

Analytics & Reporting at different levels for a CRIS based on DSpace

the use case of the Peruvian National Platform

VISION, GOAL AND STRATEGY - <https://perucris.concytec.gob.pe/>

Funded by the World Bank, **#PeruCRIS** is the Peruvian project for setting up and operating the National Information Network on Science, Technology and Technological Innovation. It is based on open source software and open standards, specifically as an extension of the already existing network of Peruvian open access repositories, and it is led by Concytec, the Peruvian Council for Science, Technology and Innovation.

The vision for **#PeruCRIS Project** is:



Vision

To have an **information ecosystem** in science, technology and technological innovation in Peru to provide value, accessibility and development for all people.

#PeruCRIS Project has as main goal:



Main Goal

To develop an **open, interoperable and integrated national network** that provides timely and reliable information services on the actors, activities and results of science, technology and technological innovation in Peru, in favor of the scientific and academic community, the public and private sector, and the general public.

In order to accomplish the goal for **#PeruCRIS Project**, Concytec implements the following strategy:

Main strategy

1



Articulate, operate and regulate the National Information Network on STI

2



Strengthen the National Network of Open Access Repositories (RENARE)

3



Launch the #PeruCRIS platform and interoperate with institutional research management systems

4



Provide value-added services for accessing and managing information on STI in the country



EXPECTED BENEFITS

- Timely statistics and reports on national R&D activities.
- Better monitoring and evaluation for public funding and national open access policy.
- Disseminate research results, analyze trends and impact.
- Sharing and discovery for innovative technologies and ideas, new markets, competitors and partners.
- Better decision making at different levels: national, local, institutional, private sector and general population.



**Scientific
community**

—
Researchers, technological
developers, innovators,
entrepreneurs, managers
and planners in STI



**Institutions &
businesses**

—
Universities, public
research institutes,
funding agencies,
companies



Government

—
Ministries, regional or local
governments

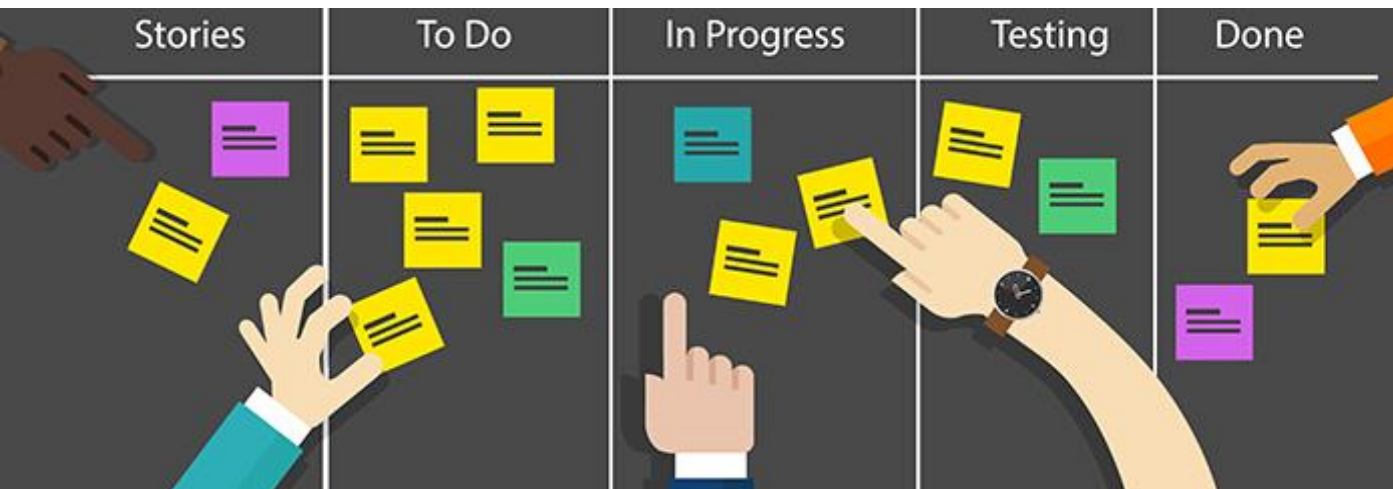


**General
population**

Project pillars

Sustainability:

- Open-Source Software
- Agile Methodology



Interoperability:

- Open standards
- Best practices

CERIF (Common European Research Information Format)
CERIF XML & CERIF API

OpenAIRE Guidelines for CRIS Managers
OpenAIRE Guidelines for Literature Repositories

