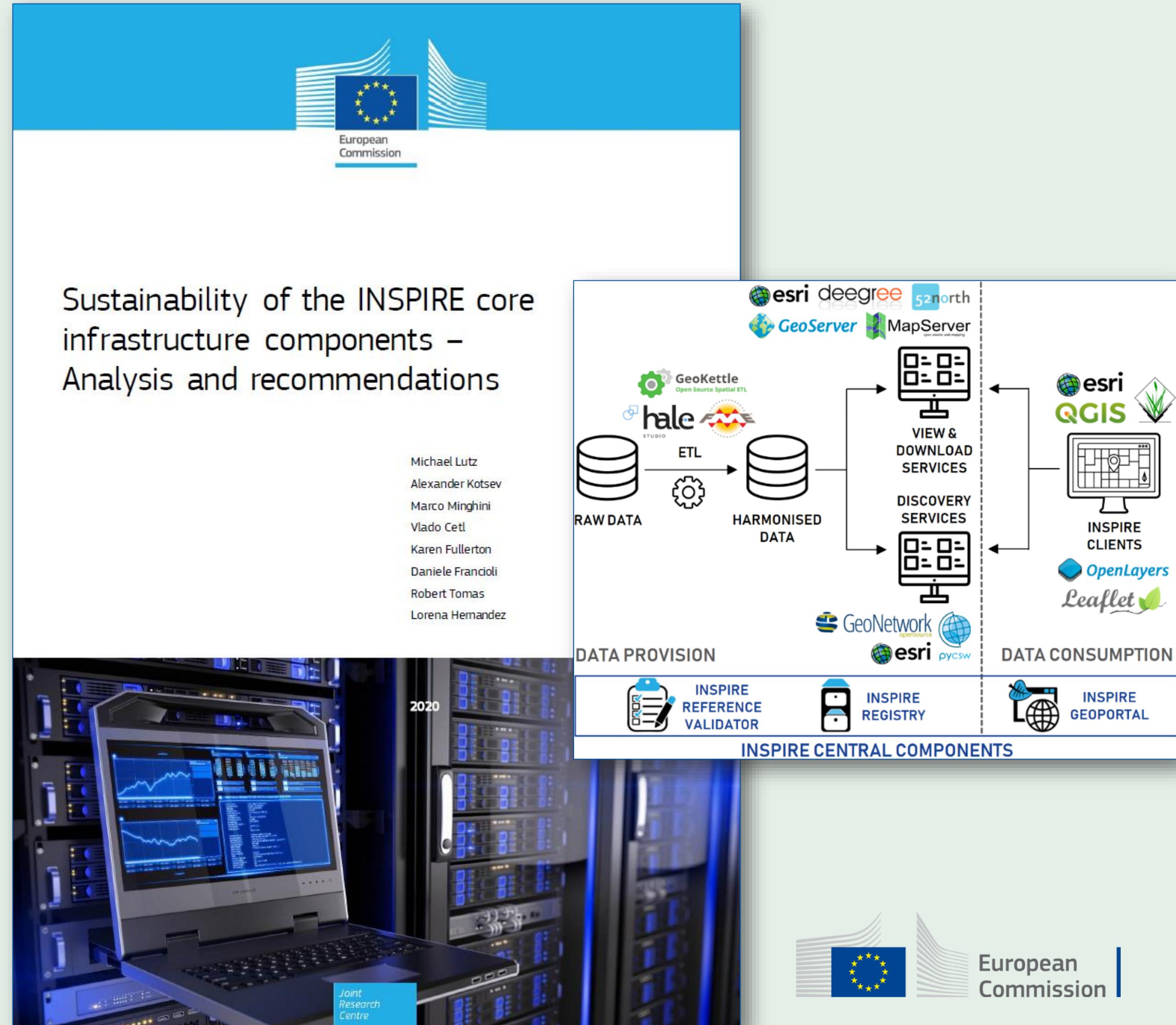
Public Sector
geospatial,
statistics.



Central INSPIRE Components

Sustainability approach

- Sustainability of the INSPIRE core infrastructure components
 - **Support by tools is the default.**
 - **Build strategic partnerships with communities: OSGeo, GeoNetwork.**
 - Focus on the INSPIRE-specificity and not on mainstream tool development.
 - Harmonise the approaches for helpdesk.
 - Decouple tools from infrastructure.
 - Extensive use of the cloud.



INSPIRE Geoportal

What we have

Main access point to INSPIRE Infrastructure tools and resources.

- Developed by JRC.
- Main focus: INSPIRE Monitoring and Reporting.
- Being revamped using a GeoNetwork-based backend.

