

Breaking Public Administrations' Data Silos: the Case of Open-DAI, and a Comparison between Open Data Platforms.

Original

Breaking Public Administrations' Data Silos: the Case of Open-DAI, and a Comparison between Open Data Platforms / lemma, Raimondo; Osella, Michele; Morando, Federico. - In: EJOURNAL OF EDEMOCRACY AND OPEN GOVERNMENT. - ISSN 2075-9517. - STAMPA. - 6:2(2014), pp. 112-122.

Availability:

This version is available at: 11583/2573941 since:

Publisher:

Danube-University Krems

Published

DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)



Breaking Public Administrations' Data Silos

The Case of Open-DAI, and a Comparison between Open Data Platforms

Raimondo Iemma

Nexa Center for Internet & Society at Politecnico di Torino, DAUIN – Corso Duca degli Abruzzi 24, 10129 Torino (Italy), raimondo.iemma@polito.it, +39 011 090 5954.

Federico Morando

Nexa Center for Internet & Society at Politecnico di Torino, DAUIN – Corso Duca degli Abruzzi 24, 10129 Torino (Italy), raimondo.iemma@polito.it, +39 011 090 5954.

Michele Osella

Istituto Superiore Mario Boella – Via Pier Carlo Boggio 61, 10138 Torino (Italy), osella@ismb.it, +39 011 227 6201.

Abstract: *An open reuse of public data and tools can turn the government into a powerful 'platform' also involving external innovators. However, the typical information system of a public agency is not open by design. Several public administrations have started adopting technical solutions to overcome this issue, typically in the form of middleware layers operating as 'buses' between data centres and the outside world. Open-DAI is an open source platform designed to expose data as services, directly pulling from legacy databases of the data holder. The platform is the result of an ongoing project funded under the EU ICT PSP call 2011. We present the rationale and features of Open-DAI, also through a comparison with three other open data platforms: the Socrata Open Data portal, CKAN, and ENGAGE.*

Keywords: Public Sector Information, Service-Oriented Architecture, Open Data platforms, Benchmarking, Exploitation.

Acknowledgement: This paper was drafted in the context of the Open-DAI ICT-PSP project (GA n°297362). The authors acknowledge the support of the European Commission and are grateful to the project members for their support.

By allowing external actors to reuse government data and tools, new services can be provided to citizens and by citizens (e.g., Nam, 2012; Linders, 2012). In this way, the government can be turned into a powerful "platform" also involving innovators (e.g., O'Reilly, 2011). At the same time, by using common open repositories, public administrations can save time and money from the automatization of internal data exchange, while increasing their degree of transparency (Stiglitz et al., 2000). Not by chance, 'open by default' is becoming one of the foundational principles of open data-related pieces of legislation, including the recently updated European Directive on Public Sector Information (PSI).

However, the typical information system of a public agency is not open by design. The general public can frequently access to services based on software applications. But raw data and/or granular data services are typically not available to the general public. Actually, a low level access to the system is reserved to a small number of public officials. Apart from technicians, the most frequent category of users consists of external service providers, or of bulk data re-users. In both cases, the conditions and purposes of access typically result from well formalised agreements. A huge amount of relevant public sector information is stored in proprietary formats (see, e.g., the background analysis for the UK Action Plan 2013 related with the G8 Open Data Charter). Data streams are usually fragmented, with information only flowing vertically, and rarely between departments (Tapscott et al., 2008). The same kind of issue applies to interaction between agencies at different administrative levels, with the additional aspect of semantic interoperability.

As a consequence, open data dissemination is typically not yet embedded in the ICT management strategy as a step of the data life-cycle (e.g., Fioretti, 2011).

Making public agencies' information systems open is arguably a challenge for the medium and long term (see, e.g., the UK Open Standards Principles, 2012). In the short run, it seems useful to track endeavours aimed at smoothing the process of data publication, e.g., in the form of middleware layers operating as 'buses' between data centres and the outside world. In fact, several public administrations have started adopting technical solutions in this respect. At the same time, policy contributions set requirements in terms of openness and interoperability.

