Facilitating Data-Driven Innovation using VOICE Observatory Infrastructure*

Eugene Siow
University of Southampton
Southampton, UK
eugene.siow@soton.ac.uk

Xin Wang
University of Southampton
Southampton, UK
xwang@soton.ac.uk

Thanassis Tiropanis
University of Southampton
Southampton, UK
tt2@soton.ac.uk

ABSTRACT

Data is the new capital of the global economy and data analysis can increase efficiency and reduce costs for innovation. Finding and accessing both public and private data is essential to exploiting this data capital. In this paper we describe a platform, VOICE Observatory (VO), that provides an infrastructure and analytical resources to facilitate data-driven innovation. The VO curates a catalogue of analytical resources and provides secure access to them. Through the VO, users can efficiently share, find and use curated analytical resources to create new analytics. Furthermore, the VO provides a well-defined API for authorised applications to access datasets on the behalf of certain users.

CCS Concepts

•**Information systems** → *Service discovery and interfaces;*

Keywords

web observatory, innovation, data

1. INTRODUCTION

Data-driven innovation forms a key pillar of recent economic growth [3]. The convergence of several trends, including the increasing migration of socio-economic activities to the Web and the decrease in the cost of data collection, storage and processing, are leading to the increasing value and utility of data. These data sets that are being collected are becoming a core asset to the economy in terms of fostering new ideas, supporting new business, and creating significant competitive advantages.

VOICE (Virtual Open Incubation Ecosystem) is an open incubator that helps bridge the gap in the broader entrepreneurial development ecosystem by providing an international, virtual innovation and entrepreneurship ecosystem based on open innovation, crowdsourcing and co-creation principles. One unique aim of VOICE is to provide insights into innovation activities of multiple, different stakeholders. This requirement is addressed by adopting a platform called the VOICE Observatory (VO), which helps gather, analyse and assess the anonymous, open innovation data that depict user activity, engagement, collective activity, co-creation environment, enablers and obstacles in the innovation process.

The rest of this paper is organised as follows: Section 2 gives an overview of the VO; Section 3 provides details on the process of producing analytics using the VO; Section 4 describes use cases; while conclusions and future plans are discussed in Section 5.

2. OVERVIEW OF VOICE OBSERVATORY

The VO is a catalogue of analytical resources including datasets, metrics and analytics. It keeps detailed metadata of these resources and thus helps users to share, find and access them. The relationships among resources and essential metadata are shown in Figure 1. Metadata not only facilitates accurate discovery but also provides links among related resources. For example, datasets and the analytics that are produced from them are linked at the metadata level. By traversing these links, users can browse through a large network of analytical resources more effectively. Once related resources are identified, users and applications can access them via the VO in a secure way. The VO adopts a decentralised architecture in which resources can be accessed in-situ without the additional need of gathering them centrally.

Users can publish to the VO resources from outside the VOICE ecosystem while data generated within the VOICE ecosystem such as system logs, user information, innovation activities etcâre made available as datasets and listed in the VO. The VO deploys an access control mechanism so that private datasets and analytics containing sensitive information are protected and only accessible to authorised users and applications.

The VO allows users to access datasets and analytics via a web interface. It also provides an API that allows applications to programmatically access analytical resources. The API is protected by OAuth 2.0 [2] to ensure that access to any resource is controlled by the resource owner and can be delegated to applications without having to reveal the underlying credentials. Applications can also access multiple datasets simultaneously, through the API, to perform complementary analytics and information fusion. It is also possible to access live data streams and build real-time an-

*This work is partly supported by the EU project VOICE (GA 621137).

DDI'16, May 22-25 2016, Hannover, Germany

ISBN 978-1-4503-4360-2/16/05...$15.00

DOI: http://dx.doi.org/10.1145/2911187.2914584

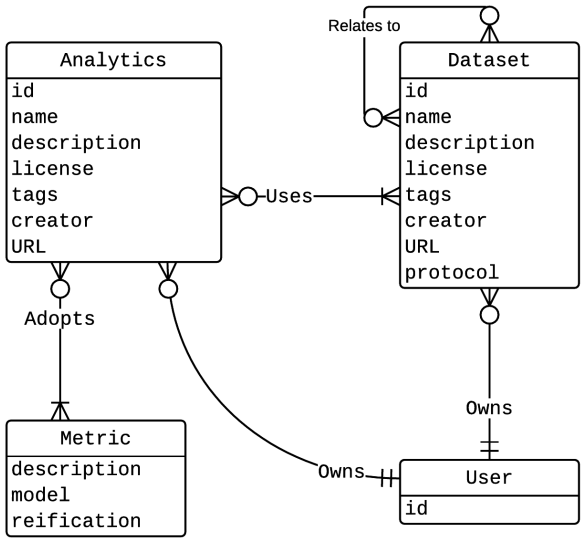

Figure 1: ERD diagram of resources cataloged by the VO.

alytics through standards like AMQP[1].

3. ANALYTICS FORMATION USING THE VIRTUAL OBSERVATORY

Analytics in the VO are in the form of applications, usually comprising visualisations. The basic steps to make analytics include identification of datasets and metrics, as show in Figure 2. The process goes as follows: a user chooses metrics and searches amongst datasets in the VO to inform the metrics. Metrics can come from the user's own experience, literature, online resources or other applications in the VO. The user then implements the metrics as applications that programmatically access chosen datasets using the VO API. It is recommended that metrics used in the analytical application be explicitly defined in metadata, in which case they can be adapted and stacked with new datasets to produce new analytics.