<https://inspire-geoportal.ec.europa.eu>

The screenshots illustrate the current state of the INSPIRE Geoportal:

Left Screenshot (Home Page):

- INSPIRE GEOPORTAL**: Enhancing access to European spatial data
- European Commission > INSPIRE > Geoportal
- Home | Priority Data Sets Viewer | Thematic Viewer | Harvesting status | Find out more about
- Welcome to the INSPIRE Geoportal**: The INSPIRE Geoportal is the central European access point to the data provided by EU Member States and several EFTA countries under the INSPIRE Directive. The Geoportal allows:
 - monitoring the availability of INSPIRE data sets;
 - discovering suitable data sets based on their descriptions (metadata);
 - accessing the selected data sets through their view or download services.
- The metadata used in the Geoportal are regularly harvested from the discovery services of EU Member States and EFTA countries. The status of harvesting is available [here](#).
- Feedback regarding the functionality as well as data set availability is welcome [here](#).
- Priority Data Sets Viewer**: Application displays priority data sets for environmental reporting.
- INSPIRE Thematic Viewer**: Application displays all MS data sets falling under the scope of INSPIRE Directive.
- INSPIRE Reference Validator**: Application helps data providers check data set requirements.

Right Screenshot (EU & EFTA Country overview):

- INSPIRE GEOPORTAL**: Enhancing access to European spatial data
- European Commission > INSPIRE > Geoportal
- Home | Priority Data Sets Viewer | Thematic Viewer | Harvesting status | Find out more about
- INSPIRE Data Sets - EU & EFTA Country overview**: Map of Europe showing data availability by country.
- INSPIRE Geoportal Data Set Statistics**:
 - 84729 Metadata records
 - 57494 Downloadable Data Sets
 - 57704 Viewable Data Sets
- Spatial scope coverage:**
 - National
 - Regional
- Select a COUNTRY**: List of countries with their data availability statistics.
- Select the whole EUROPE**: Download stats button.
- Version: 1.6.2

INSPIRE Geoportal

Revamp objectives

- **Migration to a GeoNetwork-based architecture.**
 - Maintaining the functionality of the current INSPIRE Geoportal.
 - Improved metadata harvesting processes & performance (use of micro-services).
 - Harvesting dashboard protected through EU Login-based authentication.
 - Automated translations of metadata using the APIs offered by EC DGT Translation unit.
- **Migration to a cloud environment:** assure system availability / scalability.
- **Contribution to GeoNetwork core development.**