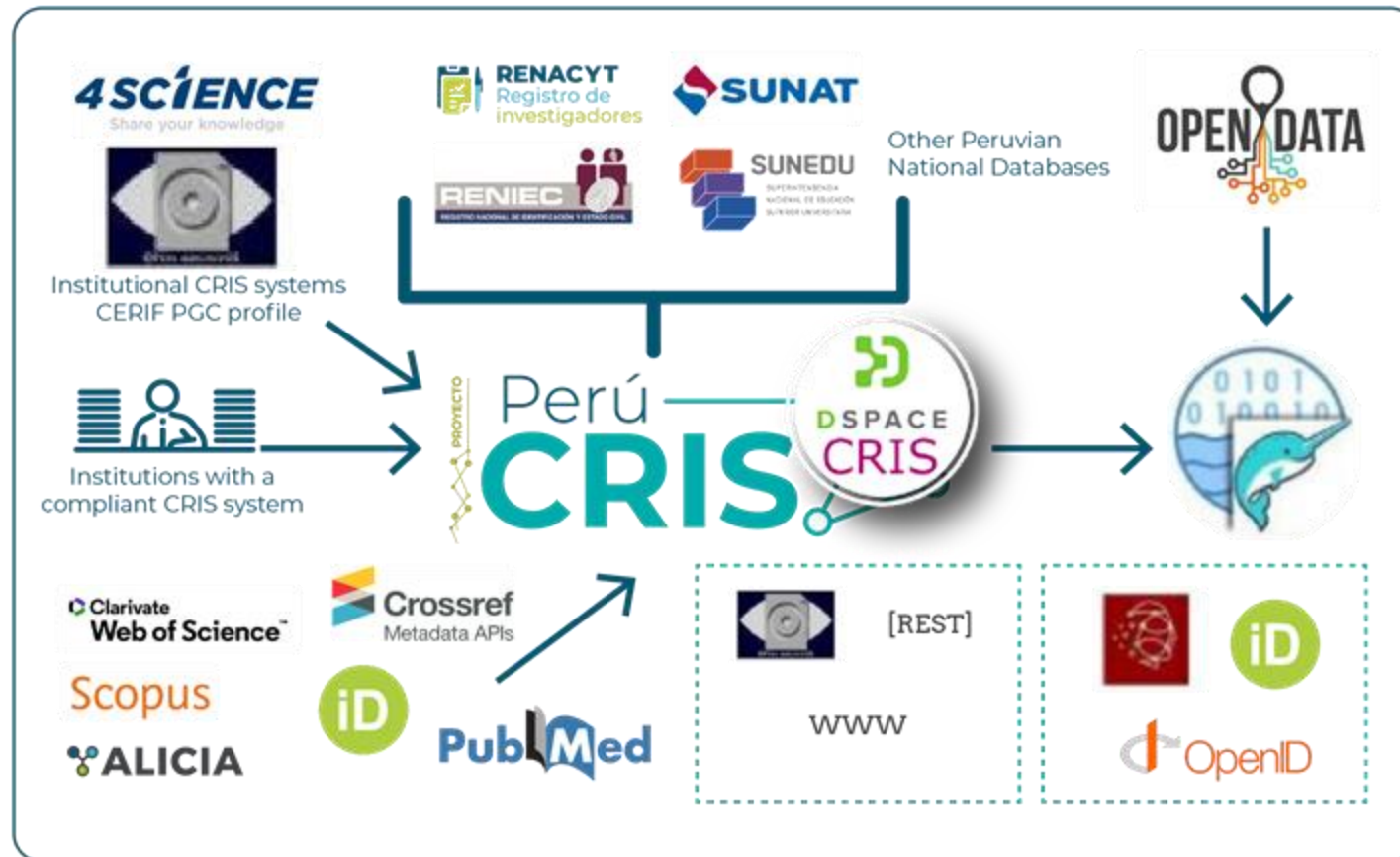
OAI-PMH Harvesting Protocol

Persistent Identifiers
ORCID, ROR, DOI, Handle

#PeruCRIS Guidelines
(application profile, semantics, syntax and scope of data)

GENERAL SOFTWARE ARCHITECTURE

Main Directories will be populated through the following flows:



1. A subsystem allows for institutions to have private entities collections under their own private communities, where they are able to bulk import or directly edit information to be submitted to the main public directories.
2. OAI-PMH harvesting with PeruCRIS custom CERIF XML profile, designed to be the preferred way for data ingesting.
3. Custom connectors for national interoperability (personal national records - RENIEC, national degrees database - SUNEDU, national institutions tax database - SUNAT) and other bibliographic sources (Scopus, Web of Science, and Latin American SciELO collections).
4. A new version of the national profile CV system is being developed as its own DSpace community,
5. PeruCRIS Platform goes beyond mere aggregation, supporting data normalization, enrichment and curation for collected data, and direct editing in the Directorios.

The design and development of the platform is based on DSpace-CRIS and carried out by 4Science, <https://www.4science.it/>.

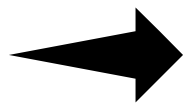
Analytics components

OpenSearch is a community-driven, **open source search and analytics suite** derived from Apache 2.0 licensed Elasticsearch 7.10.2 & Kibana 7.10.2

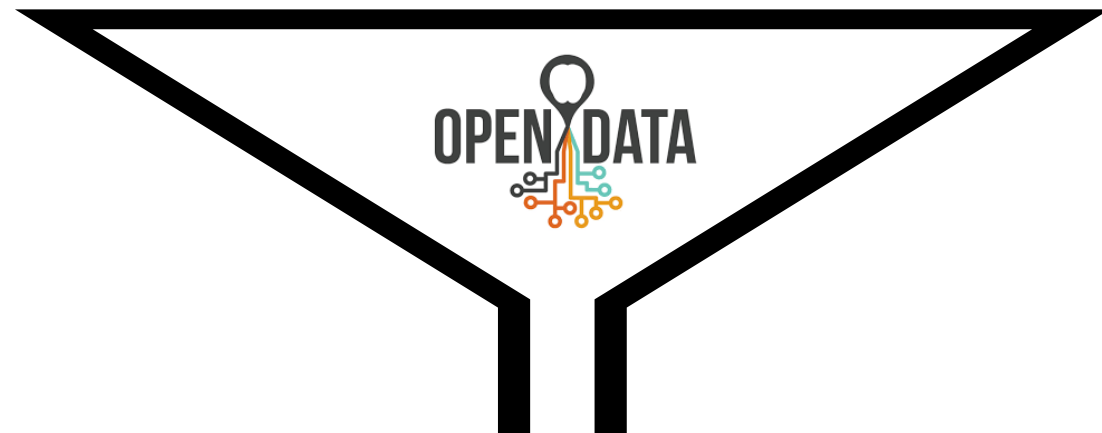
Dremio is a **data lake engine** that creates a semantic layer and supports interactive queries.

Apache Superset is a modern **data exploration and visualization** platform

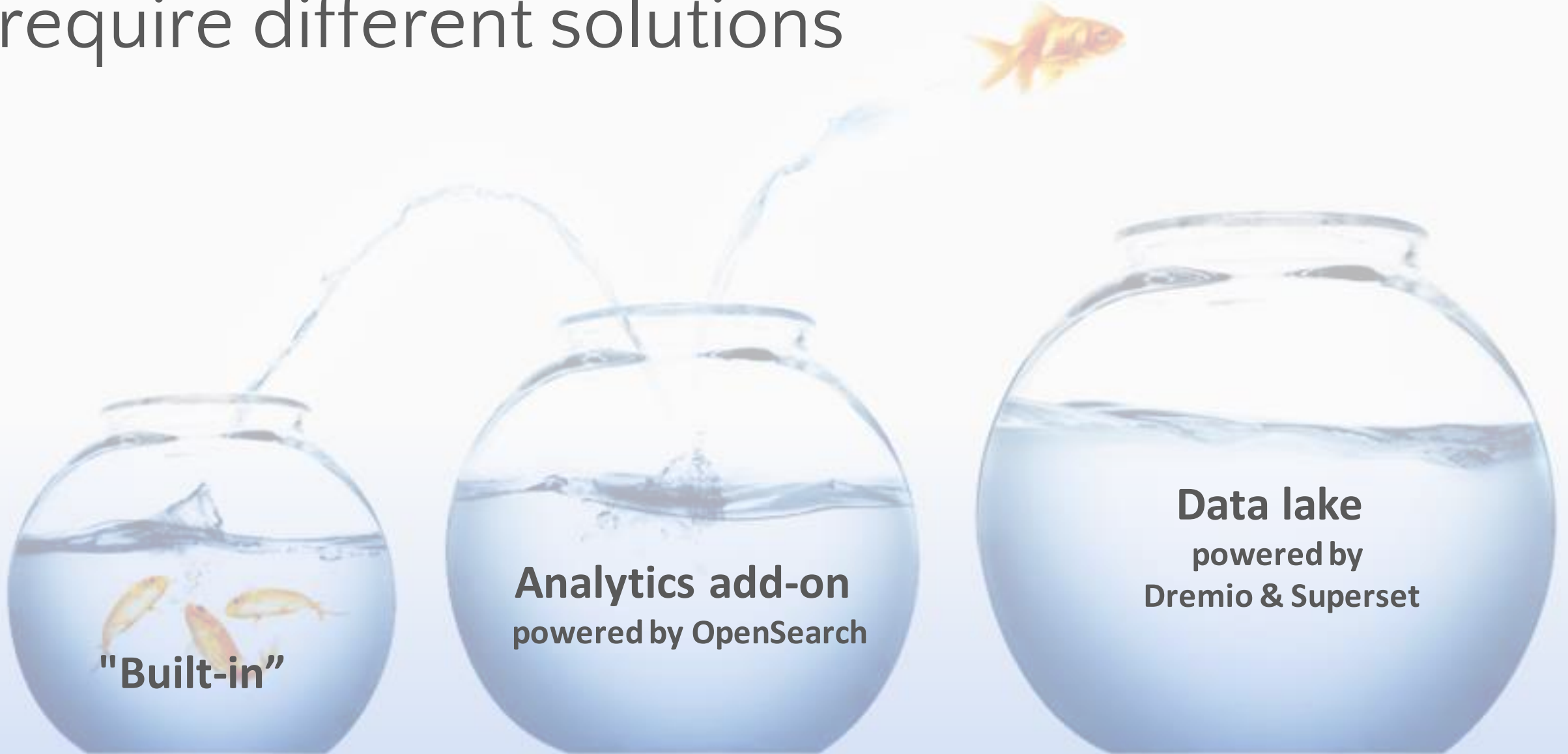




"former"



