

## Chapter

# 5

## Information Systems based on (Linked) Open Data: From Openness to Innovation

*Sean W. M. Siqueira, Ig I. Bittencourt, Seiji Isotani, Bernardo P. Nunes*

### *Abstract*

*There is a trend towards making data and services available worldwide. As reuse is encouraged and transparency becomes a requirement, an important characteristic that has been discussed is providing open data and services. An evolving cycle is therefore created as new open data and applications based on the previous ones are developed. In addition, open data and services are moving towards a linked approach, which improves semantic expressivity and provides better machine processing. However, publishing and consuming linked (open) data and services in the Information Systems context present several challenges in the development, management and evolution of linked open data innovative ecosystems. In this chapter we present the context and challenges for these ecosystems and call the Information Systems community to embrace them.*

### **5.1. Introduction**

Open data is “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike” [Open Data Handbook 2016]. Government agencies seeking benefits in terms of public transparency and improved services to the population, in a pioneering way, support, fund and help to disseminate initiatives that leverage open publication and consumption of data.

Recently, open data initiatives have been encouraged by global movements, such as the campaign from the International Aid Transparency Initiative<sup>1</sup> on the transparency of spending records. It is also noticeable that the concept of public (or governmental) open data [Chignard 2013] is considered as a common property, just like scientific ideas. The Open Government

---

<sup>1</sup> <http://www.aidtransparency.net/>

Partnership<sup>2</sup> has made the open data initiative become a strong trend, and open data portals of several governments are available<sup>3</sup>, such as *data.gov*, *data.gov.uk*, *open.canada.ca*, *data.gov.ru* and *dados.gov.br*. The Brazilian Federal Public Administration (BAPF), through the Brazilian Access to Information Law<sup>4</sup>, in its 8<sup>th</sup> article, recognizes the need for making government data available in open format.

Although governments are publishing open data, they are using separate portals, in different locations and in multiple formats, hindering the access and use of information efficiently by users. The main issue cited as a barrier that limits the ability of stakeholders to interact with new information is the ability to search for and discover the right information [Bizer 2009]. As a result, it is necessary to focus on how to effectively generate, structure and search (linked) open data [Isotani and Bittencourt 2015].

However, once the data is open, further developments and actions are needed to solve problems, innovate and reach the full value potential. Therefore, it is necessary to invest in strategies, initiatives and tools to increase the availability of data open for exploitation by citizens, government agencies, industries, scientific communities and other stakeholders. According to Isotani and Bittencourt (2015), several companies, governments and research institutes have driven efforts to make data available and produce Web technologies that allow the creation of an ecosystem of data production and consumption with the objective of speeding up the discovery of new knowledge and adding value to any information freely available on the Internet. Moreover, they also point out that an ecosystem based on open data will only succeed when the data are adequately represented, structured and linked. It would allow both the automation of the production of these data and the consumption of them by machines, adding value and meaning to the data collected.

Open Data Ecosystems have not been formally defined as well as linked open data ecosystems or innovation<sup>5</sup> ecosystems (based on linked open data). The idea described by McKinsey (2009) is to transform open data into valuable tools. Gama and Lóscio (2014) consider the collaborative nature and the diversity of actors (government, application developers, small and medium enterprises, startups, civil society, universities, funding agencies and investors). They also consider the open-data-as-a-service (DAAS) platform as a fundamental component which could promote the integration of the different actors, allowing the development of products and services based on open data that can provide practical benefits to users. According to Davis (2011), an open data ecosystem could help identifying and evaluating possible

---

<sup>2</sup> <http://www.opengovpartnership.org/>

<sup>3</sup> More open government data catalogues can be found on <http://dataportals.org/>

<sup>4</sup> The law is available in Portuguese at: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/112527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm), more on the law and how to access information is available at: <http://www.acessoainformacao.gov.br/>