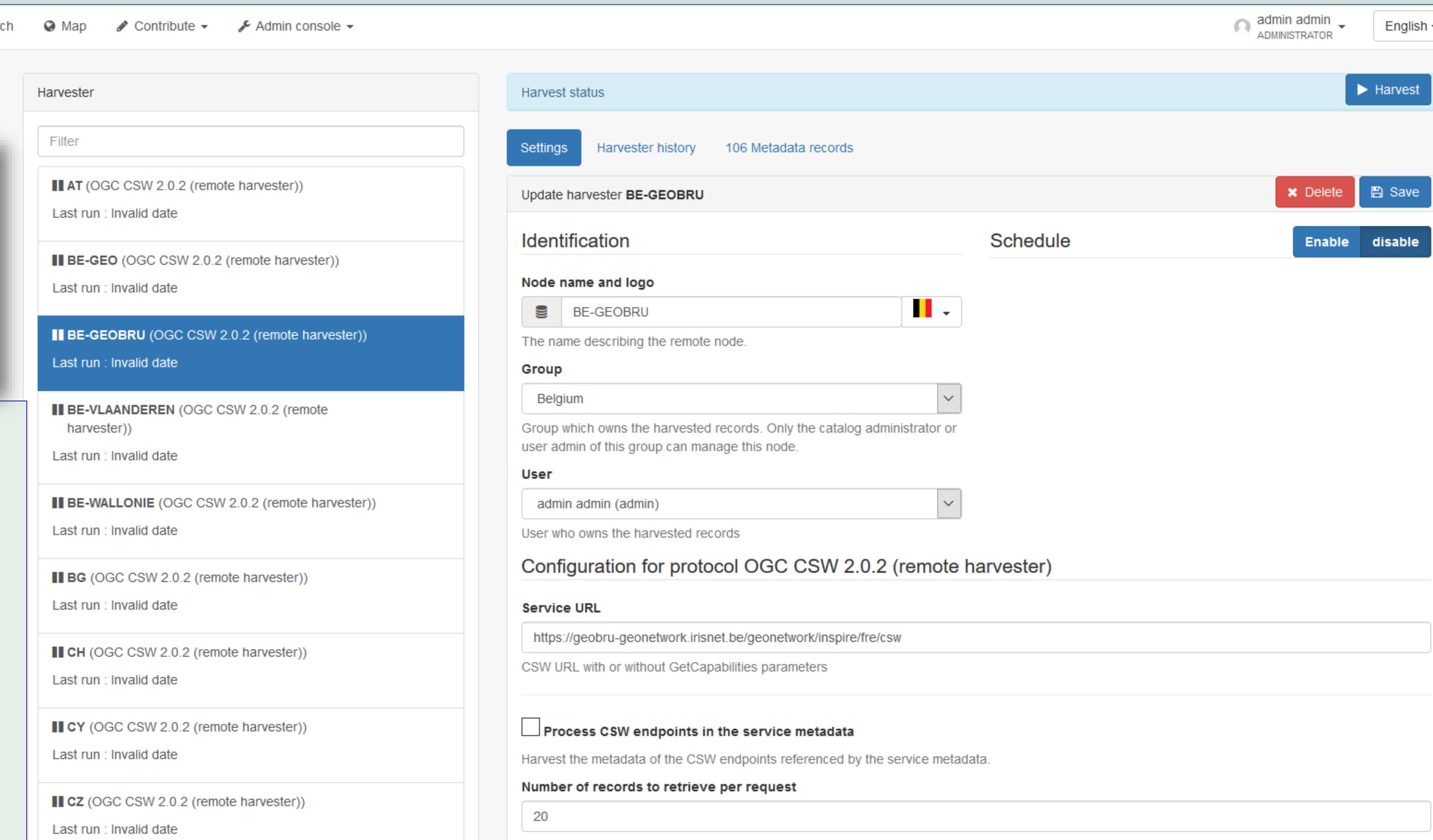


INSPIRE Geoportal Revamp – Backend: GeoNetwork (in the cloud)



GOVERNMENT GEOGRAPHIC DATA PUBLISHING

- Prototype:
Available and revised
- Working system:
Delivered
JRC Testing finished
GeoCat improving the system



The screenshot shows the GeoNetwork Admin console interface. At the top, there are navigation links: My GeoNetwork catalogue, Search, Map, Contribute, and Admin console. On the right, there is a user profile for 'admin admin' (ADMINISTRATOR) and language selection ('English'). The main content area has a header 'Harvester status' with tabs for Settings, Harvester history, and 106 Metadata records. A large blue button at the top right says '▶ Harvest'. Below this, there's a section for 'Update harvester BE-GEOBRU' with fields for 'Identification' (Node name and logo: BE-GEOBRU, Belgium), 'Group' (Belgium), and 'User' (admin admin). Under 'Configuration for protocol OGC CSW 2.0.2 (remote harvester)', there's a 'Service URL' field with the value 'https://geobru-geonetwork.irisnet.be/geonetwork/inspire/fre/csw', a checkbox for 'Process CSW endpoints in the service metadata', and a 'Number of records to retrieve per request' field set to '20'. To the left, a sidebar lists 'Catalog harvesters' and 'Catalog harvester report'. The main panel displays a list of harvesters, with 'BE-GEOBRU' highlighted in blue.

Harvester	Last run
AT (OGC CSW 2.0.2 (remote harvester))	Invalid date
BE-GEO (OGC CSW 2.0.2 (remote harvester))	Invalid date
BE-GEOBRU (OGC CSW 2.0.2 (remote harvester))	Invalid date
BE-VLAANDEREN (OGC CSW 2.0.2 (remote harvester))	Invalid date
BE-WALLONIE (OGC CSW 2.0.2 (remote harvester))	Invalid date
BG (OGC CSW 2.0.2 (remote harvester))	Invalid date
CH (OGC CSW 2.0.2 (remote harvester))	Invalid date
CY (OGC CSW 2.0.2 (remote harvester))	Invalid date
CZ (OGC CSW 2.0.2 (remote harvester))	Invalid date

INSPIRE Geoportal Revamp – User interface

- **Available** – Filtering by High Value Datasets (HVDs).

To be launched when the Implementing Act on HVDs is entering into force.

- **Almost ready** – Additional developments to integrate the new user interface with the new backend: GeoNetwork.