In this paper we discuss the features of Open-DAI, an open-source platform designed to enable organisations to expose data as services, directly pulling from their legacy databases. Open-DAI is the result of an ongoing project funded under the EU ICT PSP call 2011, Objective 4.1: Towards a cloud of public services. Amongst the expected impacts, an increase in the efficiency of administrative services, applying new architectural approaches to legacy assets. As a EU-funded project over the period February 2012 - September 2014, Open-DAI is coordinated by CSI Piemonte, the ICT in-house company of Regione Piemonte, and involves public administrations from Italy, Spain, Sweden, and Turkey¹.

This paper is organised as follows. Section 1 describes the overall architecture of Open-DAI. Section 2 contains a comparative analysis with other platforms for open data exposure. In Section 3, exploitation scenarios are presented. Section 4 draws conclusions, and discusses future works.

1. Open-DAI

1.1. Background and Objectives

When defining the optimal technological approach for Open-DAI, the EU call (ICT PSP 2011, Objective 4.1) specifications were taken into account. Two technological paradigms were adopted at the infrastructural and the architectural levels respectively: cloud computing and service-oriented architectures (SOA). Cloud computing can ensure an elastic provision of resources, with a trade-off emerging between the efficiency savings driven by a decentralisation / rationalisations of the IT estate of an organisation, involving concerns related with reliability, data protection and security. Service-oriented architectures refer to design patterns aimed at enabling exchange of information between services without the need to make any changes in underlying programmes. All these aspects are particularly relevant for the public sector (e.g., Armbrust et al., 2009).

SOA principles place the interoperability of software services at the core of the design of systems development and integration. Balzer (2004) lists as the most relevant guiding principles to direct development, maintenance, and usage of the SOA "[r]euse, granularity, modularity, composability, and componentization", together with "[c]ompliance to standards". Indeed, these are the functional equivalent for a service of the most desirable characteristics of open government data, which should not just be accessible, but also available for re- use as raw data that can be technically and legally remixed with other data and possibly semantically described, using standard vocabularies (e.g., Berners-Lee, 2006; or Heath & Bizer, 2011).

Open-DAI is designed to play the role of an 'open data hub', allowing data exposure using standard protocols, and avoiding data duplication. Its second objective is to improve interoperability, without any modification of the legacy logical and physical infrastructure.

1.2. Rationale, Architecture, and Technological Choices

At the highest level of abstraction, Open-DAI is a platform that directly extracts data from legacy databases that sit behind existing public sector applications. Under the rules defined by the data holder, Open-DAI generates a virtualised version of the database in the cloud, and a layer that

¹ The complete list of the Open-DAI partners is available at <http://www.open-dai.eu>.

exposes the transformed data as services (RESTful APIs²), therefore providing data re-users (e.g., developers) with a 'real time' connection with the legacy data.

At the architectural level, Open-DAI encompasses two interrelated components: (i) a cloud infrastructure; (ii) a SOA-compliant middleware layer operating within each private cloud owner, i.e., a data holder (to ensure autonomy scalability related with specific needs), but encompassing common components (so that the middleware is managed by the cloud provider, i.e. the Open-DAI maintainer - without any extra burden for the public agency using it). Technological choices result from the integration of 'out-of-the-box' open-source tools.

The cloud computing infrastructure is implemented through CloudStack, an open-source solution that organises virtual machines into logical groups, helps to deploy them on physical host, and provides fine-grained management features. A cloud cluster is managed by CSI Piemonte (Italy), as coordinator of the Open-DAI project. In practice, each user of the platform receives a private allocation (domain) of the cloud, isolated at the network layer for security purposes.

The middleware layer exposes data services, allowing the creation of new services, and integrating them using a SOA-compliant approach, as in Figure 1 - Schematisation of the Open-DAI architecture.

Access to legacy databases is ensured by a data virtualisation layer (the open-source component JBoss TEIID, a data virtualisation system that allows applications to use data from multiple, heterogenous data stores), using Virtual Private Network (VPN) connections, and also allowing data transformations. Using the D2RQ platform as semantic module, Open-DAI also enables exposure of linked data, with an RDF triple store coupled with a SPARQL endpoint³. Geographic data are released using GeoServer, an open-source Java J2EE application designed for that purpose.