<sup>5</sup> According to the Oslo manual [Mortensen and Bloch 2005], in summary, innovation is an invention that generates value (financial/social). Technological process innovation is the adoption of technologically new or significantly improved production methods, including methods of product delivery. Then innovation ecosystems based on linked open data must focus on generating value from linked open data.

strategies that open data initiatives can adopt in seeking the realization of the promised benefits of open data. Lee (2014) describes the open data ecosystem elements as: data audit, dataset selection, address and map data, data privacy, licensing, publishing high-quality data, data access, data discovery, supporting public bodies, engaging data users, encouraging economic reuse and evaluation. However, it is also important to sustain the ecosystem, which is based on open data ecosystem principles (support, standardized, demand-driven, transparency and related to other open data initiatives) and challenges (privacy, governance, operational change and usage) leading to the open data ecosystem impact.

Furthermore, according to Shadbolt et al. (2012), although the issue of making data available for public use and creating an open data culture and mindset is of utmost importance, the open data movement is still in its infancy in terms of achieving the expected economic benefits. Companies and industry, seeking innovation and competitive advantages, have been making private data open as well as investing in programs that foster and exploit their use [Kapoor et al. 2015]. Therefore, with the increasing adoption of open data by different entities, the benefits obtained from the open data initiative go beyond improving government operations: open data are not only capable of providing public transparency, but also of generating information for better decision-making, promoting competitiveness and innovation [Isotani and Bittencourt 2015]. According to Lee et al. (2014), open data are a resource with value and promise for goods and services, considered a prerequisite for the creation of value [Chan 2013] and for the creation of innovative products, services and business models.

Kapoor et al. (2015) suggest it is necessary to further the creation of open data portals towards innovation ecosystems, which can eventually engage all the parts of the innovation process in discussions driven by data, through the identification of problems that might be solved and require actions to lead to innovative solutions. Indeed, the most difficult part of any innovation process is to find a solution that is desirable for the users, economically feasible and technologically doable, and therefore the search for solutions is worthwhile [Kappor et al. 2015]. As we seek to solve problems for cities, societies and businesses, we need to investigate strategies, tools and initiatives that provide a better understanding of the issue for all stakeholders in the ecosystem. In order to speed up the resolution of important problems, it is critical to invest in strategies for opening data for exploitation by citizens, government agencies, providers and other stakeholders. Innovation systems will provide tools, services, market data and applications, and collaborative environments that will enable public authorities, organizations and citizens to work together with open data and accelerate the creation of economic value.

Therefore, the central challenge proposed in this work can be summarized as: the development and management of Information Systems of innovation based on (linked) open data. It is a proposal that covers all the thematic and sub-areas of information systems since such systems will also require changes (evolution) in the types of information systems as well as new types of information systems, new information technology architectures, new procedures for the use of such systems, new business innovations, advances in the development of these systems, new cultures and behaviors.

## 5.2. Background

Tim Berners-Lee proposed the “5-Star rating system”<sup>6</sup>, which ranks the degree of data openness. The more open, the greater the number of stars for the data and the easier it can be enriched:

- i. Available on the web (whatever format) but with an open license, to be Open Data
- ii. Available as machine-readable structured data (e.g. excel instead of image scan of a table)
- iii. As (2) plus non-proprietary format (e.g. CSV instead of Excel)
- iv. All the above, plus: Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
- v. All the above, plus: Link your data to other people’s data to provide context

Therefore, open data is related to the linked data concept, which can be defined as a set of good practices for publishing and connecting (linking) Web-structured datasets with the aim of creating a “Web of Data” [Bizer et al. 2006]. These practices are based on Web technologies such as HTTP (Hypertext Transfer Protocol) and URI (Uniform Resource Identifier), in order to allow the reading of the linked data automatically by software agents. The Web of Data creates a number of opportunities for the semantic integration of the data itself, motivating the development of new types of applications and tools, such as browsers and search engines.