The screenshot shows the homepage of the INSPIRE Geoportal. At the top, there is a header with the European Commission logo and a language selection link ('EN English'). Below the header, the page title is 'INSPIRE GEOPORTAL' with a subtitle 'ENHANCING ACCESS TO EUROPEAN SPATIAL DATA'. A navigation bar includes links for 'Home', 'High-Value Data Sets', 'Thematic Data', 'Harvesting status', and 'Find out more about'. The main content area features three large cards:

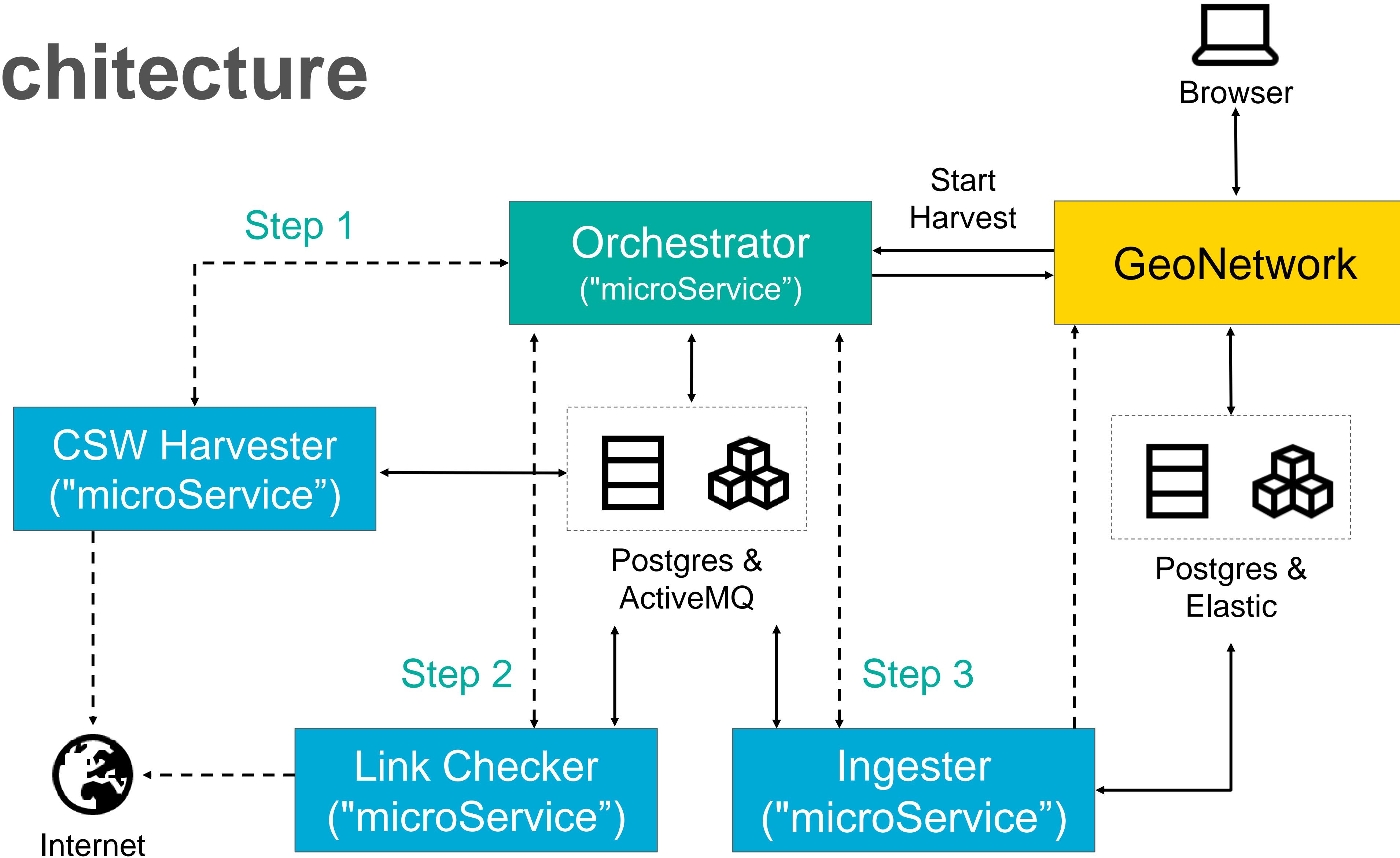
- High-Value Data Sets**: Shows a map with a green circle containing a color fan icon. Text: 'The application provides an overview and access to geospatial high-value datasets and other core data (including priority data sets for eReporting)'. A 'Browse' button is at the bottom.
- INSPIRE Thematic Data**: Shows a map with a yellow globe icon. Text: 'The application displays the availability and provides access to all EU MS data sets falling under the scope of INSPIRE Directive filtered by data themes and countries (i.e. Annex I, II and III)'. A 'Browse' button is at the bottom.
- INSPIRE Reference Validator**: Shows a map with a small globe icon. Text: 'The validator application is to help data providers, solution providers and national coordinators to check whether their data sets, network services and metadata meet the requirements defined in the INSPIRE Technical Guidelines. The validation tests are based on the Abstract Test Suites agreed between Member States and the Commission in the INSPIRE Maintenance and Implementation Group.' A 'Go to INSPIRE Reference Validator' button is at the bottom.