Different users & project scales require different solutions



"Built-in"

Analytics add-on
powered by OpenSearch

Data lake
powered by
Dremio & Superset

Built-in level

DSpace-CRIS supports a flexible search engine that allows data analysis and exploration. Results can be exported in configurable formats including CSV, XLS, PDF. Aggregation can be used to narrow the analysis (faceted browsing) and provide basic visualization

CONCYTEC **Open CRIS** SCIENTIFIC PRODUCTION PROJECTS HUMAN TALENT INSTITUTIONS INFRASTRUCTURE Log In

HOME / SEARCH

Research Outputs

Visualization by Type Visualization by Date **OCDE Knowledge Area**

ciencias n.
ingeniería.
ciencias a...
ciencias m...
ciencias s...
psicología
pediatría
matemática...
ingeniería...
ingeniería...
anatomía...
humanidade...

Search the repository ... Search

Filters

- Access Condition +
- Type +
- OCDE knowledge area +
- Date +
- Publisher +

Publications Search Results

Now showing 1 - 10 of 183

Patent
IQ Patente RD SEIS 27/12/2021
(2022-12-27) IQ APELLIDO RD UNO 27/12/2021. IQ NOMBRE RD 27/12/2021
Resumen Introduzca la descripción de la patente. Resumen Introduzca la descripción de la patente. Resumen Introduzca la descripción de la patente. Resumen Introduzca la descripción de la patente.

Publication
IQ PUBLICACIÓN NCS
(FDITORIAL) PI ANFTA 2022-12-20 IIO NOFI IA CATSII I O. NOFI IA

Communities & Collections Research Outputs Projects People Log In

Home • CRIS • Person • **Bollini, Andrea** Export

Bollini, Andrea

Preferred name Bollini, Andrea
Main Affiliation 4Science
Web Site <https://www.linkedin.com/in/andreabollini/>
<https://github.com/abollini>
andrea.bollini@4science.it
Email
ORCID [0000-0003-0864-8867](https://orcid.org/0000-0003-0864-8867)

Research outputs Projects Metrics Other

Research Output

Visualization by Type Visualization by Date Visualization Demo LTR Visualization Demo RTL

2021 2020 2019 2018 2017 2016 2015 2014 2013 2012

Now showing 1 - 10 of 35

Filters

Publication
The road (and the roadmap) for building Trusted Digital Repositories within an Interuniversity Consortium
(2018) Cortese, Claudio, Bollini, Andrea

Editor
3

Organization
The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Pfizer-BioNTech COVID-19 Vaccine - United States, December 2020
(2020-12-18) Bollini, Andrea, Lombardi, Corrado

Subject
[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7648883/](#)



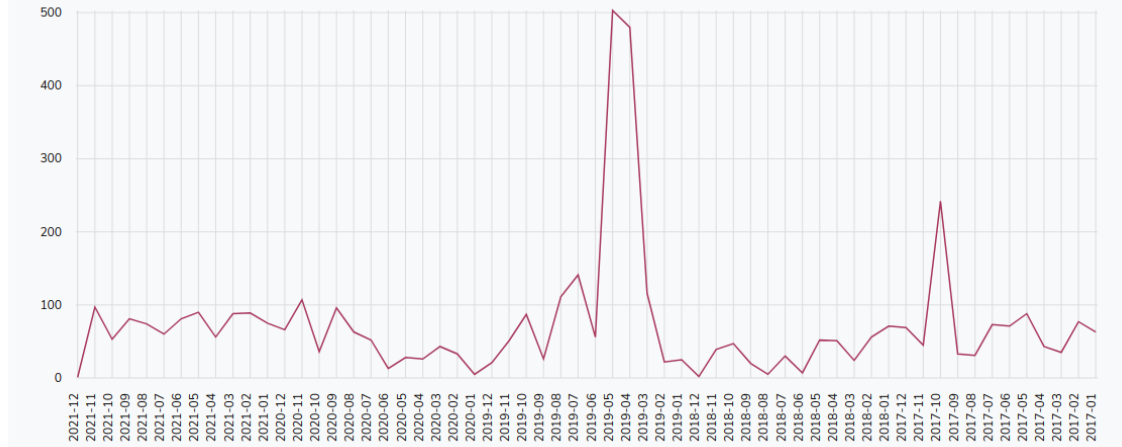
It is possible to configure the graphical visualization (bar charts, pie charts, lines) with respect to any dimension of aggregation for searchable or predetermined lists (projects of a department, of a researcher, of a certain area, etc). The visualization is interactive, and it is possible to progressively shift the focus using both graphical and numerical facets, and search queries.



Data can be extracted according to user privileges

Search Output

Visualization by Item upload



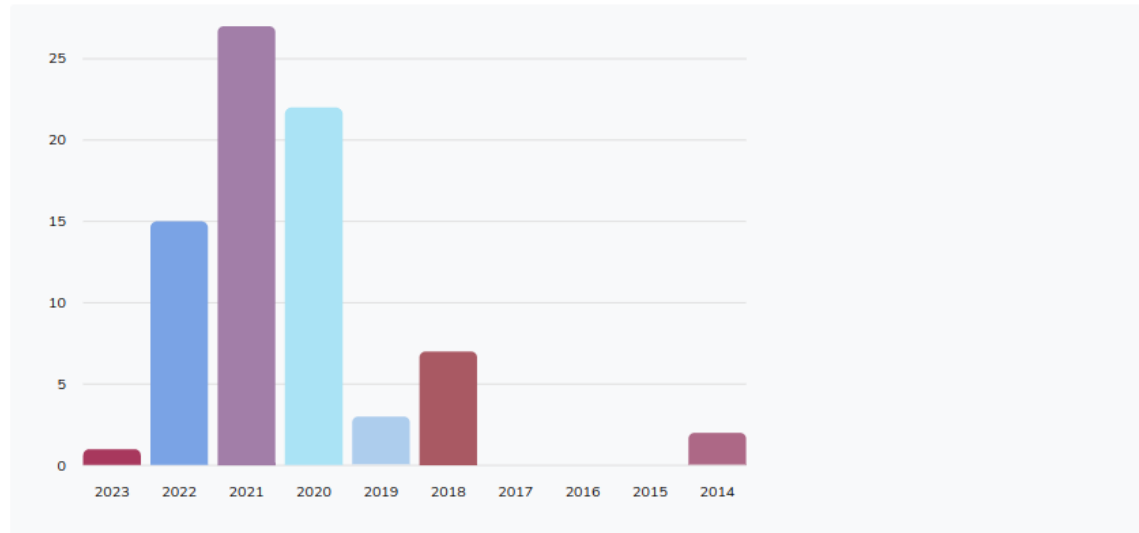
Search Output

Visualization by Type

Visualization by Date

Visualization Demo LTR

Visualization Demo RTL



Item Export Process Launcher

Export all items of a particular entity type belonging to the current search.