The task of publishing of data services (as RESTful APIs) is carried out by the open-source web server Apache, with WSO2 as Enterprise Service Bus⁴, so that the existing infrastructure (including servers, storage systems and/or relational DBs) is retained. This approach is particularly suitable for the exposure of frequently changing data. As a 'proof-of-concept' of possible data reuses enabled by Open-DAI, several pilot services were created by the project partners (see section 2.3).

A 'common components' group encompasses a set of tools which are meant to facilitate the management and monitoring activities carried out by the platform user (typically, a public administration), including features to support configuration of computer systems provided through the open-source tool Puppet.

² We call RESTful the Web service APIs that adhere to the Representational State Transfer (REST) architectural constraints, and, in particular, have: 1) a base Uniform Resource Identifier (URI); 2) an Internet media type (often JSON); 3) standard HTTP methods to call the API; 4) hypertext links to reference state; 5) hypertext links to reference related resources.

³ SPARQL is a recursive acronym for "SPARQL Protocol and RDF Query Language", a query language for databases able to retrieve data stored in Resource Description Framework (RDF) format, which is based upon the idea of making statements about resources (in particular web resources) in the form of subject-predicate-object expressions.

⁴ An enterprise service bus (ESB) is a software architecture model used for designing and implementing communication between mutually interacting software applications in a service-oriented architecture (SOA)

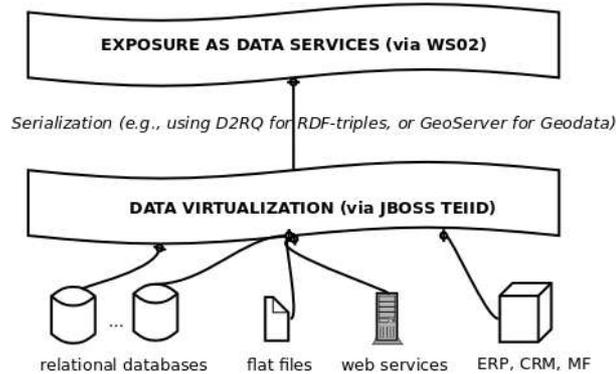


Figure 1: Schema of the Open-DAI architecture

1.3. Examples of Public Data Re-use: Pilot Services

Pilot services were developed within the project, in the form of mobile or web applications. These services represent a proof-of-concept of possible data reuses enabled by the platform. Prior to the actual design of the pilot services, an assessment of the datasets made available by the public administrations involved in the project was performed. This activity included a description of the structure and fields of the datasets, as well as further scrutiny aimed at clearing Intellectual Property Rights (IPRs) and at managing the existence of personal data. Beyond existing technical and legal constraints, datasets were selected according to the expected value in their reuse, assuming for instance the possibility of geo-referencing, and the presence of real-time updates, as some of the key features in this respect. Apps are designed to provide real-time information on: air quality (Piedmont Region and Barcelona Municipality); road accidents (with the future opportunity to also gather real-time data from citizens) (Piedmont Region and Lleida Municipality); location of points of interest (Karlshamn Municipality and Ordu Municipality)⁵.

2. Comparative Analysis

2.1. Requirements Definition

In order to compare Open-DAI with other solutions for data publication, we engaged in the selection of meaningful parameters, e.g. capturing specific features related with the functioning of a platform. We decided to derive such parameters from requirements (explicitly or implicitly) expressed in several public documents. This first set of sources encompass: legislation at European (e.g., the PSI Directive, the INSPIRE Directive), national (e.g. the Italian Code for the Digital Administration) and local (e.g. Piedmont legislation on PSI) level; strategic plans related with the implementation of the EU Digital Agenda (e.g., EU eGOV action plan, European Interoperability Framework, the “Connecting Europe Facility” proposal of regulation; the “Open Data Support” initiative by DG CONNECT); national guidelines on public sector information management (e.g. UK Open Data white paper; UK Public information management principles; UK Open Standards Principles; Guidelines on public sector information reuse by the Italian Agency for digital policies; Open Data *vademecum* by Formez); tender specifications for open data portals (e.g., the call for tender for the EU open data portal; the call for tender for the open data portal of the Lazio Region); studies (e.g., “Study on persistent URIs” by ISA; Garcia & Pardoet al., 2005, “E-government success factors: Mapping practical tools to theoretical foundations”; others cited throughout this paper).