INSPIRE Geoportal

Revamp based on GeoNetwork opensource

- Project to replace the existing Geoportal with GeoNetwork opensource
- 4 new components developed
- Adapted GeoNetwork v4 to work with those components
- Fast harvester and link analyzer with configurable, parallel processing

Architecture



The Orchestrator



Orchestrator

- Orchestrate the working of different components to trigger their execution
- The harvester console uses the orchestrator to monitor the execution of the different components and report the progress to the user

CSW Harvester

(“microService”)



Input

- Configured by the administrator using the GeoNetwork UI
 - URL
 - CSW Filter
 - What to do when an error occurs
 - Parallelism hint (to the remote CSW)
 - Number of records per request
 - Look for Nested Discovery Services (i.e. Poland)

Output

- PostgreSQL database with all the harvested XML records

Link Checker

(“microService”)



Inputs

- DB with Harvested XML Records (previous step)
- Configured by GeoNetwork User Interface
 - HTTP Cache usage
 - Max number of runs to store
 - How many download links to check
 - How many Atom sub-feeds to check

Outputs

- PostgreSQL database with a complex data model
 - Information about all the Capabilities Documents
 - Service documents have indicators attached to them
 - Dataset documents have indicators attached to them
 - All the “links to actual data” found
 - Results for all the data attempted to be downloaded (or “viewed”)

Processing Steps:

A. Reads all the Harvested Metadata records (“Find Links”)

- Parses all the XML Documents
- Saves them in the DB object Model

A. Go through all Service and Dataset Records and follow all the links (“Check Links”)

- Downloads (and parses) all the Capabilities documents (and Atom Feeds)
- Saves them in the DB object Model

Processing Steps:

C. Searches for Indicators in the DB (“Post Processing”)

- Makes connections between Dataset Records and Service “layers” (OGC/Atom)
- Computes Indicators
- Saves them in the DB object Model

D. Verifies Downloadability and Viewability (“Data Downloading”)

- Attempts to download data from WFS/Atom feeds
- Attempts to view maps from WMS/WMTS services

Ingester

(“microService”)



Inputs

- DB with Harvested XML records (1st step)
- DB with Link checker results for the Harvested XML records (previous step)
- DB with GeoNetwork metadata.

Processing Steps:

A. Process all the Harvested Metadata records

- Check if the metadata exist in the GeoNetwork DB and has changes to update / create it if necessary
- Retrieve the associated indicators calculated in the Link Checker and store them in the GeoNetwork DB object Model

B. Remove from the GeoNetwork DB the metadata that is no longer available in the harvested server

C. Index the metadata to be searchable

Outputs

- GeoNetwork database / ElasticSearch index updated with:
 - Metadata harvested for the member states.
 - Metadata associated indicators, calculated in the Link Checker.

GeoNetwork

The new Harvester component



Harvester configuration

- Configuration of the orchestrator microservice end-point in the GeoNetwork settings page:

The screenshot shows the 'Harvesters' configuration page in GeoNetwork. It contains three main settings:

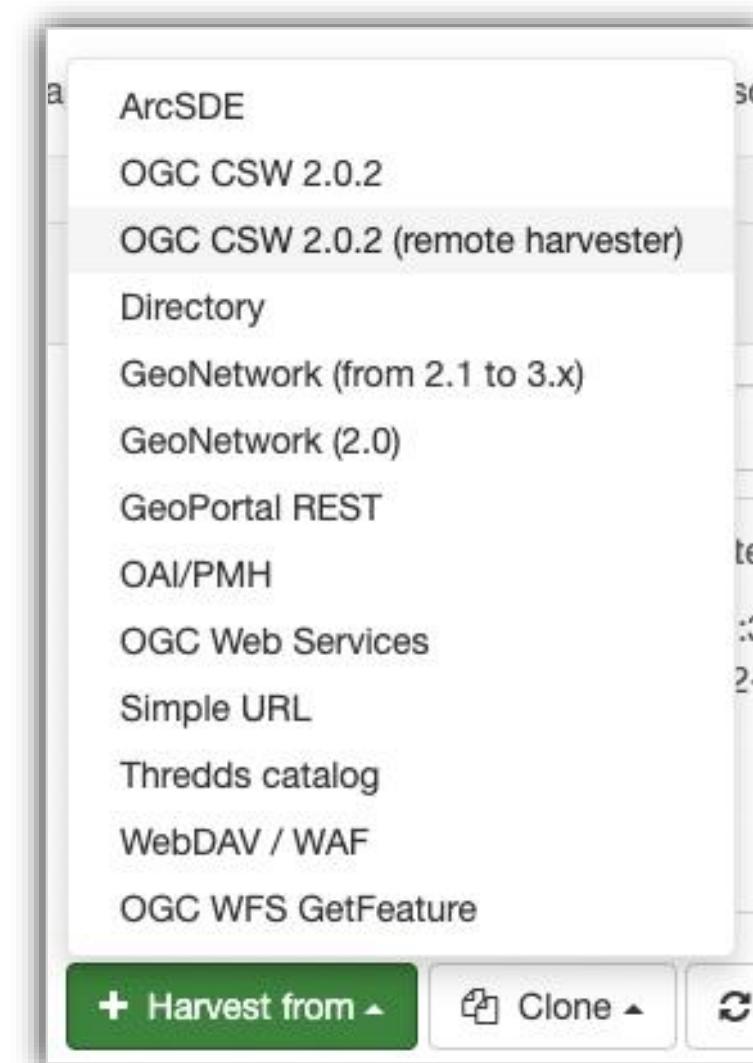
- Allow editing on harvested records**: A checkbox that is currently unchecked.
- Disabled harvester protocols**: A text input field containing a list of disabled harvested protocols, such as 'arcsde, csw, filesystem, geonetwork, geonetwork20, geoPREST, oaipmh, ogcwxs, thredds, wfsfeatures'.
- Remote harvester API endpoint**: A text input field containing the URL 'http://jrc-sandbox2.geocat.live:5555/api'.