In this way, the Linked Data standards, recommendations, guidelines and good practices<sup>7</sup> allow anyone to publish the data in a way that can be read by people and processed by machines. This is possible because data that was previously “hidden” in the “Web of Documents” is now becoming accessible thanks to the use of the standards for linking data. As a result, everyone (humans and machines) may work more efficiently together.

The W3C Data on the Web Best Practices<sup>8</sup> [Lóscio et al. 2016] provides best practices related to the publication and usage of data on the Web designed to help to support a self-sustaining ecosystem: As data should be discoverable and understandable by humans and machines, the best practices facilitate interaction between publishers and consumers.

According to the ‘Data on the Web Best Practices’ [Lóscio et al. 2016], the main benefits of applying the best practices are:

- **Comprehension:** humans will have a better understanding about the data structure, the data meaning, the metadata and the nature of the dataset.
- **Processability:** machines will be able to automatically process and manipulate the data within a dataset.
- **Discoverability:** machines will be able to automatically discover a dataset or data within a dataset.

---

<sup>6</sup> <http://5stardata.info/en/>

<sup>7</sup> For following the news on the W3C activity on Data on the Web, please access: <https://www.w3.org/blog/data/>

<sup>8</sup> First public working draft published on February 24<sup>th</sup> 2015. Latest published version available at: <http://www.w3.org/TR/dwbp/>

- **Reuse:** the chances of dataset reuse by different groups of data consumers will increase.
- **Trust:** the confidence that consumers have in the dataset will improve.
- **Linkability:** it will be possible to create links between data resources (datasets and data items).
- **Access:** humans and machines will be able to access up to date data in a variety of forms.
- **Interoperability:** it will be easier to reach consensus among data publishers and consumers.

### 5.3. Related Worldwide Issues

The challenge is related not only to publishing and using (linked) open data, or the whole (linked) open data ecosystem, but also to its evolution to sustainable (linked) data open systems of innovation. In order to achieve such a scenario, it is necessary to advance on the state of the art and practice, not only from a technological point of view, but also from different perspectives such as government, business and community. Although it might be considered a worldwide challenge, the Brazilian IS research community can contribute and lead some of the solutions.

Despite the benefits of open data, there is a major problem when it is necessary to relate different data or when the amount of data is excessive for human interpretation. If the data were structured and linked in such a way that a computer could process them, the necessary tasks could be automated [Barros et al. 2011]. In this case, some issues arise, and can be summarized in how to manage, collect, model, standardize and consume data properly.

Cowan et al. (2014) list some issues related to open data:

- ensuring that all data that should be open is open
- finding and accessing open data
- providing the right tools to use open data
- keeping open data current
- ensuring privacy of individuals and property
- capturing open data sources
- supporting data redundancy
- sustaining the cost of storage, delivery and maintenance of open data
- governance of restricted data that is open for the purposes of a secure community of practice application.

Considering that the use of the open data serves some purpose for someone, Lindman et al. (2013) present some research questions regarding (linked) open data services from different perspectives: technologies, information, processes and activities, products and services, participants, customers and environment. Their questions cover several issues such as: availability, interoperability, heterogeneity of platforms as well as of users' devices and contexts, upgrades, security, privacy, standards' definition and compliance, timeliness, data format transformation, conceptualization of data and data services, business drivers, quality evaluation of apps, services categorization and application areas, motivations for data owners and providers, business models, service innovation and co-creation, incentives to developers, information asymmetry, skills and knowledge, description and discovery of services,

differentiation of services, and so on.

Open data initiatives (including open data ecosystems) focus on making data open and (re)useable. On the other hand, Kapoor et al. (2015) state that it is critical to invest in strategies that will open not only the data, but also the space of problems, i.e. open data ecosystem need to evolve to support the innovation workflow, thus becoming systems of innovation. Consequently, some issues arise, such as the need for analytics, social and collaborative capabilities, ideation, evaluation and creation of solutions, which would require a system of engagement and system of insight.