As a result of an extraction carried out in two steps, i.e., the first one meant to elicit a long-list of preliminary requirements drawing on aforementioned sources, and the second one aimed at

⁵ The mentioned pilots are described in the Open-DAI website (<http://open-dai.eu/>)

distilling the short-list of refined requirements adopted as criteria for benchmarking purposes). Finally, 18 requirements have been finally obtained with the idea of capturing basic recurrent features characterizing a 'state-of-the-art' platform for data publication. Those requirements were organised in four categories, defined from the point of view of a data holder adopting and using the platform, and describing: (i) publication features (capturing, e.g., the process through which data are published) [A1 to A8]; (ii) data features (e.g., in terms of standards supported by the platform) [B1 to B5]; (iii) the platform architecture, or other general features [C1 to C3]; (iv) add-ons [D1 to D2]. Arguably, this categorisation is just one amongst the many possible, also considering that the impact of most of the features can be reflected in several aspects at the same time. For instance, specific platform features are supposed to also maximise data reusability (beyond enabling data holders to effectively engage in a publication strategy).

2.2. Benchmarking

Platforms subject to benchmarking were chosen so to ensure a reasonable coverage of the existing solutions, still preserving comparability. We then included in our benchmarking activity: a commercial, widely adopted platform (Socrata Open Data portal); a 'community-based', widely adopted platform (CKAN); two platforms deriving from the work carried out within European projects, therefore with a limited user base so far, but with considerable potential, such as ENGAGE and Open-DAI.

Socrata is a U.S. company founded in 2007, providing social data discovery services for opening government data. Its 'Open Data Portal' provides a cloud-based service for data publishing, metadata management, data catalogue federation, and exposure of data as services. Data can be published manually, or through dedicated APIs. Search APIs allow queries at the dataset level. Data reuse is also enabled through developers APIs (in a 'freemium' logic). Currently, around 50 out of 330 public data catalogues worldwide use the Socrata software (figure derived from <http://www.socrata.com/customer-spotlight/>). In early 2013, Socrata launched the "Community Edition" of its Open Data portal (free and open-source).

CKAN (acronym for Comprehensive Knowledge Archive Network) is an open-source data management platform maintained by the Open Knowledge Foundation. Currently, it is used by around 50 out of 330 data catalogues worldwide (figure derived from <http://ckan.org/instances>), including the recently issued European Open Data portal (<http://open-data.europa.eu/>), developed by the Belgian company Tenforce. CKAN is released under several versions that differ from each other in terms of features and service level. While the download and usage of CKAN are free, the CKAN team offers deployment services. CKAN furthermore allows catalogue federation through its APIs.

ENGAGE is a combination of CP & CSA project funded under the European Commission FP7 Programme. Its main goal is the development and use of a data infrastructure, incorporating distributed and diverse public sector information (PSI) resources, capable of supporting scientific collaboration and research, particularly for the Social Science and Humanities (SSH) scientific communities, while also empowering the deployment of open governmental data towards citizens.

The main results of the benchmarking are reported in Table 1.

