Towards a sustainable Open Data Ecosystem: First steps for optimizing findability and user feedback

Abdul Aziz, Dagoberto José Herrera-Murillo, Javier Nogueras-Iso, F. Javier Lopez-Pellicer

Advanced Information Systems Laboratory (IAAA)
Aragon Institute of Engineering Research (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: {abdul.aziz, dherrera, jnog, fjlopez}@unizar.es

Summary

ODECO is a 4-year Horizon 2020 Marie Skłodowska-Curie Innovative Training Network initiative (grant agreement 955569) whose main aim is to train the next generation of creative and innovative early-stage open data researchers, to unlock their creative and innovative potential to address current and future challenges in the creation of user driven, circular and inclusive open data ecosystems. This paper describes the first steps performed for optimizing findability and user feedback.

Introduction

Open data are data that are free of charge, openly licensed, machine readable and provided in an open format. In the past decade, open data initiatives have resulted in a greater availability of open data, with the expectation of realizing ambitions such as improved efficiency and effectiveness of public services, increased transparency, accountability, citizen participation, and economic and social value creation [1]. However, there are still many challenges that prevent open data to live up to its potential and sustainability:

1. Lack of skilled people to use open data and create added value.
2. Existing open data ecosystems are neither user-driven nor balance demand-supply matching. Provisioning data from governments to end users is often made through a one-way data portal without differentiating between distinct types of users.
3. Existing open data ecosystems are linear. In the current linear open data ecosystem users of open government data typically capture value without adding value back to the ecosystem.
4. Current open data ecosystems are exclusive. Research and practice are focused on a limited selection of open data providers and user groups, namely governments as data providers and companies as open data users, while many other stakeholders are ignored.

To fully realize the benefits of open data, traditional "one-way street", open data practices and initiatives should be transformed into an open data ecosystem [2]. The overall objective of ODECO is to train a new generation of creative and innovative 15 early-stage open data researchers, able to face current and future challenges, in the establishment of user driven, circular and inclusive open data ecosystems (see Figure 1).

In order to provide an effective training of early-stage researchers, ODECO project is organized in the same way as a research project in terms of work packages and deliverables. Apart from a work package devoted to providing the basic training on the core concepts of Open Data ecosystems (WP1) and the necessary work packages for dissemination (WP6), management (WP7) and ethics (WP8), the project also integrates 4 work packages towards the fulfilment of a sustainable open data ecosystem. These work packages are focused on the research on theories and methods for the development of new user-driven open data portals (WP2) that encourage the circularity (WP3) and inclusiveness of stakeholders (WP4) into a fully sustainable ecosystem (WP5).

The University of Zaragoza contributes as a partner in ODECO with two early-stage research (ESR) projects focused on the necessary technology to provide user-driven tools for optimizing the findability, circularity and inclusiveness of different user groups. This work describes the first steps performed for optimizing the findability and the circularity.

First steps for optimizing findability

Current information retrieval systems behind Open Data catalogs are based on the indexing of the textual and spatial properties of Open Data metadata standards. The text used for metadata is usually short

and more specific properties, such as spatial coverage, lack coordinates to facilitate spatial indexing. As a result, ranking algorithms for estimating the relevance of a dataset with respect to a user query perform poorly, especially in the case of large Open Data catalogs containing millions of metadata records in different languages. The information retrieval systems are not able to retrieve the most relevant results in the first positions of the ranking, and some relevant results are omitted because their descriptions do not overlap with the vocabulary of user queries.

In order to overcome these difficulties, it is believed that the introduction of new approaches at the front-end and the back-end of information retrieval systems can report important benefits with respect to findability. At the back-end new strategies can be applied to increase the performance of textual, spatial, or temporal indexes. At the front-end, user interfaces can be redesigned to be user-centric and facilitate transparent query expansion mechanisms.

As an initial work in this direction, we are designing a search user interface prototype that includes several query expansion mechanisms: theme-based and keyword-based filters are extended with translated terms and related terms available in knowledge organization systems; spatial filters are extended with place names in the area selected by the user; and temporal filters are translated into more specific filters on metadata properties based on dates.

First steps for optimizing user feedback

The portals of some Open Data initiatives have proposed communication channels with users through dedicated discussion forums or web forms where the different communities of users can report their experience in the reuse of data provided through the portals. However, this kind of user feedback is very heterogeneous and the feedback from different initiatives cannot be compared automatically.

Therefore, as an initial step to promote circularity and see how to encourage the involvement of users, we have decided to study the feedback from users on social networks: a common forum where different stakeholders express their opinions about any type of activity or organization.

We have proposed a methodology to compile a series of variables describing both the main features of the Open Data initiatives and their associated Twitter activity in order to infer potential relationships and trends from both the coordinators of open data initiatives, the suppliers and the user communities. The methodology has been assessed by analysing 27 European Open Government Data portals and their activity on Twitter in 2021.

Conclusions

ODECO aims at achieving the involvement of users in the development of circular and inclusive open data ecosystems.

The objective of this work has been to present the initial steps that are necessary from a technological perspective to enhance the findability of existing data and obtain more actionable feedback from users.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955569.

REFERENCES

[1]. JETZEK, T. The value of Open Government Data. Geoforum Perspect. 2013, 12(23), 47-56.

[2]. POLLOCK, R. Welfare gains from opening up Public Sector Information in the UK, University of Cambridge. 2011, vol. 4.

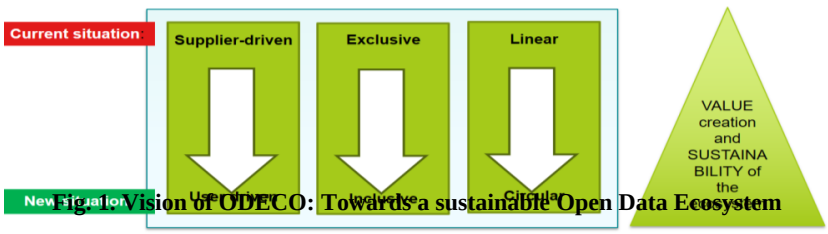

Fig. 1. Vision of ODECO: Towards a sustainable Open Data Ecosystem