Below the 'Remote harvester API endpoint' setting, there is a descriptive note: 'Remote harvester API endpoint, to execute the harvesting in a external process. For the time being only CSW harvester is supported.'

Harvester console

Added a new harvester type: Remote CSW harvester, to configure the CSW servers of the member states to harvest

- URL of the server to harvest
- CSW filter configuration
- Error conditions handling
- etc.

A screenshot of a configuration page for a harvester named 'ES'. The top navigation bar includes 'Settings', 'Harvester history', and a count of '542 Metadata records'. Below the navigation are several buttons: 'Update harvester ES' (red), 'Delete' (red), 'Save' (grey), 'Link checker' (grey), 'Harvest' (green), and 'Push metadata' (green). The main content area is titled 'Identification' and contains a 'Node name and logo' section where 'ES' is listed next to a small Spanish flag icon. A descriptive text below says 'The name describing the remote node.' Another section titled 'Group' shows 'Spain' selected in a dropdown menu. A note below states: 'Group which owns the harvested records. Only the catalog administrator or user admin of this group can manage this node.'

Harvester console

Provides access to the harvester execution history

The screenshot shows a web-based harvester history interface. At the top, there are three tabs: 'Settings' (blue), 'Harvester history' (white with blue border), and '540 Metadata records' (blue). Below the tabs, the title 'Harvester history' is displayed, followed by a red 'Delete' button. The main content area lists three harvested jobs, each represented by a green circular arrow icon and a summary box:

- 540 record(s) harvested in 4047 seconds**
🕒 a month ago
◦ total: 540
- 543 record(s) harvested in 1079 seconds**
🕒 a month ago
◦ total: 543
- 543 record(s) harvested in 796 seconds**
🕒 2 months ago
◦ total: 543