Table 1: Open Data Platforms Benchmarking

Requirement	Open-DAI	Socrata OD Portal	CKAN	ENGAGE
A1. Uses an automatic process to expose data stored in legacy databases.	Yes. The platform pulls data (virtually in real time) from legacy DBs, with standard connectors available (for most DBs).	Automatable, using 'Publish' APIs made available to data holders.	Automatable, using 'Publish' APIs made available to data holders.	Automatable, using 'Publish' APIs made available to data holders.
A2. Uses APIs at the data level (e.g., transformations).	Yes (e.g., CSV to JSON), and data filtering.	Yes (e.g., CSV to JSON).	Yes, with its 'Data storer' plugin.No.	No.
A3. Promotes the use of standard metadata.	Under development, with the aim of following the same approach as CKAN in this respect.	Yes, in the 'Community' edition, using standard vocabularies such as DCAT (W3C).	Yes , using standard vocabularies such as DCAT (W3C).	Yes, three-layer metadata architecture: discovery (e.g., Dublin Core, eGMS, CKAN), context (e.g., CERIF), detail (i.e., subject-specific or topic-specific).
A4. Enables catalogue federation (with CKAN-API as de facto standard).	Not yet, ongoing development.	It allows federation, using CKAN metadata (but not the API) as a standard.	Yes (trivially).	Yes, using CKAN-API as standard.
A5. Allows to perform a data quality check (and related data refinement).	Not directly.	Not directly, but it enables quality check, e.g. identifying data types for values.	Not as embedded functionality, but a OpenRefine extension is available for CKAN.	Not as embedded functionality, but ENGAGE provides a strong integration with OpenRefine). Besides, data curation by the community is encouraged.
A6. Is designed to be integrated with (or includes) a front-end / open data portal.	Yes, i.e. it is planned to integrate Open-DAI as a back-end of the Open Data portal of the Piedmont Region.	Yes, but poorly customizable. Ongoing attempts by third parties, e.g. an integration point with Drupal is being developed, still in Alpha mode.	Yes. On top of the standard front-end, there are well experimented modules for Drupal and Wordpress.	Yes. A full-fledged front-end is included.
A7. Releases APIs to reuse data.	Yes, RESTful APIs.	Yes, RESTful APIs.	Yes, with its 'Data storer' plugin.	No. RESTful APIs are implemented only at metadata level.

Requirement	Open-DAI	Socrata OD Portal	CKAN	ENGAGE
A.8 Enables browsing at the data level.	Yes.	Yes.	Yes, with its 'Data storer' plugin.	Yes.
B1. Designed to expose Open Data.	Yes.	Yes.	Yes.	Yes.
B2. Designed to publish dynamic data.	Yes.	Yes.	No, only static files.	Yes.
B3. Designed to expose geospatial data.	Yes.	Yes.	No, but allows to expose georeferenced metadata.	Yes.
B4. Designed to expose Linked Open Data (meaning at least RDF triple store+SPARQL endpoint+other features, e.g., ontology mapping).	Yes.	No. The 'Community edition' allows data exposure as RDF, but with no triple store, nor SPARQL endpoint.	No, just linked metadata.	Yes.
B5. Presents prototypes of data reuse.	Yes, e.g., Open-DAI pilot services.	Yes.	Yes, e.g., tabular previews.	Yes, derived datasets .
C1. Released as open-source software.	Yes.	Not the standard edition (Yes, in case of the 'Community Edition').	Yes.	Not yet. However, the consortium is inclined to release the basic engine under the MIT License.
C2. Available in a cloud environment.	Yes, at all levels of abstraction.	Yes, SaaS.	No.	Yes, SaaS.
C3. Available "on premise" by the data holder (i.e., as a DB independent from the provider's API).	Yes.	No.	Yes (but has a 'hosted' option).	Yes.
D1. Allows to gather feedback on data (also in terms of 'forked' datasets).	Yes, in the case of service pilots that enable data flow in both directions.	Yes, users can manipulate files and save their edits.	Yes (through the 'datahub.io' portal).	Yes. Derived datasets are welcome and are tracked by the system.
D2. Encompasses a ticketing system.	No.	No.	No.	Yes, the issue tracking system covers bug, license issues and general

Requirement	Open-DAI	Socrata OD Portal	CKAN	ENGAGE
				questions/suggestions. Moreover, users may place a new request for data not available on the portal.

2.3. Discussion

Open-DAI can be conceived as a ‘bus’ that, by federating governmental data repository, breaks silos existing among governmental agencies making data available for a twofold goal: on one hand, Open-DAI becomes a propellant for a fluid flow of data (even in case of confidential data not bound to be published) among public bodies and, on the other hand, allows the exposure of Open Government Data to the outside world.

At this level of abstraction, Open-DAI holds several common points with other solutions designed with the same purpose. However, considering specific functionalities, differences may emerge as significant, and therefore worth exploring.