Entity Type:

Format:

- Choose the export format
- publication-chicago
- publication-xls**
- publication-cerif-xml
- publication-json
- publication-csv
- publication-apa

Analytics Add-on module

4Science engineered and generalized the solution used in the PeruCRIS project and offers it to institutions for a share of the initial cost of design. 4Science engineers will support the institution to configure the add-on module to meet the institution specific needs, extracting the most from data. It provides self-service analytics capabilities, rows and columns level security, automatic periodic generation of reports. During the ingest, data are de-structured to provide easy analysis from different perspectives. Data can be accessed via JDBC / ODBC as they are in a traditional database ANSI SQL, allowing further reuse in external tools such as PowerBI, Qlik Sense, Tableau or even Excel.

Dashboards

+ Create dashboard

Search...

<input type="checkbox"/> Title	Description	Actions
<input type="checkbox"/> OrgUnits		
<input type="checkbox"/> Person		
<input type="checkbox"/> Projects		
<input type="checkbox"/> Publications		
<input type="checkbox"/> Publications - single contribution view		
<input type="checkbox"/> Publications - current affiliation view		

Rows per pag

Dashboards

+ Create dashboard

Publications

<input type="checkbox"/> Title	Description	Actions
<input type="checkbox"/> Publications		
<input type="checkbox"/> Publications - single contribution view		
<input type="checkbox"/> Publications - current affiliation view		