According to Kapoor et al. (2015), “systems of innovation will provide tools, services, data/app marketplaces and collaborative environments, which allow public authorities, organizations and citizens to work together in the open data landscape, and accelerate the creation of economic value”.

## **5.4. The Challenges**

This chapter addresses two main challenges that include the issues raised worldwide and specific Brazilian specific demands: (i) Enriched and Linked Open Data and (ii) System of Linked Open Data Innovation. Although complex, the latter challenge can be seen as a consequence (or evolution) of the former.

### **5.4.1. Enriched and Linked Open Data**

This challenge focuses on evolving open data portals (such as *dados.gov.br*) and portals of transparency in order to produce data with quality and improve transparency. It aims at improving open data quality and integration through a careful process of enrichment and data linking. In this challenge, all the issues related to linked open data ecosystems are considered.

Suggested projects:

- Integrated search and interoperable access to (linked) open data
- Systematic mapping of data portals (such as *dados.gov.br*, portals from several states and municipalities, as well as from industry) and their datasets
- Quality models for (linked) open data and open data services
- Quality analysis of the open datasets
- Conceptual framework for the Linked Open Data
- Development and evaluation of ontologies for open data enrichment
- Open data transformation to 4 and 5-stars
- Quality analysis of Linked Open datasets
- Open Data as a Service (DAAS) solutions
- Validation of DAAS solutions
- Analysis of acceptance and use of (linked) open data.

### 5.4.2. System of Linked Open Data Innovation

This challenge focuses on boosting the system of innovation based on linked open data. Therefore, it aims at providing the necessary infrastructure for supporting the investigation and understanding of problems and opportunities, then the actions to elaborate the solutions, the development of the related products and services, and finally the feedback and evolution of the system. In this challenge, all the issues regarding systems of innovation are considered.

Suggested projects:

- Analysis of requirements to systems of linked open data innovation
- Development of the workflow for a system of innovation
- The linked open datasets integration to the system of innovation
- Projecting the self-sustainable system of innovation
- Modeling and implementing the innovation indicators within the system of innovation
- Building a control and feedback mechanism within the system of innovation.

### 5.5. Progress Evaluation

The assessment of the progress of this challenge may consider: (a) monitoring the research evolution on the challenge (number of publications and their correspondence to the related goals and issues); (b) monitoring the dissemination of linked open datasets (data and services, including respective quality, use and innovative products and services); (c) monitoring the technological state of practice (software landscape evolution); (d) monitoring the innovation aspects (software registers and patents); and (e) establishment of standards, recommendations, guidelines and good practices within the context of this challenge.

Perhaps a maturity model for evaluation of the system of innovation based on linked open data could also support the assessment. However, in addition to considering the evolution of open datasets through the 5-star system, it is also necessary to analyze the degree of innovation it promotes.

### 5.6. Final Remarks

Once open data is disseminated, whether by government initiatives or industry, a new movement is expected to leverage public services (promoted by democratic governments or citizen demands) and technological advances that will propitiate the evolution of business and public management. The progress is aligned with the evolution of the open data towards the innovation promoted by linked open data, i.e. the system of innovation.

The proposed challenge is aligned to the “Brazilian grand challenge: productivity”. As presented in the reports of the 3<sup>rd</sup> Seminar of Grand Challenges in Computer Science Research in Brazil<sup>9</sup>, Prof. Silvio Meira suggested that “in order to become competitive in the world market, we must create and develop technology [...], aligning importance and relevance, in order to not waste time and resources developing useless technology and knowledge”. By investing in systems of innovation based on linked open data we have the chance to effectively accomplish

---

<sup>9</sup> <http://www.sbc.org.br/documentos-da-sbc/send/141-grandes-desafios/802-grandesdesafiosdacomputaonobrasil>

this. The availability of open data in Brazil, together with services that make effective use of these data, will promote innovation generated by the demand of citizens, government agencies, industries, scientific communities and other stakeholders. Open data is mentioned in several parts of these reports.