The process under which data are extracted from legacy DBs is arguably one of the distinctive features of Open-DAI. In fact, Socrata OD Portal, CKAN and ENGAGE enable data exposure in a ‘push’ mode, i.e. using “publish” APIs available to data holders, who set them according with their needs (e.g., in terms of frequency of update), while Open-DAI - as already explained - ‘pulls’ data from DBs of legacy applications. From the point of view of developers, the data they get using Open-DAI is a transformation of (a query on) a legacy database, while using other platforms developers get the most recent version of the published data. Depending on the optimal frequency of update of a specific dataset (from the point of view of its meaningfulness, and actual reusability), this aspect could turn out to be more or less relevant. Moreover, Open-DAI provides a broad set of services/formats, and fine-grained API management (through WSO2), which is not always the case for the platforms used for this comparison.

Currently, Open-DAI is not integrated with a ‘traditional’ portal, although, for instance, there are plans to expose its APIs on the Open Data portal of the Piedmont Region. Together with catalogue federation, this aspect represents one of the future developments foreseen for Open-DAI. Both CKAN and ENGAGE encompass a full-fledged front-end (a ‘data hub’, in the first case), while the CMS of the Socrata Open Data portal has advanced data preview features, but is perceived by its users as poorly customisable. Open-DAI is a potential substitute of ‘traditional’ (e.g., not exposing data as services) open data portals, but it can also be seen as a complement to these pieces of software. In fact, to serve the broader “data portal” market, Open-DAI needs a front-end: it can get it through integration with an open data portal and/or with CKAN, composing, in this way, the same kind of offering as softwares such as Socrata Open Data portal.

Although with some differences in the way they are implemented, all platforms exposing data as services feature advanced solutions in terms of data exposure, e.g. related with specific formalisms or categories of data, while CKAN usually enables these kinds of features only at the metadata level. In particular, among the considered platforms, currently only Open-DAI and ENGAGE are designed to expose (and allow standard queries on) Linked Open Data.

3. Exploitation Scenarios

In light of the comparative analysis above, and of the actual incentives and constraints of the partners, four exploitation scenarios were drafted for Open-DAI.

Looking at Scenario 1, the partial reuse of project outputs as components is a default and worst-case scenario. Under this scenario, when Open-DAI ends as a EU-funded project, nobody

maintains it as a unique platform and the various exploitation items become reused in other contexts. Obviously, this is a sub-optimal scenario.

Under Scenario 2, Open-DAI would be maintained as an open-source platform by one of the partners of the former consortium (most probably, the project leader). Benefits could be experienced at different levels, not only in terms of tangible legacy, but also for third parties willing to engage in further developments. Moreover, adoption costs for interested PAs would be reasonably low if compared with market offerings. In addition, the maintainer could achieve a potentially high return, especially in terms of economies of scale and scope within its organisation.

Under Scenario 3, a “data cloud” offer (essentially equivalent to the Open-DAI platform) would be promoted, as part of a major public procurement action, e.g. by a national / local public group purchasing organization (GPO) able to capture significant scale and scope economies. However, in order to become a service purchased by PAs on a regular basis, this “data cloud” should be defined and evaluated under standard terms, which is currently rather complex. Moreover, competition concerns may, although some standard remedial/mitigation actions could be foreseen (e.g., avoiding the ‘winner takes it all’ approach).

Scenario 4 captures a market approach, defined through a detailed business plan, also considering the comparative analysis performed. Possible sources of revenue are identified as being mainly related with (i) start-up and integration of the platform, and (ii) supply of Open-DAI as a service (with re-users served in a “freemium” mode). Realistic cost and demand scenarios make Open-DAI economically sustainable even at the level of a single European country and with a single software maintainer. In any case, the incentive to offer Open-DAI to public administrations, even if barely reaching break-even, would be strong, also in relation with potential positive externalities. In particular, while some of the benefits could be internalised by the PSI holder, some others are spillover effects related with the blossoming of new business opportunities in the Open Data ecosystem (see, e.g., Ferro & Osella, 2013).

It emerges that Scenario 2 is reasonably feasible, and, given the willingness expressed by some of the partners, it represents the most likely alternative. Scenario 3 is possibly granting a higher chance of internalising externalities deriving from a standardised adoption of Open-DAI, but weak in terms of autonomy from decisions of external stakeholders. Finally, the market exploitation by some of the partners, e.g. interested in providing services around Open-DAI, is to be considered as arguably likely.