Rows per page: 20

< 1 >

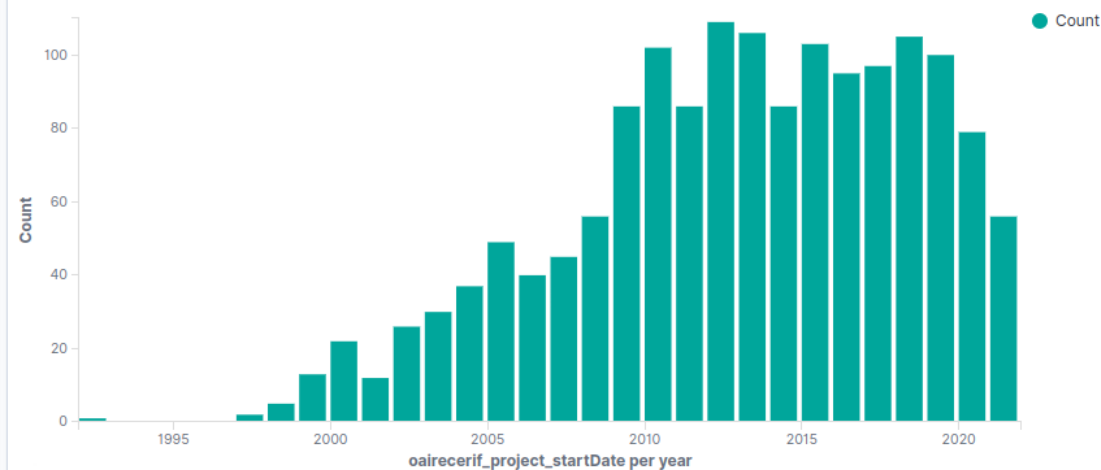
Projects dashboard

+ Add filter

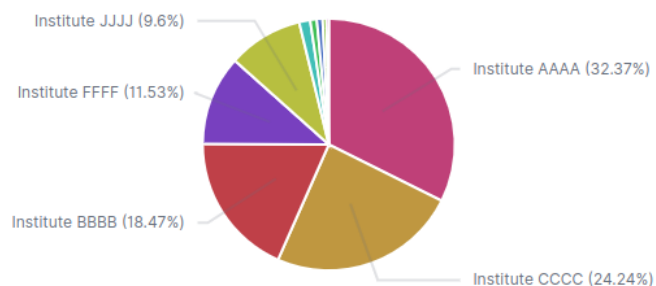
Projects count

1,558
Count

Projects by Start Year

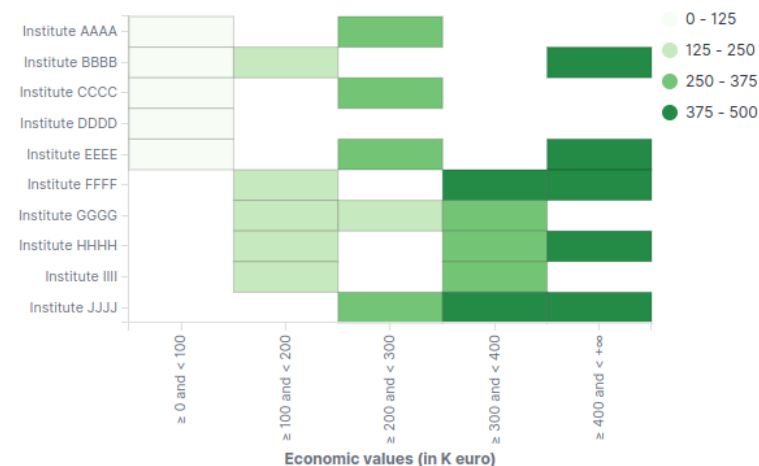


Projects by Dept

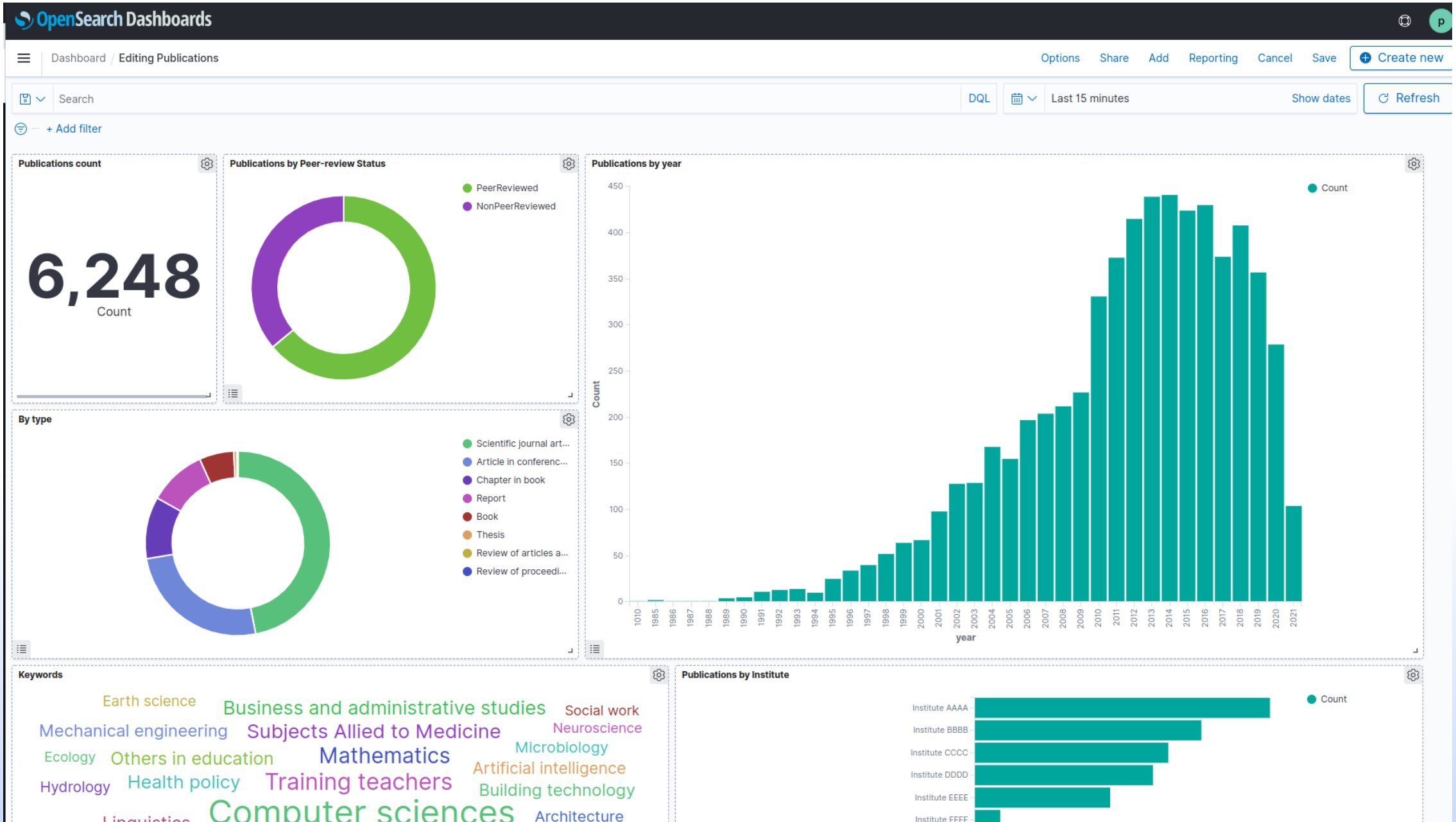


- Institute AAAA
- Institute BBBB
- Institute CCCC
- Institute DDDD
- Institute EEEE
- Institute FFFF
- Institute GGGG
- Institute HHHH
- Institute JJJJ

