

# REDI: A Linked Data-powered Research Networking Platform

Xavier Sumba<sup>1</sup>, José Segarra<sup>1</sup>, José Ortiz<sup>1</sup>, Boris Villazón-Terrazas<sup>2</sup>, Mauricio Espinoza<sup>1</sup>, and Víctor Saquicela<sup>1</sup>

<sup>1</sup> Computer Science Department, University of Cuenca, Ecuador

<sup>2</sup> Fujitsu Laboratories of Europe, Madrid, Spain

**Abstract.** Research networking is a difficult part of academics in spite of the multiple benefits that the Web has brought within this field in recent years. Even though scientific and business social networks provide a medium to discover peers worldwide, their usefulness meets its limits when real-world requirements come in. The broad audience of those tools and other bibliographic databases lead them to ignore cultural and geographical aspects such regional indexes, organizational structures, among others. On this poster, we introduce REDI, a Linked Data - powered research networking platform which combines both local (institutional/regional) and external (Web) scholarly sources in a consolidated knowledge base. Moreover, REDI leverages on its knowledge base to cluster authors within similar research areas easing networking and unveiling a variety of new information from data for multiple purposes.

**Keywords:** Data Integration, Linked Data, Data Mining, Research Networking

## 1 Introduction

The web not only has opened up a new whole world of opportunities to share scientific work but has eradicated the barrier of communication. However, there is still an open door for the process of research networking. Senior researchers usually create their collaborative networks through the years. Technology and tools such as ResearchGate have simplified this process but are still limited. Indeed, young or novice researchers need a mechanism to locate possible collaborators. For example, some grants are allocated in an inter-institutional approach, so there is no tool to find potential collaborators in a certain area or institution. Filling that gap in social research networks will ease the research community the creation of cross-cultural and interdisciplinary projects.

In addition, the availability of scholarly literature on the web has given access to researchers a considerable amount of articles or other digital academic resources. This extremely large amount of scientific publications turns the process of finding an article into a challenging task due to the overwhelming results of well-known search engines or bibliographic repositories. As a consequence of

the variety of bibliographic repositories, articles are spread out in different digital repositories, text files, or bibliographic databases, and researchers have to search in a variety of sources.

Further, in some cases, institutions are regulated by a control system. This control system is carried out for external or internal institutions in order to approach a quality assessment of education. For this evaluation, institutions need to generate research-based indicators. Digital repositories provide some indicators, but they target a specific index (e.g. Scopus). On the other hand, bibliographic databases (e.g. Google Scholar) cover a broad range of indexed literature, and as a trade-off of their broad audience, they cannot provide specific indicators for institutions. As a result of the lack of useful features for decision making, institutions need to get indicators manually.

In this poster, we present and describe REDI a research networking platform, outlining its components and main functionalities.

## 2 What is REDI?

REDI<sup>3</sup> is an open source platform<sup>4</sup> for scholarly information that is built on top of Apache Marmotta<sup>5</sup>. It allows storing information of authors along with their publications in a *Linked Data* approach. In addition, it supports the discovery of similar areas of knowledge and potential collaboration networks. Finally, it has a web application to navigate information in the repository and generate statistics and reports. The main modules that make REDI are illustrated in Fig. 1. For a detailed description of each component, we refer to [1].

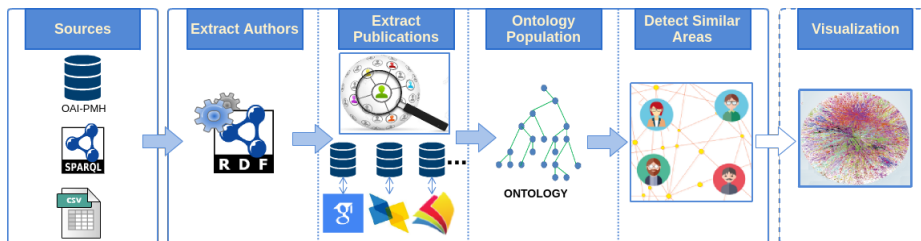


Fig. 1. Main components of the REDI platform.

**Authors Sources.** As a starting point, there is the need to register a candidate researchers source. These authors will be used for the population of the platform. Researchers can be imported through a *CSV file*, *SPARQL Endpoint*, or *DSpace repository*.

<sup>3</sup> REDI is a spanish acronym, “Repositorio Ecuatoriano de Investigadores”, and stands for the project name where the idea was born (<https://redi.cedia.edu.ec/>).

<sup>4</sup> <https://github.com/ucuenca/redi>

<sup>5</sup> <http://marmotta.apache.org/>

**Authors Extraction.** Given an authors source, this module extracts authors information and transforms to *RDF* if necessary. The information is stored in the repository using FOAF<sup>6</sup> ontology.