4. Conclusions and Future Work

The platforms for open government data publishing share a set of common features. All of them allow their adopters to reach high-level policy objectives related with enabling data reuse by third parties, in standardised (and therefore, at least partially, interoperable) ways. Yet, differences may also be identified. These are related, on the one hand, with the type of integration (if any) with the legacy systems of public administrations; the way this aspect is managed by Open-DAI is arguably its plus. On the other hand, features improving data discoverability (also through multiple catalogues), and reusability, e.g., integration with existing data portals, are supposed to maximise the expected value in the use of the platform by developers and/or other interested parties. In this respect, reaching a critical mass of public administrations adopting the platform would entail an increase the available data in volume, variety and quantity, and would attracting more data re-users as a consequence. This is arguably the main challenge for Open-DAI. Generally speaking, in fact, interaction with potential re-users could be improved in any of the examples taken into account. For instance, a properly sustainable model for a ‘public data versioning’ seems is not yet available yet, although several attempts have been carried out.

We submit that the benchmarking exercise drafted in this paper could be further developed, even in the short run, and thus become a useful reference for practitioners, policymakers and, of course, public administrations facing a choice between open data platforms. In particular, the set of references used to elicit requirements could be broadened. More in general, requirements could be

expressed within a richer taxonomy in such a way to explicitly capture interrelations, providing a comprehensive framework for further benchmarking. In fact, the authors of this paper are considering enriching the analysis by taking stock of previous research endeavours in the same vein (such as Zuiderwijk et al, 2013a, Zuiderwijk et al, 2013b, which use as sources primary research involving stakeholders, as well as a broad literature review) so to consolidate the set of requirements to be taken into account. From the point of view of open data platforms, a possible strategy would be to encompass in the analysis specific software modules (instead of, or in addition to, full-fledged softwares integrating several modules), so to be able to isolate specific features. As a final goal, we plan to set a roadmap for the creation of a tool allowing stakeholders – such as public decision makers interested in open data – to identify the best set of software solutions given their needs and priorities. As an intermediate step, the authors are constructing and populating a semantic Wiki, so to expose the results of the benchmarking as response to queries, e.g., which softwares meet a specific requirement, and to which extent.

References

- Chadwick, A. (2006). *Internet Politics: States, Citizens, and New Communication Technologies*. Oxford: Oxford University Press.
- Agenzia per l'Italia Digitale (2013). *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico (secondo semestre 2013)*. Available at: http://archivio.digitpa.gov.it/sites/default/files/allegati_tec/LG_Val_PSI_v1.0.pdf
- Archer, P., Goedertier, S., & Loutas, N. (2012). Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. *Interoperability Solutions for European Public Administrations*.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. & Matei, Z. (2009). *Above the clouds: A Berkeley view of cloud computing*. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS, 28.
- Balzer, Y. (2004). *Improve your SOA project plans*. IBM Global Services.
- Berners-Lee, T. (2006). Design issues: Linked data.
- Bonaccorsi, A., & Rossi, C. (2006). Comparing motivations of individual programmers and firms to take part in the open source movement: From community to business. *Knowledge, Technology & Policy*, 18(4), 40-64.
- DigitPA (2013). *Codice dell'amministrazione digitale*. Available at: <http://archivio.digitpa.gov.it/amministrazione-digitale/CAD-testo-vigente>.
- European Commission (2013). Directive of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>.
- European Commission (2011). *Tender specifications. Implementation of a European Commission open data portal*. Smart 2011/0050. Available at: http://www.eu-service-bb.de/download/000227/tender_specifications_final.pdf.
- European Commission, COM(2010) 744 final. Annex 2 to the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions 'Towards interoperability for European public services. Available at: http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf.
- European Commission, COM(2010) 743 final. *The European eGovernment Action Plan 2011-2015*. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0743:FIN:EN:PDF>.
- European Commission, COM(2010) 665 final. *Proposal for a regulation of the European Parliament and of the Council establishing the Connecting Europe Facility*. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0665:FIN:EN:PDF>.
- European Commission (2007). Directive of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>.
- Ferro, E., & Osella, M. (2013 R). *Eight Business Models Archetypes for PSI reuse*. Open Data on the Web Workshop, London.
- Fioretti, M. (2011). *Open Data: Emerging trends, issues and best practices*. Laboratory of Economics and Management of Scuola Superiore Sant'Anna, Pisa.
- Gil-García, J. R., & Pardo, T. A. (2005). *E-government success factors: Mapping practical tools to theoretical foundations*. *Government Information Quarterly*, 22(2), 187-216.

