

# A comprehensive architecture to support Open Data access, co-creation, and Dissemination

Adegboyega Ojo\*, Arkadiusz Stasiewicz\*, Lukasz Porwol\*, Andrea Petta\*, Donato Pirozzi\*\*, Luigi Serra\*, Vittorio Scarano\*\*, Luca Vicidomini\*\*

\*Insight Centre for Data Analytics, National University of Ireland (NUIG), Ireland

\*\*Dip. di Informatica, Università degli Studi di Salerno, Fisciano, Italy

## ABSTRACT

Contemporary data infrastructures are yet to afford easy access to available data, better understanding of these data, engagement around data to drive collective sense-making and knowledge co-creation. This work synthesizes the knowledge gained through a 40-month research and innovation project which led to designing and implementing an architecture which addresses those challenges.

## CCS CONCEPTS

• **Information systems** → **Collaborative and social computing systems and tools**; • **Applied computing** → **E-government**;

## KEYWORDS

Open Data, Open Data Infrastructure, Open Government

### ACM Reference format:

Adegboyega Ojo\*, Arkadiusz Stasiewicz\*, Lukasz Porwol\*, Andrea Petta\*, Donato Pirozzi\*\*, Luigi Serra\*, Vittorio Scarano\*\*, Luca Vicidomini\*\*. 2018. A comprehensive architecture to support Open Data access, co-creation, and Dissemination. In *Proceedings of dg.o 2018: Proceedings of the 19th Annual International Conference on Digital Government Research, Delft, Netherlands, May 30-June 1, 2018 (dg.o '18)*, 2 pages. <https://doi.org/10.1145/3209281.3209411>

## 1 INTRODUCTION

Open Data (OD) initiatives usually are implemented by exposing an Open Data catalogue over Internet. However, there is a growing pressing on the theme of citizens engagement and on the ways to establish a constructive communication channel among Public Agencies and citizens. Moreover, citizens are more interested in information rather than the data itself. Indeed, data needs proper skills to be processed and exploited. Citizens are interested in answering their questions like who, what, where, and when, about specific topics of interest as well as in discussing those themes.

This paper contributes in this context by proposing a technological stack named ROUTE-TO-PA<sup>1</sup> which encompasses layered services designed to engage various target audiences at different levels.

<sup>1</sup><http://routetopa.eu/>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
*dg.o '18, May 30-June 1, 2018, Delft, Netherlands*  
© 2018 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-6526-0/18/05.  
<https://doi.org/10.1145/3209281.3209411>

Data skilled users, such as a data journalists, that desire to exploit open data in the traditional way can use a CKAN data portal enhanced with an easy and pleasant User Interface as well as a set of advanced tools to manipulate datasets. ROUTE-TO-PA provides a collaborative and social layer that enables citizens in forming on-line communities together with PAs to discuss topics of interest around OD and their visualisations. The knowledge gained by communities can be showcased through a blog (i.e., Wordpress Blog).

Data, information, and knowledge layers are available to PAs as an all-in-one configurable and easy to update virtual machine, which can be shaped to support various scenarios. According to its needs, every PA can decide to 1) install only some layers of the ROUTE-TO-PA 2) decide which layers should be private (within the PA Intranet) or publicly available to anyone, 3) or choose the deployment policies. ROUTE-TO-PA easily integrates with other PA systems (e.g., existing CKAN or blog platform installations). ROUTE-TO-PA is based on a set of Non-Functional requirements (NFR), reported in literature and collected through an extensive requirements elicitation activity over five pilots PA across Europe.

## 2 THE ARCHITECTURE

This Section introduces the main ROUTE-TO-PA architectural layers, which are motivated by and mapped onto the *data-information-knowledge* hierarchy [1]. Thus, the architecture has three main layers: the data, collaborative, and presentation layers.

### 2.1 Data layer TET

The data layer is responsible for managing the data catalogues, search and discovery of datasets, as well as the exploration and analysis of datasets. This layer is designed to extend the basic data services of data platforms such as CKAN with a set of so-called Transparency Enhancing Toolsets (TET) or extensions. We describe below the major elements of this layer as well as the specific features (or extended services) provided through their interaction.

**2.1.1 TET Sub-layers.** There are three sub-layers in the data layer of our architecture. These include storage, service and user interface sub-layers.

The *Storage layer* is concerned with persistence of data and information about datasets. It also provides all the tools for storing and efficient retrieving of Open Data. Specifically, this is responsible for storing the files, structured data, tables and trees as well as the indexes and the metadata. Data can be stored directly in file system storage or in the structured data store.

The *Service sub-layer* provides services on top of Storage layer that could be exploited by the user interface layer described below.