**Publications Extraction.** Each researcher is assigned to his publications. Publications are extracted from digital repositories or bibliographic databases, at the time of writing the ones used are Academics Knowledge, DBLP, DSpace, Google Scholar, Scielo, Springer, and Scopus. Information of each provider is stored in its own graph of the triple-store platform using BIBO<sup>7</sup> ontology.

**Ontology Population.** Information of each provider and authors is consolidated in a *central graph*. It implies that information is verified; e.g. assigning correctly publications to authors, combine repeated information found in different providers, and disambiguate authors, publications, and co-authors. For a detailed explanation of the author disambiguation process refer to [2].

**Similar Areas of Knowledge.** Using the information of the central graph, we classify documents based on the UNESCO nomenclature in order to identify knowledge areas. This is done using graph mining, clustering, and DBpedia as a knowledge base, the process is similar to the work of [3]. Once identified those areas, we build potential collaboration networks.

**Visualization.** Finally, results are shown in a web application that allows exploring information about authors or publications. Most importantly, it suggests potential collaboration networks and similar knowledge areas that authors are working. It shows similar authors working in similar areas even if they have not co-authored before.

### 3 Platform Overview

REDI gives the facility to store research data of an institution or country. These data are relevant for researchers, regulatory agencies, government agencies, or the general public. The platform has a set of user-friendly interfaces in order to facilitate data accessing and its consumption. These interfaces show information that is collected and summarized, and it is presented to final users in a graphic and intuitive approach through mechanisms such as statistics, author graph-like navigation, authors or publications search, among others. Next, we highlight some features of the platform.

**Academic Knowledge Base** REDI collects a large amount of information related to the academic and scientific field. Information that might act as a source of interest for private and public institutions that deal with academic information. The available information is processed, stored and published following the Linked Data principles, thus facilitating the process of data sharing, as well as improving its interpretation as it has a defined ontology and allows linkage through repositories.

**Visualization of authors and their associated elements** The tool has the ability to search for authors and visualize their data. These data are collected

<sup>6</sup> <http://xmlns.com/foaf/spec/>

<sup>7</sup> <http://bibliontology.com/>

from different sources of information, so users can find interesting notions about data that they cannot find when exploring a single source. For instance, resources of scientific works of an author and co-authors can be expanded when exploring a graph, and probably discovering new information given that these associations are not always explicit.

**Knowledge Discovery** In the repository of researchers, it is possible to find new information that is not available in other repositories. For example, it is possible to identify the articles of publications that have Latindex index as detailed in [4]. This information is valuable, both for researchers and control agencies because it can help calculate the research impact, for example.

**Building potential collaboration networks** Another outstanding functionality of REDI is the identification of possible collaborative networks among researchers. Through this process, researchers can discover people who are working in similar areas of knowledge, or whom have common characteristics for the formulation of joint projects, or simply visualize the knowledge developed in a specific area of knowledge within a region. These networks can also be used by government agencies to visualize the research areas that have greater coverage within the country, and those that have few participants and need to be driven.

## 4 Conclusions

We introduce REDI a platform for collaboration and networking. Currently, the platform is used in Ecuadorian academic institutions, but we are aiming to expand and adopt through South America. At the time of writing, we are starting three nodes: Colombia, Costa Rica, and Mexico. In addition, we want that each REDI talks to each other through a centralized node and exploit all advantages of Linked Data. For example, every country or institution can host its own data or choose what they share in its networking platform.

**Acknowledgments.** This work is supported by RED CEDIA in the project “Repositorio Ecuatoriano de Investigadores”.

## References

1. Sumba, X., Sumba, F., Tello, A., Baculima, F., Espinoza, M., Saquicela, V.: Detecting similar areas of knowledge using semantic and data mining technologies. *Electronic Notes in Theoretical Computer Science* **329** (2016) 149 – 167
2. Ortiz, J., Segarra, J., Sumba, X., Cullcay, J., Espinoza, M., Saquicela, V.: Authors semantic disambiguation on heterogeneous bibliographic sources. In: *Simpósio Latinoamericano de Manejo de Datos e Información (SLMDI)-JAIHO 46* (Córdoba, 2017). (2017)
3. Chicaiza, J., Piedra, N., Lopez-Vargas, J., Tovar-Caro, E.: Domain categorization of open educational resources based on linked data. In: *International Conference on Knowledge Engineering and the Semantic Web*, Springer (2014) 15–28
4. Cullcay, J., Ortiz, J., Sumba, X., Sumba, F., Saquicela, V.: Identificación automática de artículos indexados en latindex. *Maskana* **8** (2017) 103–111