- Heath, T., & Bizer, C. (2011). *Linked data: Evolving the web into a global data space*. Synthesis lectures on the semantic web: theory and technology, 1(1), 1-136.
- Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly*, 29(4), 446–454.
- Nam, T. (2012). *Suggesting frameworks of citizen-sourcing via Government 2.0*. *Government Information Quarterly*, 29(1), 12-20.
- O'Reilly, T. (2011). *Government as a Platform*. *Innovations*, 6(1), 13-40.
- Regione Lazio (2013). *Gara a procedura aperta ai sensi del D. Lgs. 163/2006 e s.m.i. per l'appalto di servizi finalizzati alla realizzazione e all'esercizio del portale dati.regione.lazio.it ed allo svolgimento delle attività a supporto delle iniziative connesse all'apertura ed al riutilizzo delle informazioni del settore pubblico della Regione Lazio*. Available at: http://www.filas.it/Downloads/Bandi/Capitolato_Tecnico_regione_lazio_dati.pdf.
- Regione Piemonte (2011). *Legge regionale n. 24 del 23 dicembre 2011*. Available at: <http://arianna.consiglioregionale.piemonte.it/ariaint/TESTO?LAYOUT=PRESENTAZIONE&TIPODOC=LEGGI&LEGG E=24&LEGGEANNO=2011>.
- Tapscott, D., Williams, A. D., & Herman, D. (2008). *Transforming Government and Governance for the Twenty-First Century*. nGenera Insight.
- UK Cabinet Office (2011). *Information principles (version 1.1)*. HM Government, London.
- UK Cabinet Office (2012). *Open Standards Principles*. HM Government, London.
- UK Cabinet Office (2012). *Open Data White Paper*. HM Government, London.
- UK Cabinet Office (2013). *G8 Open Data Charter – UK Action Plan 2013*. HM Government, London.
- Zuiderwijk, A., Janssen, M., & Jeffery, K. (2013a). *Towards an e-infrastructure to support the provision and use of open data*. In Conference for E-Democracy and Open Government (p. 259).
- Zuiderwijk, A., Janssen, M., & Parnia, A. (2013b). The complementarity of open data infrastructures: an analysis of functionalities. In Proceedings of the 14th Annual International Conference on Digital Government Research (pp. 166-171). ACM.

About the Authors

Raimondo Iemma

Raimondo Iemma holds a MSc in Industrial engineering from Politecnico di Torino. He is PhD candidate in "Management, production and design" at Politecnico di Torino. He performs research and policy support in the field of Open Government Data and innovation within (or fostered by) the Public Sector. He is Managing Director & Research Fellow at the Nexa Center for Internet & Society at Politecnico di Torino (DAUIN).

Federico Morando

Federico Morando is an economist, with interdisciplinary research interests at the intersection between law, economics and technology. He holds a Ph.D. in Institutions, Economics and Law from the Univ. of Turin and Ghent. He is the Director of Research and Policy of the Nexa Center for Internet & Society at Politecnico di Torino (DAUIN). From Dec. 2012, he leads the Creative Commons Italy project.

Michele Osella

Dr. Michele Osella is the Head of Business Model & Policy Innovation Unit at Istituto Superiore Mario Boella, being in charge of a multidisciplinary research team studying economic, social and policy implications of ICTs. In the academic sphere, he is Adjunct Lecturer in Innovation Management at the Polytechnic of Turin. Prominent strands of research on which he is focused are social computing, open data, smart city governance as well as Internet economics, with a keen interest for groundbreaking business models.