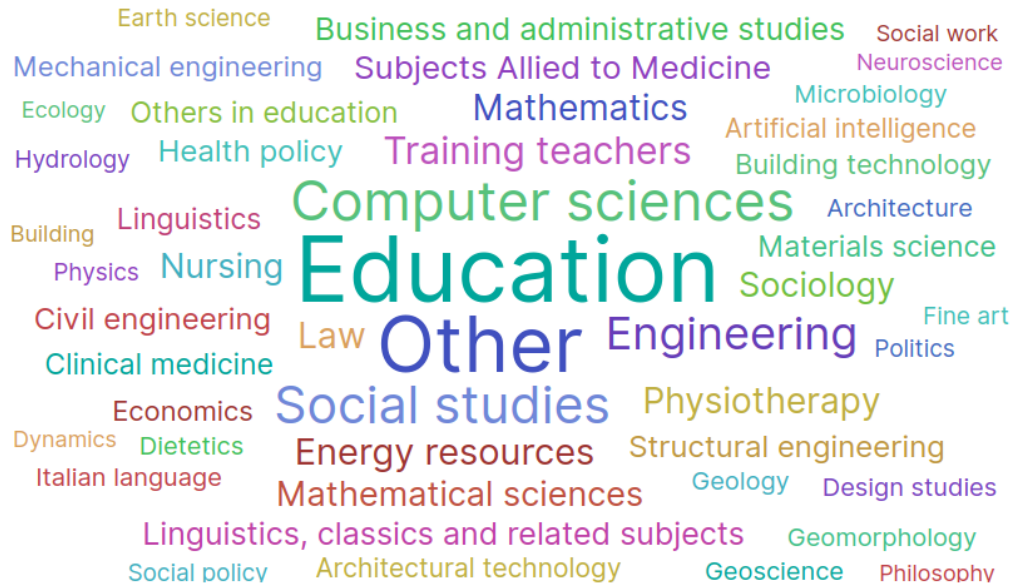
Projects Dept Heat Map



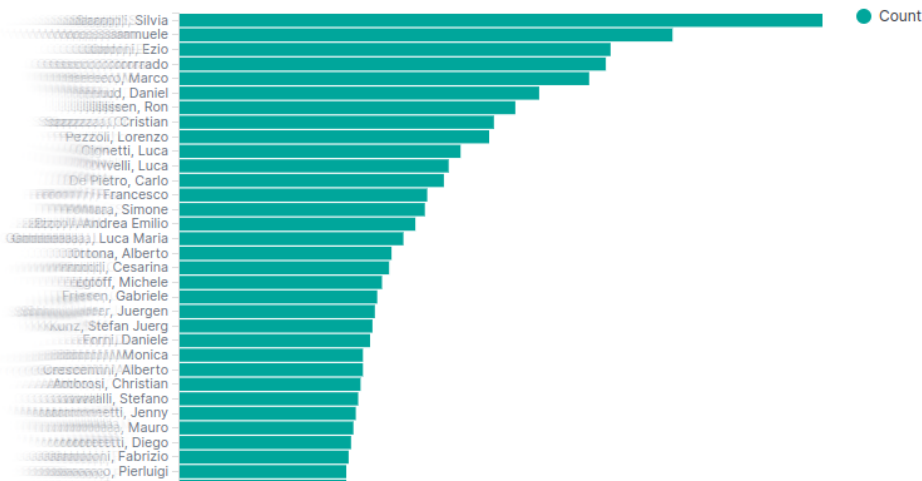
Publications dashboard



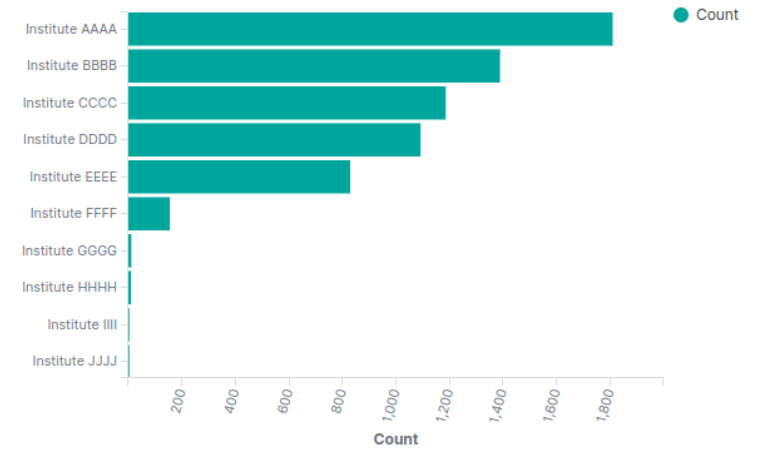
Keywords



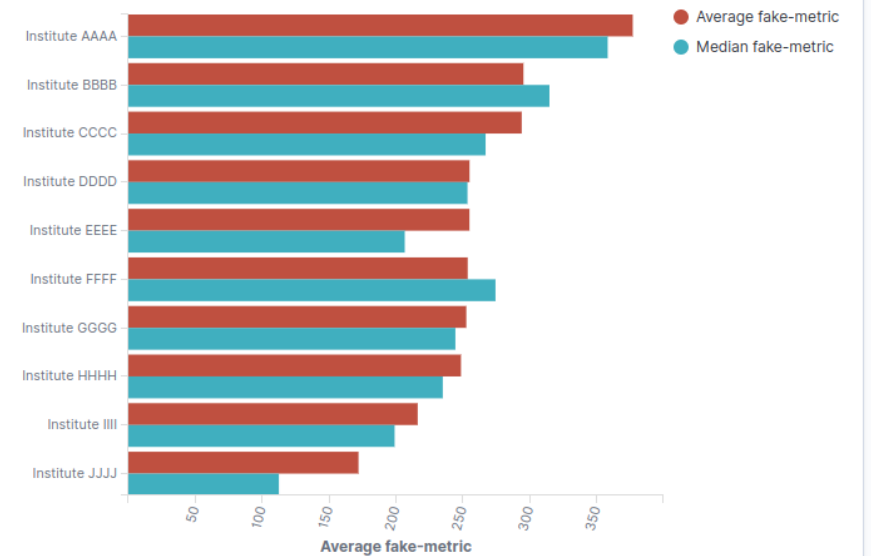
Top authors



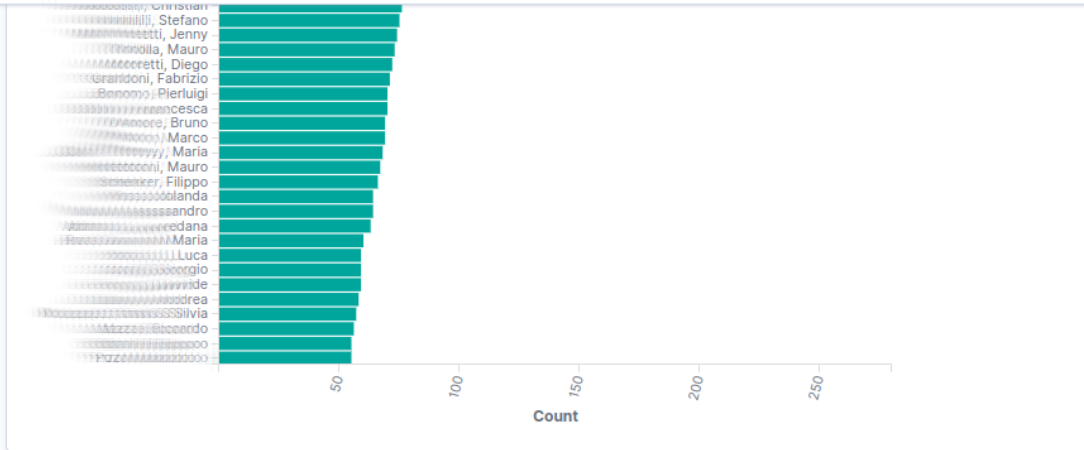
Publications by Institute



Publications performance by Dept



HeatMap publication by dept / metric

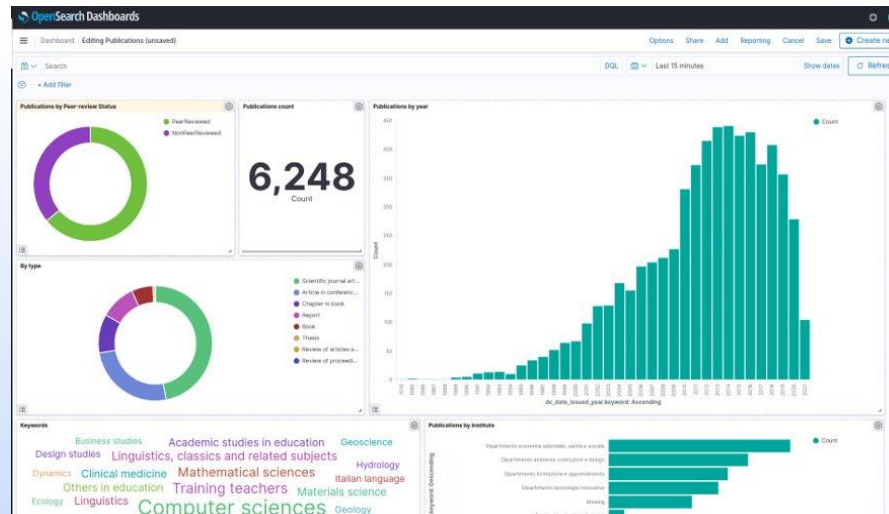
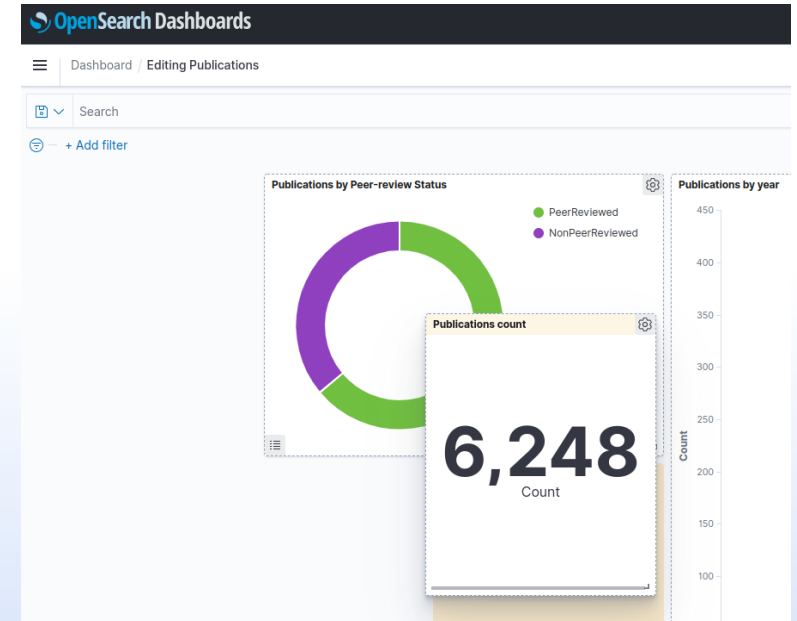
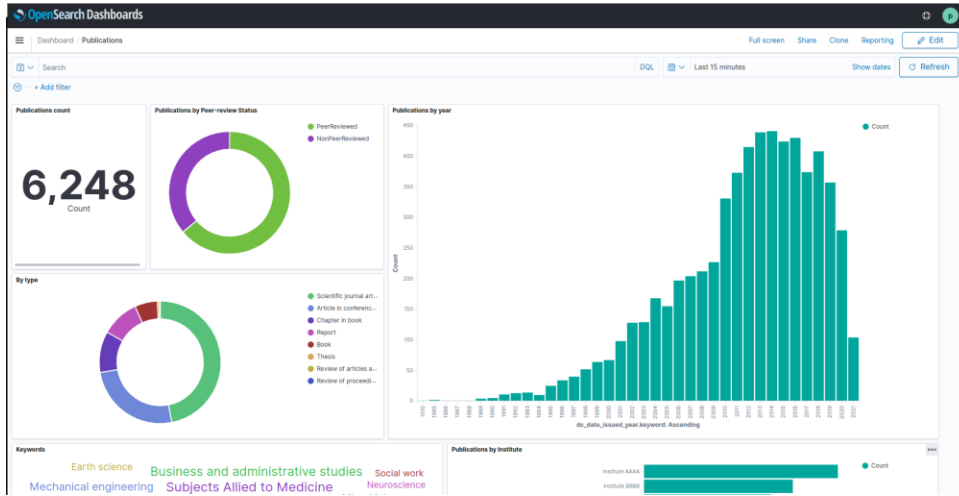
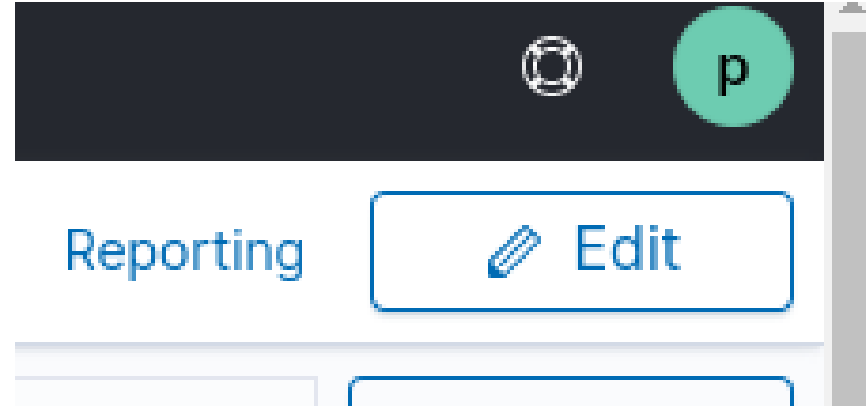