Data catalogue services are used to list the details of datasets and associated metadata stored in the platform. Search service uses the index to search relevant content. Platform extensions services allow external applications to use the platform services. All these services here have corresponding support in the interface layer.

*User Interface sub-layer* provides basic portal functions such as access to the datasets, search, personalisation and customization features. The search feature allows users to find datasets in the catalogue which are stored in the storage layer or publishers servers. The sub-layer also supports analysis and visualizations features which allow users to explore, analyse and visualize various types of datasets including tabular and geospatial data. Various APIs allows external applications to consume services offered by the sub-layer.

**2.1.2 Extended Features.** Here we list a number of features provided by the components and services of the three data sub-layers above:

- **Mobile Interface**
- **Automatic Charts Generation**
- **Exploratory Statistics**
- **Anomaly Detection**
- **Dataset Recommendation**
- **Dataset Merging**
- **Advanced Query and Dataset Splitting**
- **Multi-dimensional Data Analysis**
- **Collaborative Layer Interface**

## 2.2 Collaborative layer SPOD

The collaborative layer goal is to engage citizens, stakeholders, and PAs in forming community around topics of interest, through Data-Driven Discussions. Collaborating around open data allows the collectively generation of information and knowledge; indeed, discussions involve the understanding, combination, reasoning and interpreting data.

SPOD stands for Social Platform for Open Data, it is a virtual place where citizens can meet, forming on-line communities of interests and discussing topics using Open Data together with Public Administrations (PAs). In this sense, SPOD can fill the gaps and reduce distances between citizens and their PAs. Indeed, discussions can be created by PAs or citizens.

**2.2.1 Data Co-Creation.** They are virtual rooms where small groups of participants can meet and collaboratively create datasets, contributing to the open data world. Each room has a shared spreadsheet that allows users to build new datasets from scratch or change existing datasets.

Citizens with their engagement, participate in the creation of really valuable datasets for their communities and for public administrations, increasing public value. PAs could participate in the discussions, intercepting valuable datasets, validating them and publishing over their OD portal. Contributions of citizens have an impact outside the technological platform SPOD, recognizing their effort. This is a concrete impact of social and collaborative activities on PAs.

**2.2.2 Agora for public discussions.** SPOD Agora is the discussion layer, which provides functionalities to allow data-driven discussions, foster interpretation and transformation of open data.

**2.2.3 Knowledge Co-creation.** The Knowledge Co-creation room is the SPOD tool where users can remotely meet and digest a specific topic using OD and their visualisations. This means that in knowledge rooms data become information and then knowledge. A user creates the knowledge room, providing the initial sparks to ignite the creation process and invite other users. It has a collaborative shared document where users can collaboratively contribute to knowledge construction. The knowledge creation process terminates when there is an agreement on the co-created content, which means, that there is a sufficient agreement on the identified issues, generated ideas, and achieved outcomes.

## 2.3 Presentation layer Wordpress Blog

Knowledge Co-creation rooms as well as the Agora for public discussions allow users to create knowledge in a collaborative way by relaying of datasets of OD and their visualisations. SPOD provides a blog based on Wordpress, which is a showcase for the knowledge content. The blog is visible to all users over Internet, not only to the SPOD users, and is a tool to reach a wide audience as well as attract other citizens interested to the topic and their content.

## 3 ROUTE-TO-PA AS A SOFTWARE PLATFORM

The ROUTE-TO-PA consists of different pieces of software. There are many scenarios depending upon: 1) which layer is installed; 2) which layer is visible or hidden for unregistered users; 3) the deployment policy of layers. Each of these scenarios can include the Authentication Server that, if installed, acts as central, shared authentication place for every service of the stack.

ROUTE-TO-PA can be installed on a single machine or alternatively spanned on multiple machines (also named *nodes*). Each node can, of course, be physical or virtual.

## 4 CONCLUSION

The paper introduced a ready-to-use and configurable all-in-one platform to support a variety of scenarios and possibilities to engage citizens. Indeed, ROUTE-TO-PA encompasses a traditional data layer based on an enhanced version of CKAN named to support activities of advanced and skilled users (i.e., data journalists). A collaborative layer, which allows citizens and PAs to form and join communities of interests, discussing around the Open Data and generate visualisations. A knowledge layer, a Wordpress blog, as showcase of the activities of the PA and citizens. All the technologies as well as the infrastructure described in the paper is open source and released under the MIT license.

## ACKNOWLEDGMENTS

The ROUTE-TO-PA presented in this paper has been funded by European Union's Horizon 2020 research and innovation programme under grant agreement No 645860. We gratefully acknowledge the discussions with the project participants, that stimulated our work.

## REFERENCES

- [1] Chaim Zins. 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology* 58, 4 (2007), 479–493. <https://doi.org/10.1002/asi.20508>