4. USE CASES

4.1 Sharing and Visualising VOICE KPIs

Indicators are widely used by governments and companies to quickly obtain an overview of technology and innovation development. For example this technique is used in Africa to effectively track new development and inform policy making [1].

A concrete use case is monitoring the performance of VOICE itself. The VO acts as an internal innovation enhancement mechanism that will continuously monitor and identify areas for improvement within the VOICE ecosystem as a whole. One example is that VOICE uses the VO to monitor its own growth in terms of several key performance indicators (KPIs). The KPIs consist of the number of representative

[1]https://www.amqp.org/resources/specifications

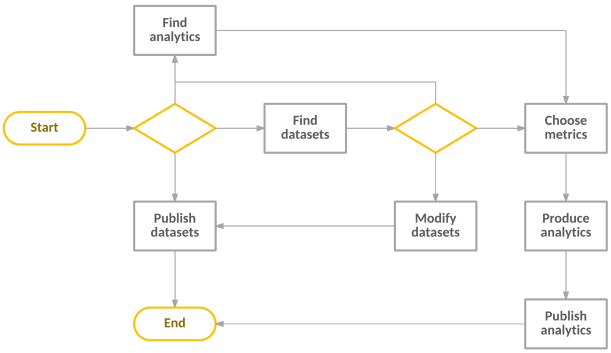

Figure 2: Work flow of producing analytics in the VO.

users, user interactions, started and graduated ideas, spin-offs from small and medium-sized enterprises, achieved investments, accumulated funding etc. These data are not only interesting to VOICE administrators but are also valuable to VOICE users. Rather than developing a specific API, like Crunchbase[2] does, to access these data, VOICE publishes these data as a dataset in the VO, and delegates access control to the VO. Users of VOICE access these data in the same way as they access any other dataset in the VO, either from the web interface or via the VO API as described in Section 2.

An application is developed by VOICE to visualise VOICE KPI in a radar chart, as shown in Figure 3. The radar chart shows VOICE KPI at different stages and indicates the development of VOICE.

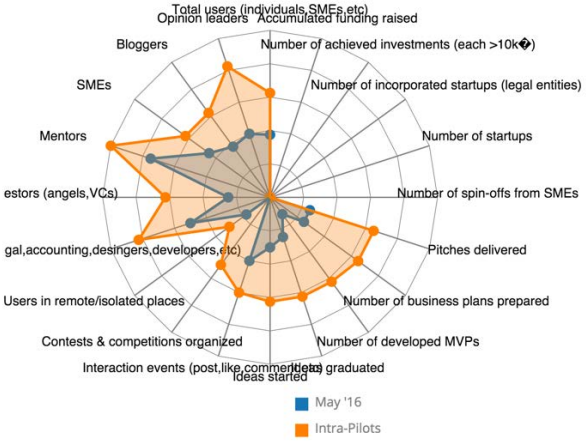

Figure 3: VOICE KPI as a radar chart.

4.2 Correlating Innovation Expenditure and Birth Rate from Eurostat Datasets

Another use case of the VO is when an application requires more than one dataset. Instead of having to find or collect different datasets, transform and load them into a central database, an application can access multiple datasets

[2]https://www.crunchbase.com

simultaneously via the VO API. Figure 4 shows a scatter plot representing multiple dimensional data including the total expenditures on innovation (x axis), the number of new startups (y axis) and the total number of startups (size of circles) grouped by countries. The data comes from three datasets: Crunchbase company data[3] and Eurostat business demography data[4] that gives the total number of new startups each year, and Eurostat innovation activities and expenditures data[5] that gives the innovation expenditure. The VO API used through this analytics application allows the data to be accessed in-situ, in a decentralised manner - allowing emphasis to be placed on analytics and application logic rather than data wrangling.

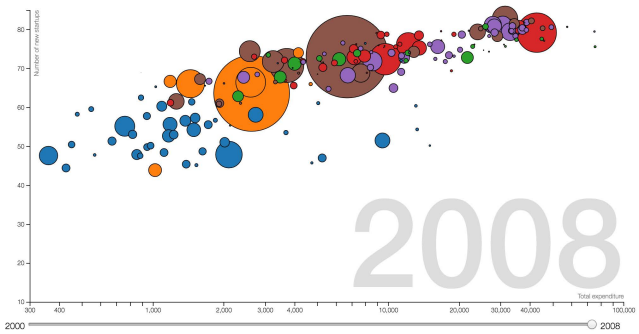

Figure 4: An example scatter plot of the number of new startups and total expenditures on innovation grouped by countries.

5. CONCLUSIONS AND FUTURE PLAN

In this paper we describe the design and usage of the VO, through various case studies. We show how the VO can be used to enable users to conveniently find, share and access private and public datasets and analytics, to facilitate data driven innovation. Furthermore, the VO curates detailed metadata of various analytical resources and provides a uniformed API to allow secure access to datasets, facilitating reuse and placing the emphasis on driving innovation.

In the future, it is essential to research further lowering the barrier of dataset and analytics discovery and access over the Web and improving the performance of accessing multiple remote datasets.

6. REFERENCES

[1] Fred Gault. Science, Technology and Innovation Indicators in a Changing World. Technical report, 2007.

[2] D. Hardt. The OAUTH 2.0 Authorization Framework. Technical Report RFC6749, 2012.

[3] OECD. *Data-Driven Innovation: Big Data for Growth and Well-Being*, volume 215. 2015.

[3] https://github.com/notpeter/crunchbase-data

[4] http://ec.europa.eu/eurostat/web/structural-business-statistics/entrepreneurship/business-demography

[5] http://ec.europa.eu/eurostat/web/products-datasets/-/inn_exp