



Semantically Mapping Science (SMS) Platform

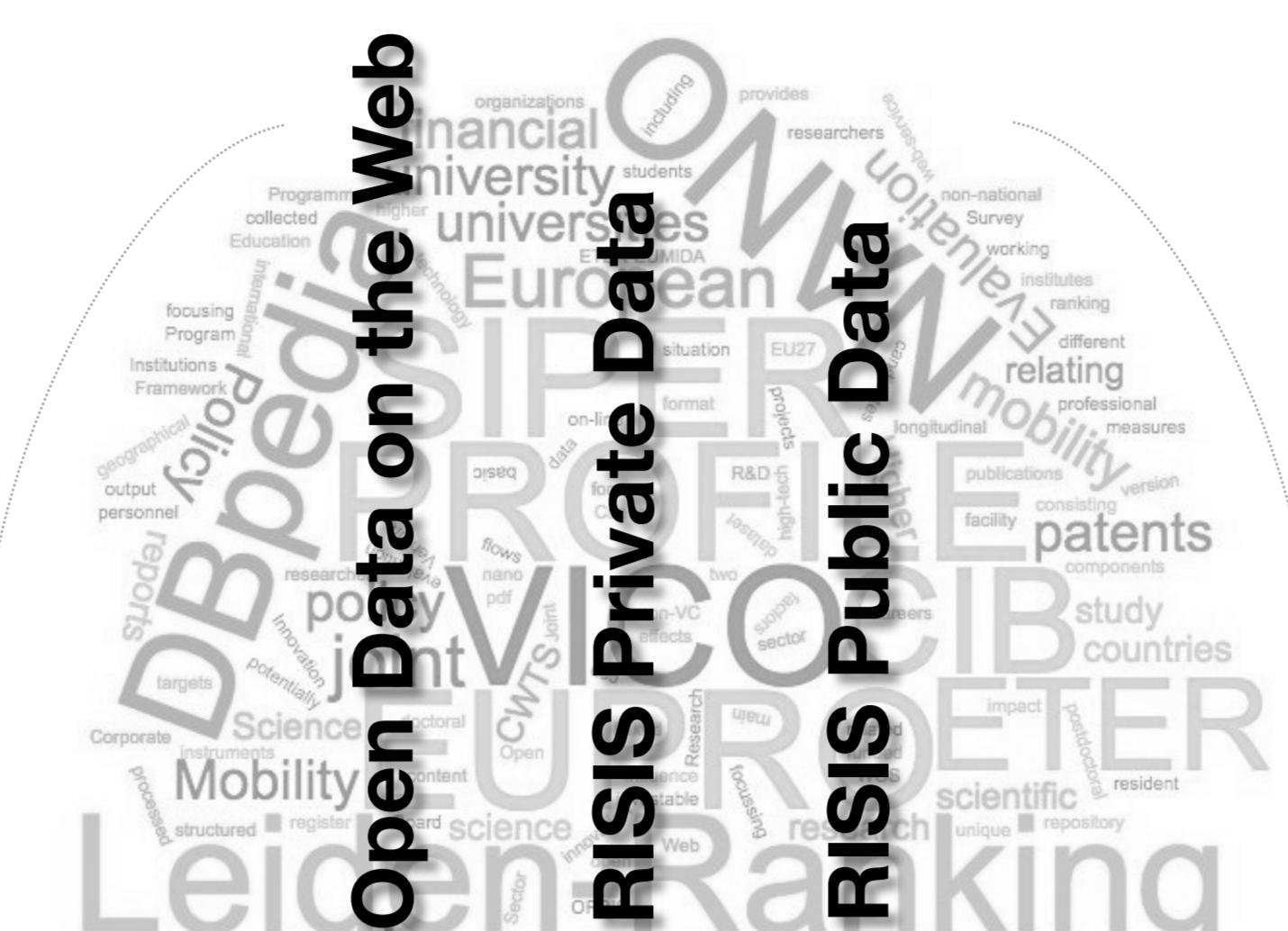
Towards an Open Infrastructure for Studying Science, Technology and Innovation