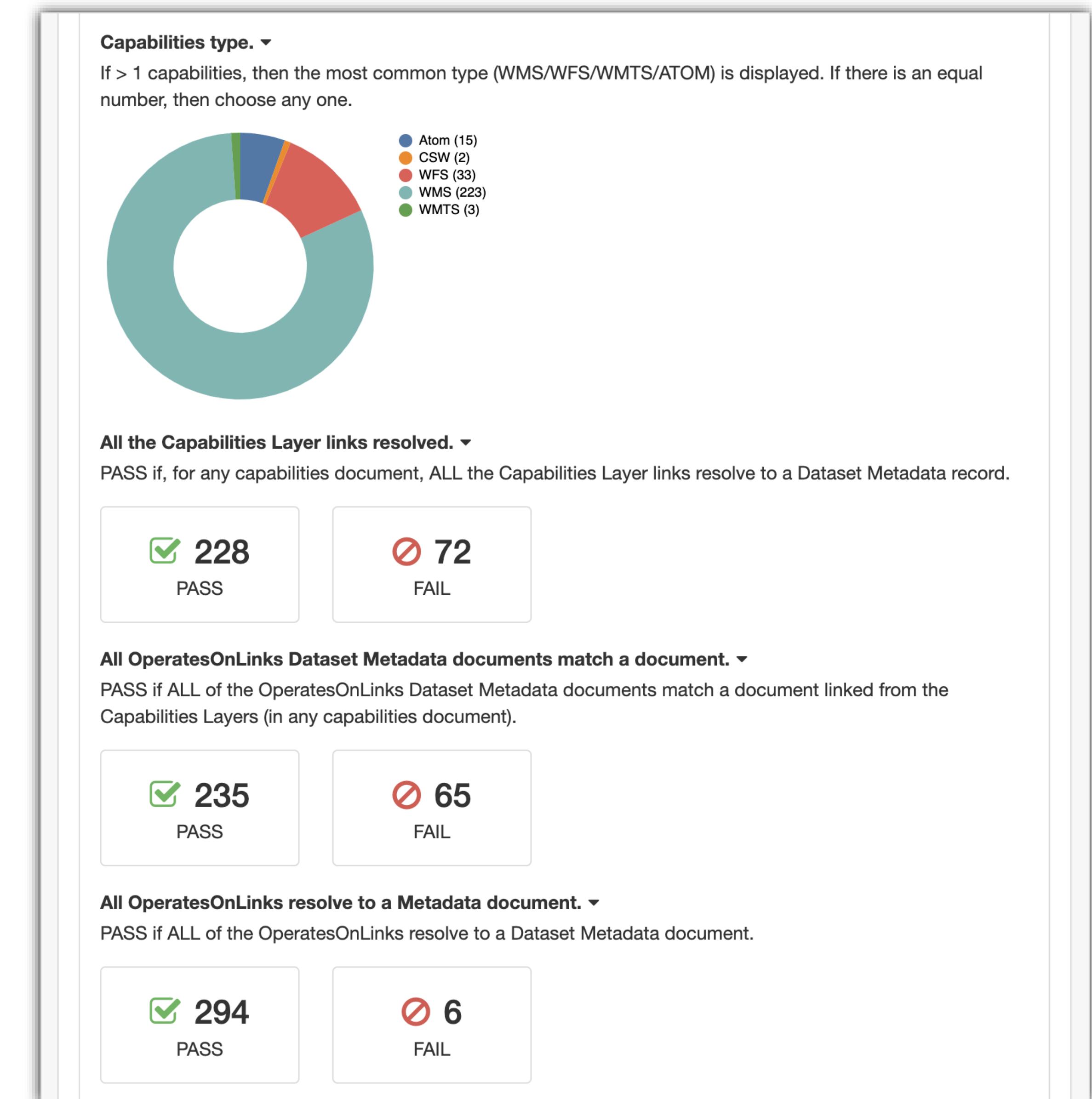
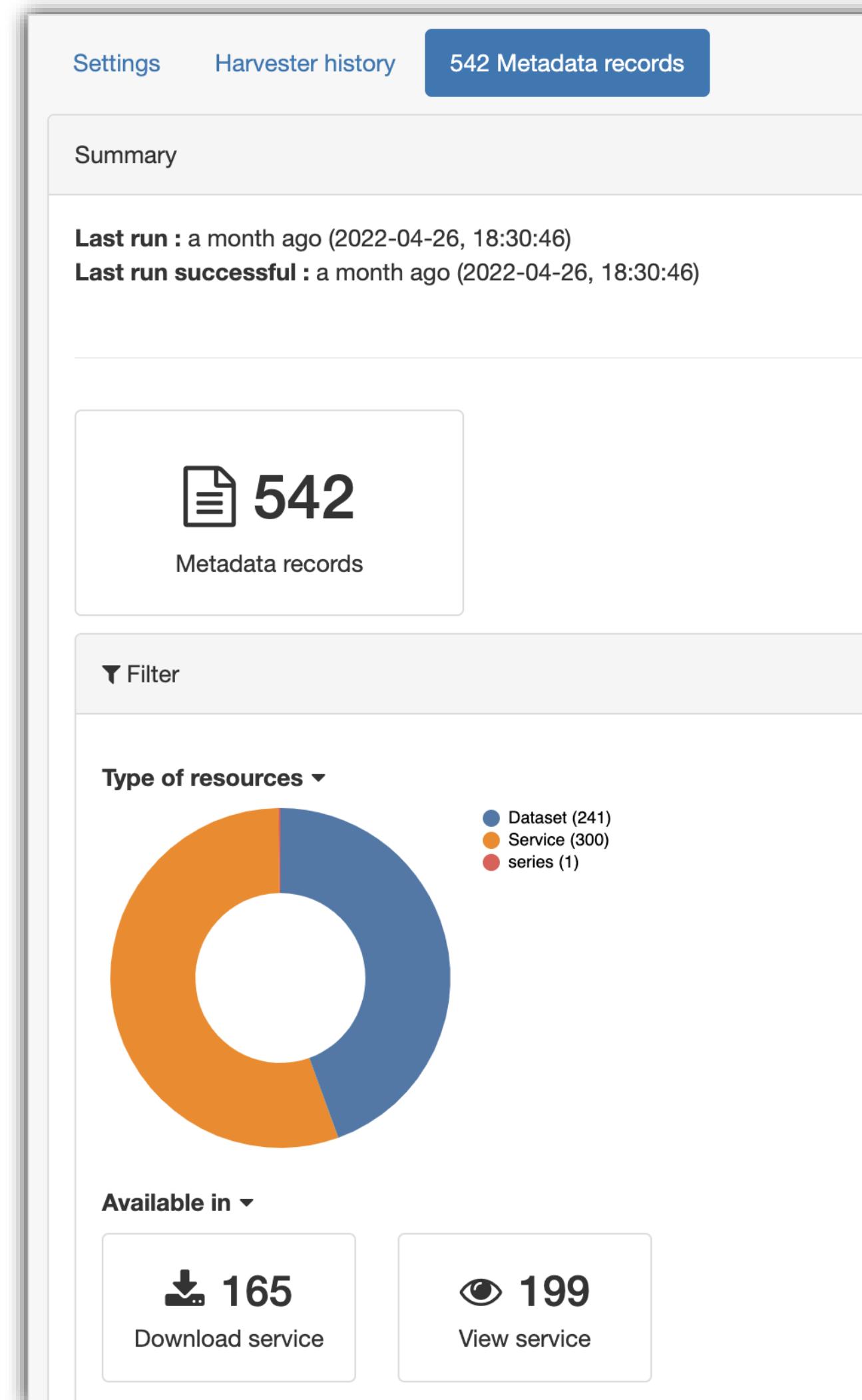
Each job summary includes a 'Log file' download button. At the bottom, there are navigation icons for back, forward, and search, along with the text '1 / 2 of 6 record(s)'.