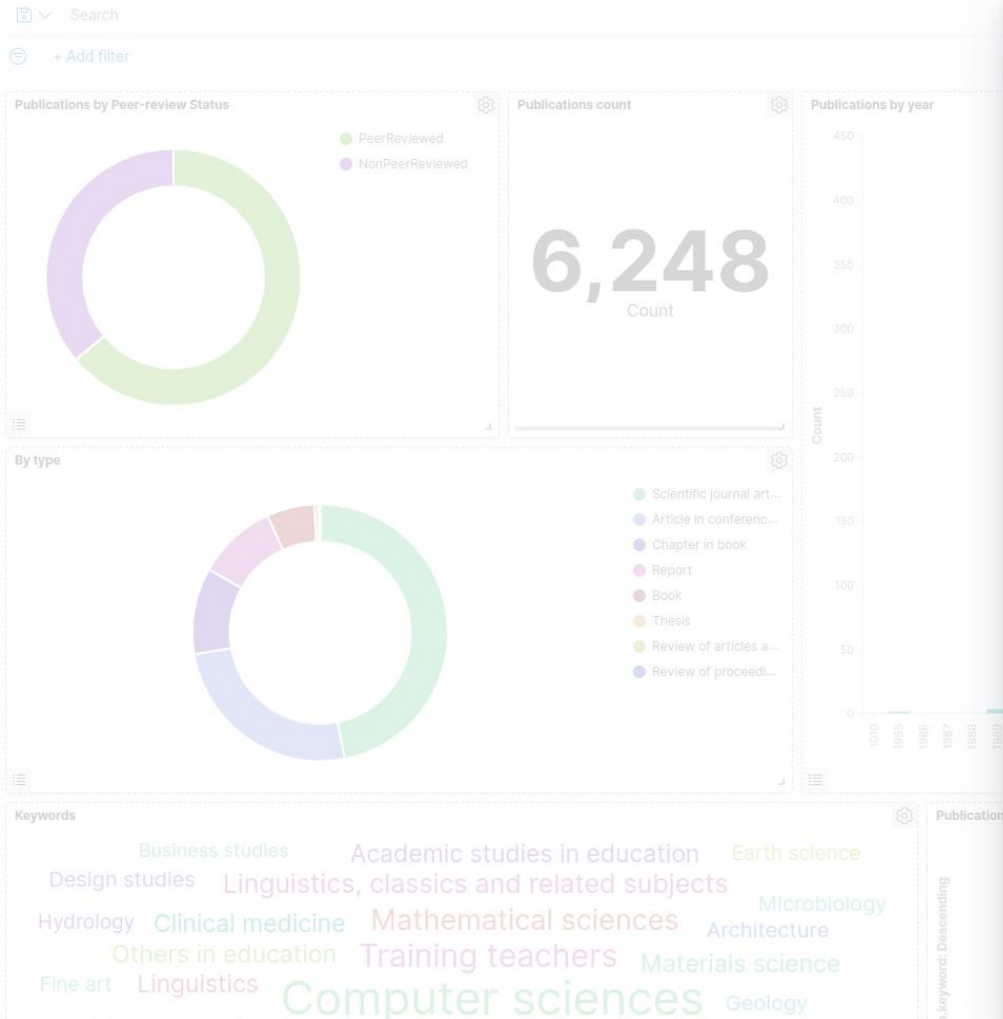
Matching publications

1-50 of 6248 < >

Year	Title	Authors	Type	Subjects
> 2016	Global life satisfaction in adolescence: The role of personality traits, self-esteem and self-efficacy	Marcionetti, Jenny, Rossier, Jérôme	Scientific journal article, Newspaper article or Magazine article	Social studies, Others in social studies, Education, Others in education
> 2006	La Balanced Scorecard tra finalità e contesti eterogenei nel confronto tra alcune esperienze internazionali	Ferrari, Domenico, Greco, Angela, Merlini, Luca	Chapter in book	Management studies not elsewhere classified

Self-service capabilities





Add panels

Search... Sort Types 2 **+ Create new**

- By type
- Departments and Institutions
- HeatMap publication by dept / metric
- Institute performances by Department**
- Institutes by Department
- Keywords
- Matching publications
- Person
- Person by Current Affiliation
- Projects by Dept

Rows per page: 10

New Visualization

Filter

Select a visualization type

Start creating your visualization by selecting a type for that visualization.

- Area
- Controls
- Coordinate Map
- Data Table
- Gantt Chart
- Gauge
- Goal
- Heat Map
- Horizontal Bar
- Line
- Markdown
- Metric
- Pie
- Region Map
- TSVB
- Tag Cloud
- Timeline
- Vega
- Vertical Bar