Peter van den Besselaar[#], Ali Khalili*, Klaas Andries de Graaf*, Al Idrissou[#], Antonis Loizou*, Stefan Schlobach* and Frank van Harmelen*

Vrije Universiteit Amsterdam, The Network Institute,

Department of Organization Sciences, * Department of Computer Science



RISIS
Research infrastructure for research
and innovation policy studies

Access Control Points (ACPs)

- provide standard interfaces to reduce technical difficulties of accessing data.
- provide a mechanism to coordinate access to data based on the user role and the datasets owner's requirements.

Scientific Lenses

When comparing two entities, depending on the user's perspective and the context of study, they might be considered the same, similar or different. Sometimes two organizations (e.g. departments) can be the same – because they are parts of the same organization (university). But if one wants to compare departments, this is not the case.

Scientific lenses support linking the entities based on the context of use and provided features of data.



Linked Data Services

expose the functionality of the SMS platform to third-party users by standard Application Programming Interfaces (APIs).

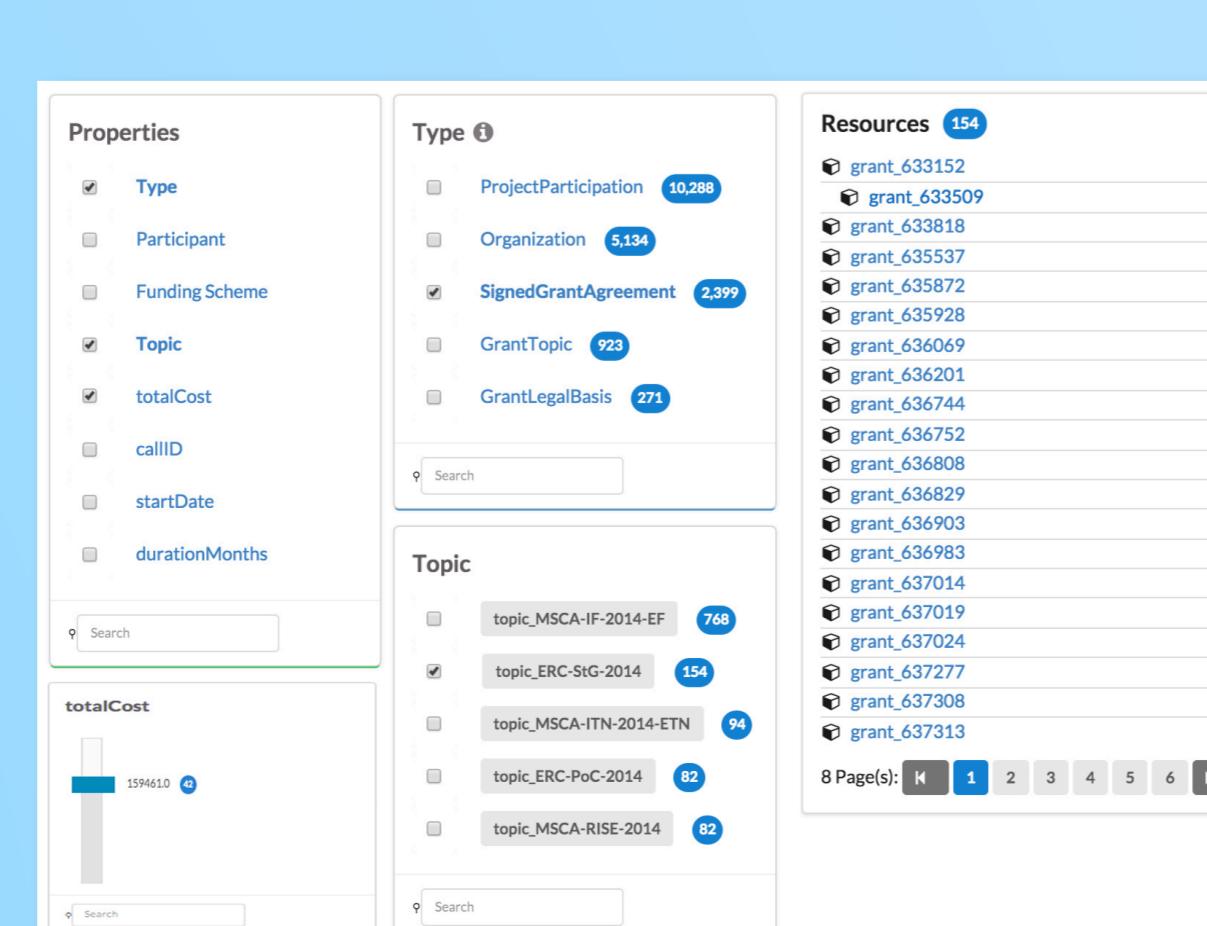
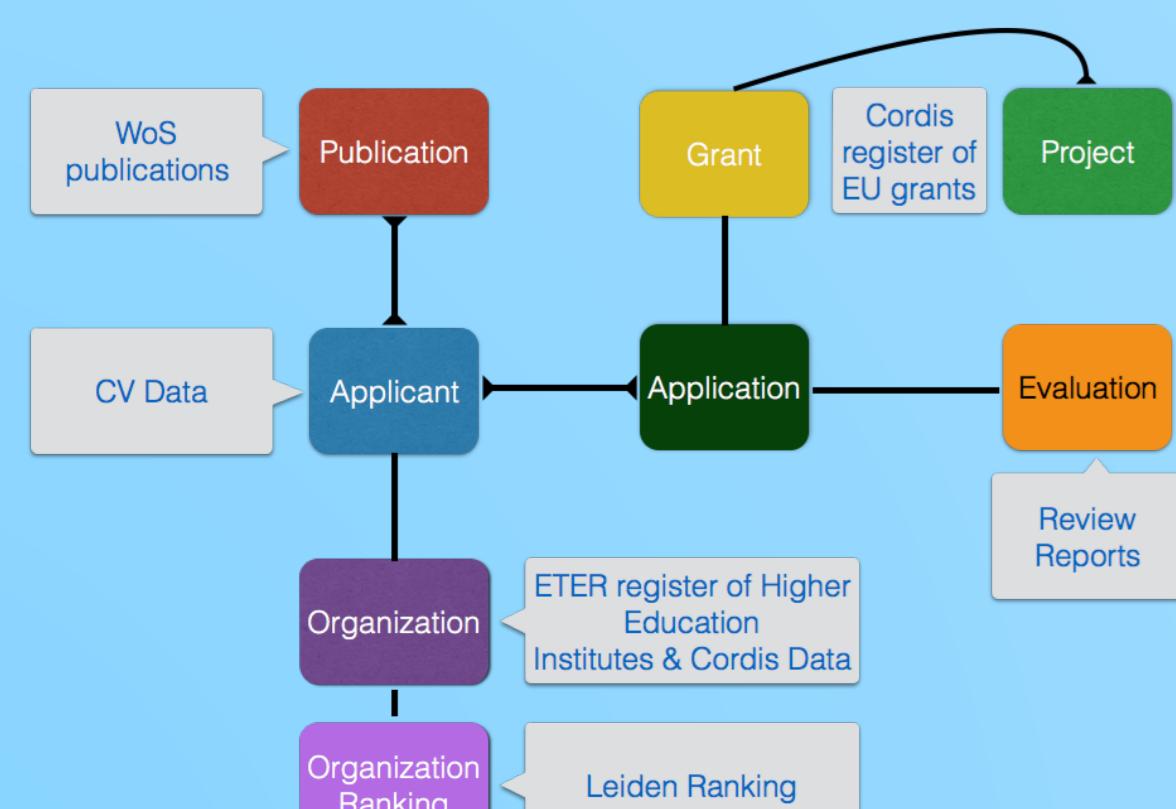
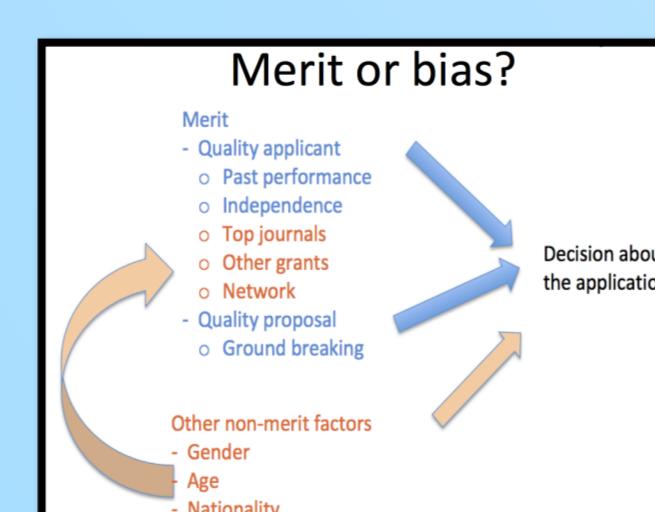
- Identity Resolution Services
- Named Entity Recognition Services
- Metadata Services
- Category Services
- Basic Geo Services
- Innovative Geo Services
- Integration Services