Harvester console

And statistics about the metadata harvested, including:

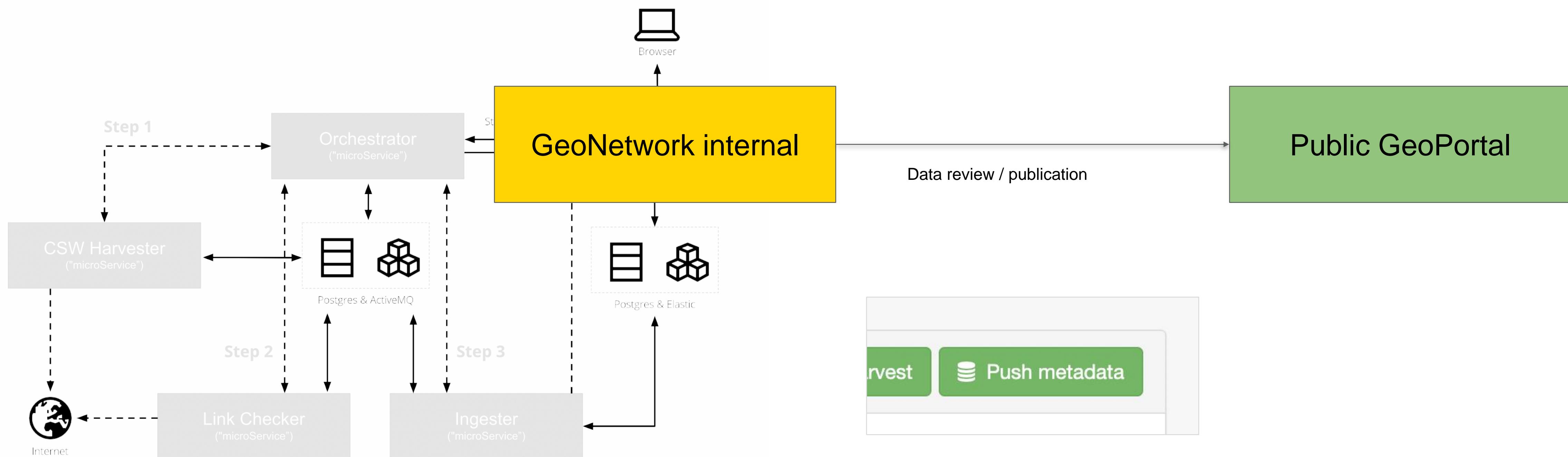
- Date of the harvesting
- Number of metadata records harvested
- Number of metadata per type: datasets / series / services
- Downloadable datasets
- Viewable datasets

Harvester console



Harvester console - data publication

Member states can review the harvester results and publish them to the public GeoPortal



Thank you!



jordi.escriu@ec.europa.eu

jeroen.ticheler@geocat.net



@JordiEscriu



© European Union 2020

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](#) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.