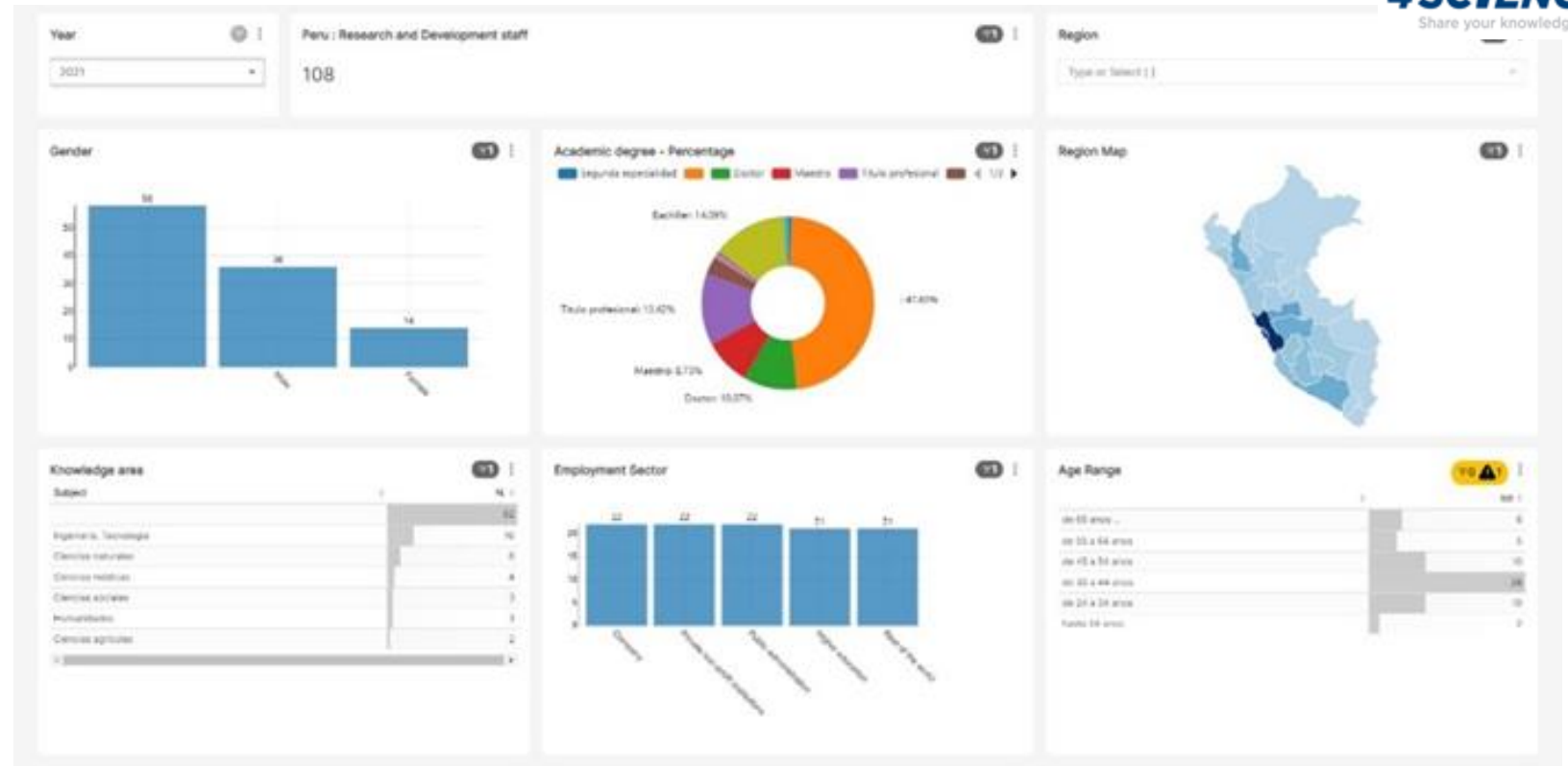
Data Lake

Dremio provides the bridge from the "known institutional" data available in DSpace-CRIS via the Analytics add-on, and other external sources (database, SPARQL endpoints, xls files, S3, etc.). The data lake can be visualized and explored via the integrated Apache Superset webapp or external tool, accessing it as a virtual traditional SQL database



dremio

Data in the national DSpace-CRIS installation are dynamically joined with external demographic sources on Dremio to provide a graphical visualization over a map of research activities



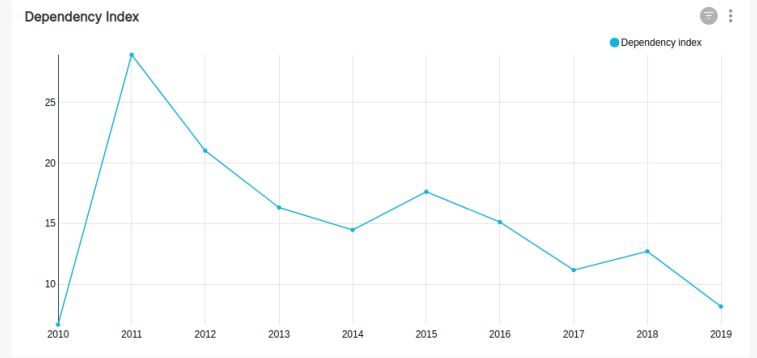
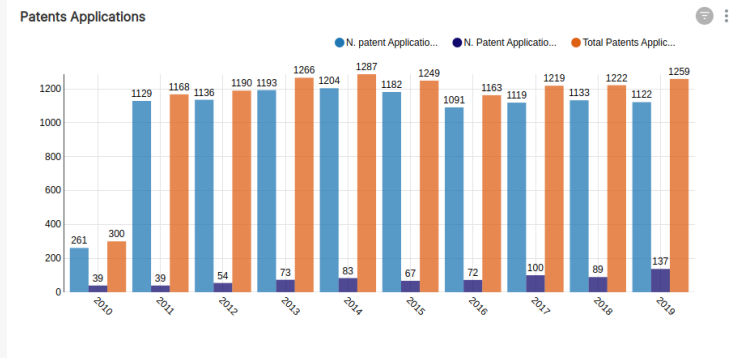
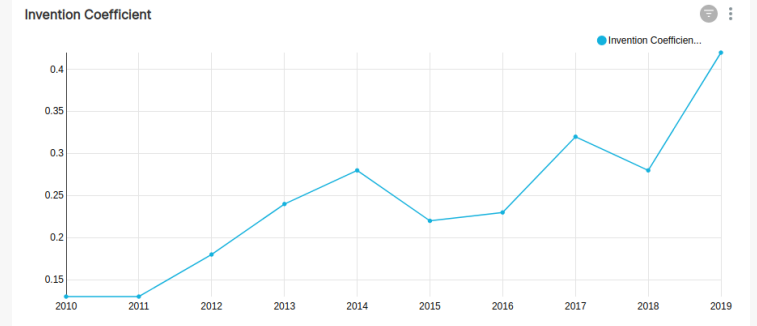
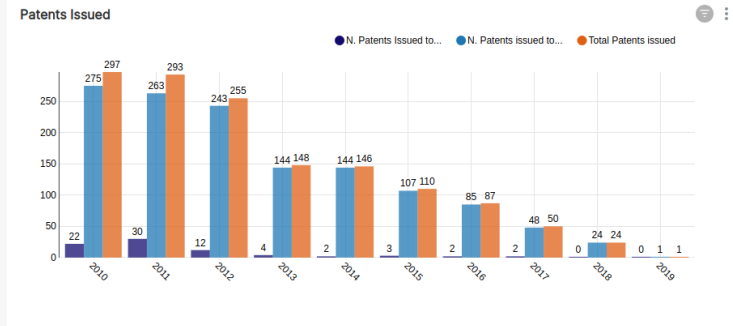
Superset provides a self-service environment with almost the same features as OpenSearch, but independent from a specific data source. In our case we connect it to the "DREMIO" virtual database


Patents dashboard

Dashboard - Patent Index Published

Year

Type or Select []





Thanks for your attention!

Andrea Bollini, 4Science
andrea.bollini@4science.it
Praxistreffen 2022