allow integration of data with RISIS public and private datasets as well as open datasets and social data available on the Web.

Gendered Dimensions in Grant Selection

In this use case, we investigated whether gender of applicants influences the grant decision. In order to answer that question, one needs a multitude of variables that may influence the grant decision - as suggested by various theories: variables representing merit (such as scholarly performance), but also variables bring in bias, such as personal characteristics (gender, age, nationality, university of PhD degree). And some in-between variables, such as the quality of the applicant's network. We use the SMS platform for preprocessing, for converting into RDF, for entity recognition, and for linking data. These variables come from a variety of data sources:

- Bibliometric performance scores: Web of Science (TXT)
- Quality of the applicant's network:
 - Organizations mentioned in the CV (PDF)
 - Ranking of those organizations from Leiden Ranking (Excel)
- Earlier grants: from CV (PDF)
- Host institution from admin file (Excel)
- Other available open datasets

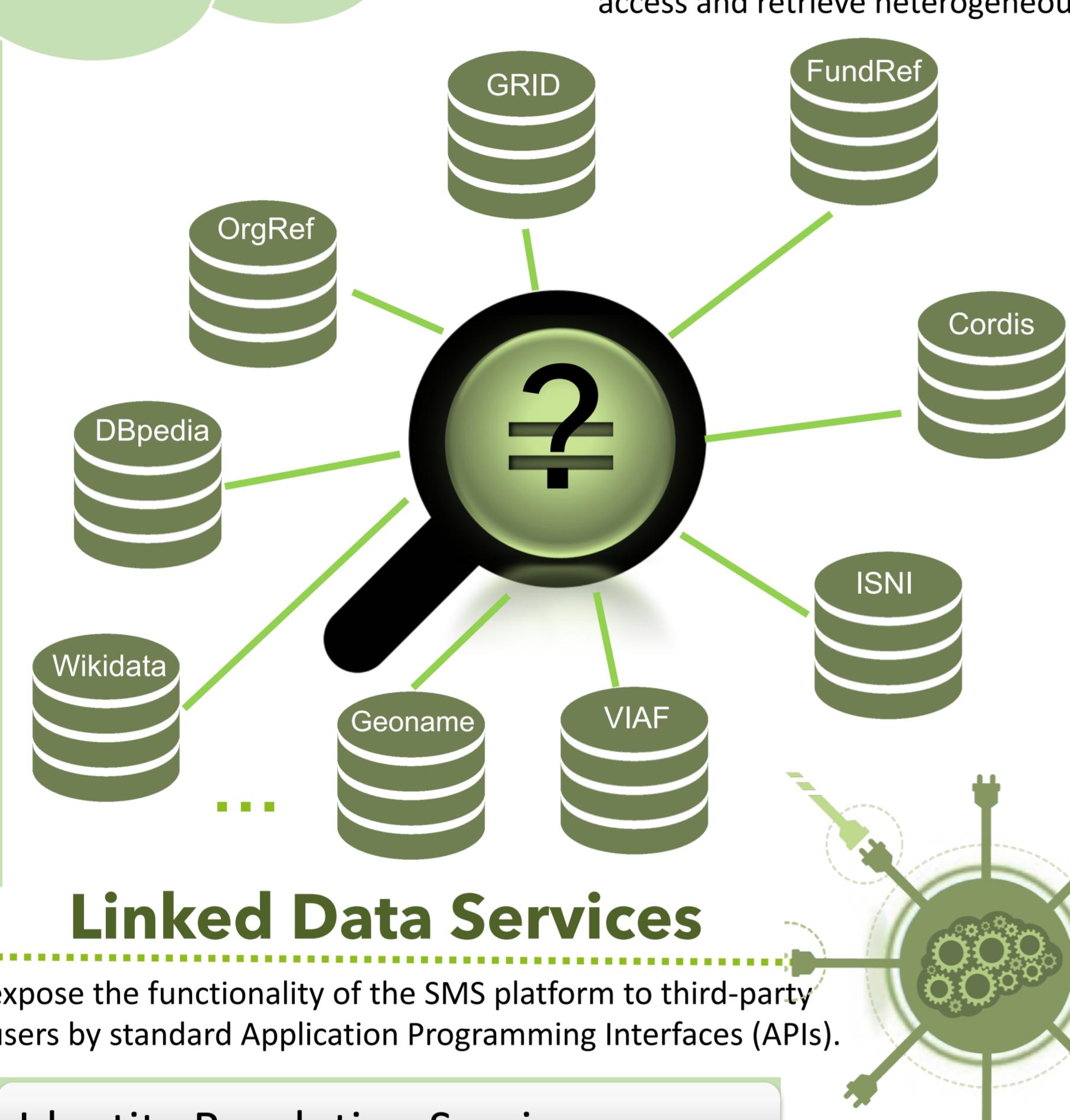


Browsing Cordis H2020 Research Grants

This article is licensed under a Creative Commons Attribution 4.0 International License

Data Ingestion

access and retrieve heterogeneous data.



Linked Data Creation

convert unstructured and structured data to RDF.

a) Dataset Metadata Editor

The RISIS dataset holders all have to describe their datasets in a detailed, consistent, and uniform way.

To achieve this goal, RDF data model is used to describe the RISIS datasets. To stimulate non-Semantic Web users to generate valid RDF metadata descriptions, we designed a novel user-friendly editor which hides the complexity of RDF from non-technical users. The metadata editor exploits the state of the art Web technologies to provide user-friendly component to view and edit the metadata.

b) RISIS Datasets Portal

In order to exploit the generated metadata, RISIS datasets portal brings a user interface to view and browse the metadata. Faceted browsing allows users to explore the dataset via multiple entry points, or when users do not know what they are looking for beforehand. The portal also handles user registration and supports the process of reviewing visit/access requests to certain RISIS datasets.