Regarding the report on the Grand Challenges in Computer Science Research in Brazil – 2006 – 2016<sup>10</sup>, the challenge described in this manuscript advances the grand challenges including:

1. **Information management in large volumes of distributed multimedia data:** the open data proposal comes precisely to support the management of information in large volumes of data available on the Web, in order to promote better structuring, representation, access and consumption of data.
2. **Computational modeling of artificial, natural and socio-cultural complex systems and the human-nature interaction:** the availability of data on the Web as linked open data has promoted the development of new computational models of complex systems.
4. **Participatory and universal access for Brazilian citizens to knowledge:** it is perhaps the challenge of computing that will be affected most by linked open data, since the availability of such data will allow and promote greater access and participation for Brazilian citizens to knowledge.
5. **Technological development of quality:** available, correct, secure, scalable, persistent and ubiquitous systems: the linked open data will promote research on environments, methods, techniques, models, devices and architectural and project standards capable of assisting designers and developers of large software and hardware systems to achieve these goals.

## References

- Barros, H., Silva, A., Costa, E., Bittencourt, I. I., Holanda, O. and Sales, L. (2011) “Steps, techniques, and technologies for the development of intelligent applications based on Semantic Web Services: A case study in e-learning systems”. *Engineering Applications of Artificial Intelligence*, 24(8): 1355-1367.
- Bizer, C. (2009) “The emerging web of linked data”. *IEEE Intell. Syst.*, 24(5):87-92, Sept.-Oct.
- Bizer, C., Heath, T. and Berners-Lee, T. (2006) “Linked Data - The story so far”. Heath, T., Hepp, M. and Bizer, C. (eds). *International Journal on Semantic Web and Information System*, Special Issue on Linked Data, 2006.
- Chan, C. M. L. (2013) “From open data to open innovation strategies: Creating e-services using open government data”. In *Proc. Hawaii Int. Conf. Syst. Sci.*, p. 1890-1899.
- Chignard, S. (2013) “A Brief History of Open Data”. *ParisTech Review*, March 29th.
- Cowan, D., Alencar, P. and McGarry, F. (2014) “Perspectives on Open Data: Issues and Opportunities”. *IEEE International Conference on Software Science, Technology and Engineering (SWSTE)*, p. 24-33.
- Davies, T. (2011). *Open Data: infrastructures and ecosystems*. Open Data Research.

---

<sup>10</sup> <http://www.sbc.org.br/documentos-da-sbc/send/141-grandes-desafios/799-grandesdesafios-ingles>



- Gama, K., & Lóscio, B. F. (2014). Towards Ecosystems based on Open Data as a Service. In ICEIS (2) (p. 659-664).
- Isotani, S., Bittencourt, I. I. (2015) “Dados Abertos Conectados: Em Busca da Web do Conhecimento”. São Paulo: Novatec. 176p.
- Kapoor, S., Mojsilovic, A., Strattner, J. N. and Varshney, K. R. (2015) “From Open Data Ecosystems to Systems of Innovation: A Journey to Realize the Promise of Open Data”, Bloomberg Data for Good Exchange Conference.
- Lee, D. (2014, October). Building an open data ecosystem: an Irish experience. In Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance (p. 351-360). ACM.
- Lee, M., Almirall, E. and Wareham, J. (2014) “Open data & civic apps: 1st generation failures”. 2nd generation improvements. Working Paper 256, ESADE.
- Lindman, J.; Rossi, M. and Tuunainen, V. K. (2013) “Open Data Services: Research Agenda”. 46th Hawaii International Conference on System Sciences.
- Lóscio, B. F., Burle, C. and Calegari, N. (2016) “Data on the Web Best Practices”. W3C Recommendation candidate. World Wide Web Consortium. <http://www.w3.org/TR/dwbp>
- Mortensen, P. S., & Bloch, C. W. (2005). Oslo Manual-Guidelines for collecting and interpreting innovation data. Organization for Economic Cooperation and Development, OECD.
- Open Data Handbook (2016) “What is open data?” In Open Data Handbook, <http://opendatahandbook.org/guide/en/what-is-open-data>
- Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W. and Schraefel, M. (2012) “Linked open government data: Lessons from data.gov.uk”. IEEE Intell. Syst., 27(3):16-24.