Applications

combine one or more SMS services with other existing services and applications to build novel and innovative applications.

c) Spreadsheet add-on

In order to enable batch processing of data, SMS provides a Google spreadsheet add-on. This add-on allows users to enrich their data directly in their spreadsheets.

d) Address to Boundaries

Given an address, this application returns the geographical boundaries from different open datasets. The results are displayed on a map.

e) Faceted browser

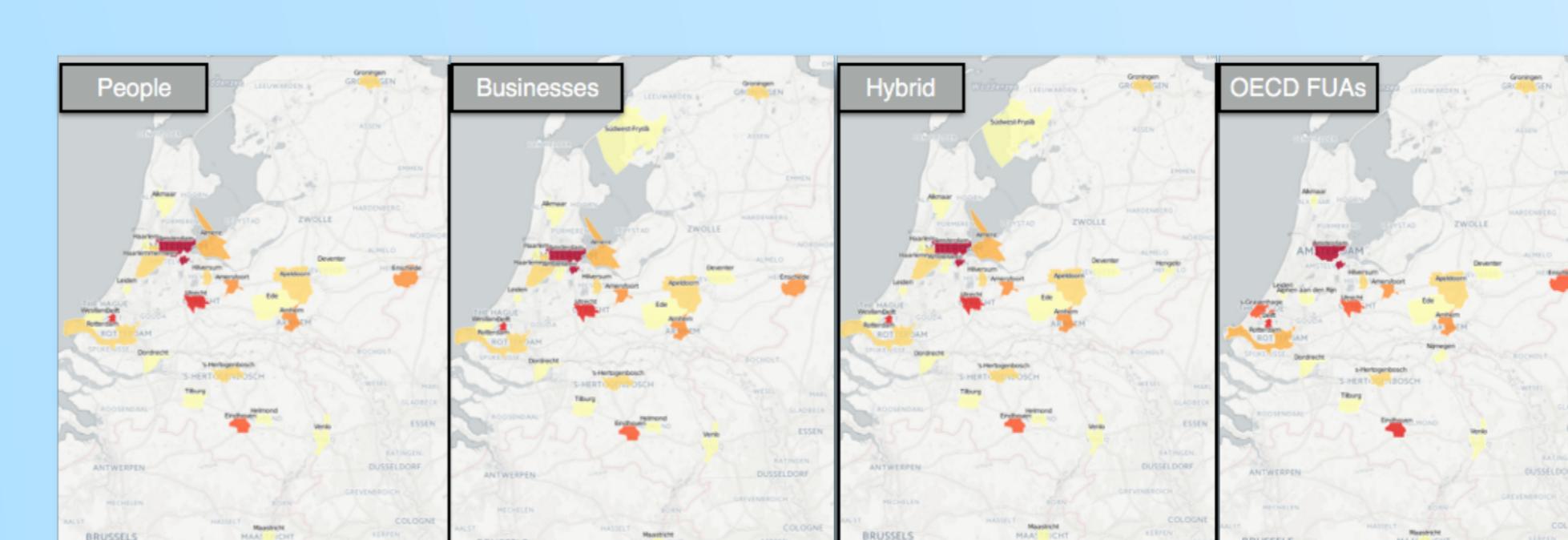
Allows browsing a dataset using a set of pre-defined facets