### **Sean Wolfgang Matsui Siqueira**

CV: <http://lattes.cnpq.br/2562652838103607>

DSc in Computer Science, Sean is an associate professor at the Federal University of the State of Rio de Janeiro (UNIRIO). He is editor-in-chief of the RBIE: Brazilian Journal on Computers in Education and co-editor of the iSYS: Brazilian Journal of Information Systems. He is the founder and coordinator of the Semantics and Learning research group and is a member of the special committees on Computers and Education (CEIE) and on Information Systems (CESI), both from the Brazilian Computer Society (SBC). His research interests are knowledge representation, web science (including social and semantic web) and advanced technologies for teaching and learning. e-mail: [sean@uniriotec.br](mailto:sean@uniriotec.br)

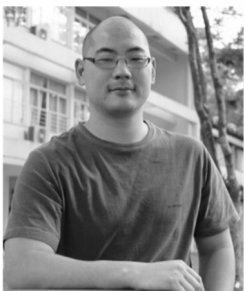


### **Ig Ibert Bittencourt Santana Pinto**

CV: <http://lattes.cnpq.br/4038730280834132>

DSc in Computer Science, Ig is an adjunct professor at the Federal University of Alagoas. He is president of the Special Committee on Computers and Education (CESI) from the Brazilian Computer Society (SBC), SBC delegate in the IFIP TC on Education (TC3) and UFAL advisory representative at W3C. Ig is founder and Co-Director of the Center of Excellence in Social Technologies (NEES.

He founded two startups awarded for innovation, one in the area of Technologies in Education (MyTutor) and another in the area of Semantic Technologies (LinKn). Ig has worked with research, development and innovation in the areas of Educational Technologies, Linked Open Data and Social Entrepreneurship. e-mail: [ig.ibert@ic.ufal.br](mailto:ig.ibert@ic.ufal.br)



### **Seiji Isotani**

CV: <http://lattes.cnpq.br/3030047284254233>

DSc in Information Engineering, Seiji is an associate professor and vice-head of the Computer Systems Department of the University of São Paulo (ICMC-USP). He is the founder and co-coordinator of the Laboratory of Applied Computing to Education and Advanced Social Technology (CAEd). He is also co-founder of two startups (MyTutor and Linkn). He has done research on the following topics:

intelligent tutors, CSCL, gamification, ontology engineering, linked open data, semantic web and educational data mining. He is ACM Distinguished Speaker, IEEE Senior Member, CNPq Productivity Fellow and Associate Editor of IEEE Transactions on Learning Technologies. e-mail: [sisotani@icmc.usp.br](mailto:sisotani@icmc.usp.br)



### **Bernardo Pereira Nunes**

CV: <http://lattes.cnpq.br/1728746187630338>

DSc in Computer Science, Bernardo is researcher and adjunct professor at the Pontifical University Catholic of Rio de Janeiro (PUC-Rio) and collaborator at the Graduate Program of Informatics at the Federal University of the State of Rio de Janeiro (UNIRIO).

He also works as senior systems analyst at the Center of Distance Learning, PUC-Rio and is associate editor of the RBIE: Brazilian Journal on Computers in Education. He has experience in Computer Science, focusing on Semantic Web, Web Science and Distance Learning. e-mail: [bernardo@ccead.puc-rio.br](mailto:bernardo@ccead.puc-rio.br)