a
b
c
d
e

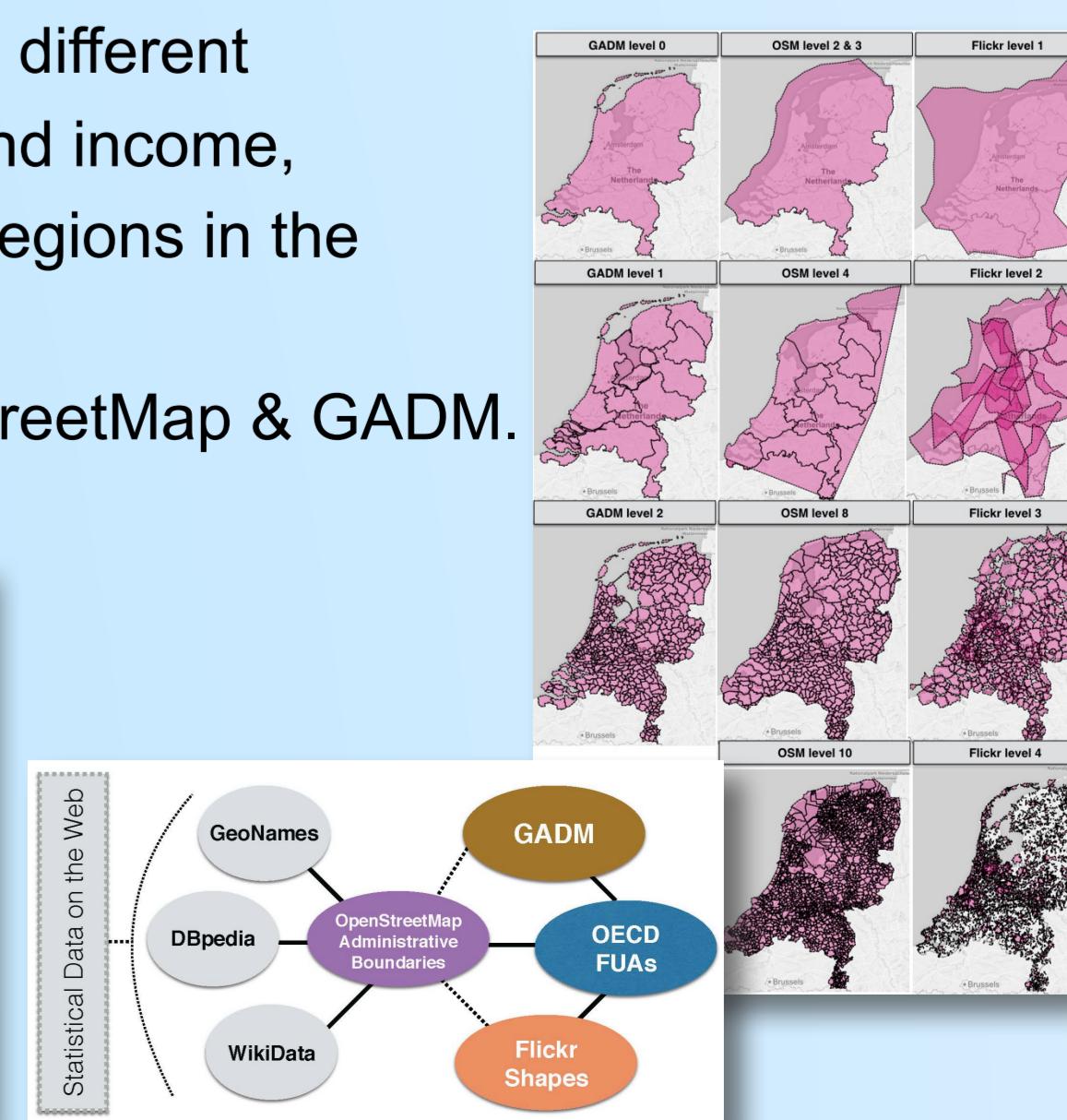
Adaptive Functional Urban Areas (FUAs) to Study Innovative Activities

In this use case, we investigated the effects of socio-economic and structural properties of urban areas on the level of innovative activities, as stimulated by recent research and innovation policies in the Netherlands. The funded projects can be considered as a useful representation of RTD (Research Technology Development) collaboration for innovation. In this case we are interested in the geographical properties of these projects. In order to investigate this, one needs data about the projects and data about the characteristics of the relevant geographical units. These data are available as open data. In this case, the following open datasets are deployed:

- RVO dataset providing a list of research and innovation projects that have received subsidies and financial support from the Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland). Projects information includes companies and research institutes that are collaborating in the project, but also the geographical coordinates of the project main applicant.
- CBS dataset published by the Statistics Netherlands provides different types of statistical information on dimensions such as labor and income, economy, society and regional aspects of municipalities and regions in the Netherlands.
- Open Datasets on administrative boundaries such as OpenStreetMap & GADM.



Amount of RVO project subsidies mapped to the dynamically delineated FUAs defined based on the CBS open statistical data and OpenStreetMap boundaries.



Integrating multiple open datasets about geographical boundaries with existing statistical data on the Web.