Proceedings of a pre-conference workshop of the 27th International Cartographic Conference:
Spatial data infrastructures, standards, open source and open data for geospatial (SDI-Open 2015)
20-21 August 2015, Brazilian Institute of Geography and Statistics (IBGE), Rio de Janeiro, Brazil

Discussing the role of university in spatial data infrastructure construction: issues and challenges for UERJ-V-SDI

José Augusto Sapienza Ramos, Carlos Eduardo Gonçalves Ferreira
{sapienza, carlos.ferreira}@labgis.uerj.br

LabGIS System - Research Center of Geotechnology of Rio de Janeiro State University (UERJ)
Multidisciplinary Institute of Human Formation with Technology of UERJ

Keywords: academic community, data producer, final-work geospatial dataset, value-added data, volunteer, open data.

1. Introduction

In the computer era, an unprecedented amount of data is being produced. This data is then organized in catalogs or repositories called "data infrastructures" (Kitchin, 2014), where information should be easily retrieved in order to be useful. It is noticeable that the performance of any new information paradigm has an impact over science and society, increasing and spreading the human knowledge more rapidly.

Geographic data can be defined as the way in which the knowledge of the Earth's surface is represented (Goodchild, 2011). Advancements in technology as well as the new paradigm in geographical information science and systems allow the production, visualization, analysis and sharing of large amounts of geographic data suitable to different kinds of audiences. To achieve that end proper scientific and technological debates are required.

With that in mind concepts such as "data infrastructure", "open data", "big data", "open governmental data" are being used to describe some of the new aspects of today's society – see Kitchin (2014) for a survey. On the other hand, data production and its usage have been decentralized from governments and companies to include the population in general, giving birth to terms such as "neocartographers" (Liu and Palen, 2010) and "Volunteered Geographic Information" (VGI) (Budhathoki and Nedovic-Budic, 2008).

Data infrastructure (DI) *"is the institutional, physical and digital means for storing, sharing and consuming data across networked technologies"* (Kitchin, 2014, p. 32). Following O'Carroll et al. (2013) and Kitchin (2014), the DIs can be divided into: (a) *data holdings*, as informal collections in data files; (b) *data archives*, as formal collections of data that are structured, curated and documented; (c) *catalogues, directories and portals*, as centralized resources linking different data holdings and archives; (d) *data repositories* which aim to ensure that each archive or holding meets a specific set of audited requirements in order to validate data integrity and ensure trust, where repositories can be single-site or multi-site; and (e) *cyber-infrastructures*, as a suite of dedicated and integrated hardware and network technologies, including interoperable software and middleware services, shared services, analysis tools, data visualization and shared policies (Cyberinfrastruture Council, 2007).

By its turn, the spatial data infrastructure (SDI) can be understood as a DI that incorporates geographic data and its specific technologies and standards, such as spatial databases, geovisualization, geoservices or geospatial metadata. At the beginning, the SDI discussion grew with a focus on implementation of cyber-infrastructures in federal

governments (National SDI or NSDI) and in larger companies. However, today, the SDI initiative spreads into other kinds of sectors, and the epistemological debates are changing to introduce, for instance, its social impacts and decentralized data production such as VGI – see Dessers (2012) for a survey of concepts and history.

2. Placing this case

The University environment is an important data producer, which includes geographic data based on scientific approach. For instance, the Sirius Network – a network of libraries from Rio de Janeiro State University (UERJ) – had, in 2013, a collection of maps with 3,896 titles and 3,553 final work documents in Engineering and Natural Sciences (DATAUERJ, 2014), covering both analog and digital spatial datasets produced and documented in several formats. The majority of higher-level education institutions (if not all) create important data repositories to store academic documents such as thesis and papers generated by their students. However, few institutions maintain similar repositories to gather another type of academic production: the spatial data created by the very same students, which results in disperse data holdings and archives (with high data loss) that make these databases unfit for further use in academic research, by governments, by businesses or by any other potentially interested sectors.

In this scenario, the public universities have a considerable spatial data production in Brazil, as they centralize an important part of the academic production in fields such as geosciences and engineering, among others. According to *Plano de Ação da INDE* (BRASIL, 2010), the Brazilian NSDI, namely Infraestrutura Nacional de Dados Espaciais (INDE) in Portuguese, the academic sector is regarded as an agent *"responsible for fomenting and developing education, qualification, training and research in SDI"* (p. 61), i.e., limited to the role of non-producer of data. The same document defines that some producers of value-added data must be additionally identified among all SDI agents, where *"it is expected that the weight and the agent's participation [as data producers] will increase considerably with the INDE evolution"* (p. 66). Today, it is noticeable that the academic sector is not perceived as a data producer by the INDE or in itself.

The "Open Government Working Group" (2015) stipulates the principles of Open Government Data, highlighting that it must be complete, primary, timely, accessible, machine-processable and license-free; the access must be non-discriminatory, and the compliance reviewable. The geographic data produced in a public university as UERJ should also follow these principles whenever possible. This way, data shall be open, so that it can be freely reused and redistributed among citizens (Pollock, 2006).

This work discusses and proposes the UERJ-V-SDI (UERJ Volunteer SDI) dedicated to the student community of UERJ, which is adherent to INDE standards, with its database searchable through the INDE portal. At the beginning, the academic community is sharing its spatial data voluntarily. However, it is expected that the university will develop further the discussion of an efficient data infrastructure for its production of spatial data. The proposal presented here is not definitive, but rather an initial discussion on this matter. An SDI is dynamic, and should evolve as the discussion evolves, but this first step is fundamental and urgent.

3. UERJ-V-SDI Components

Aiming towards the goals described above, a SDI can be addressed in different ways, and it is common to find in the literature its division into different components and objectives (Rajabifard, 2008; Dessers, 2012). This paper follows the major division proposed in Warnest (2005), which is adherent to INDE components. The Figure 1 shows the division used in this work, i.e., to an academic context.

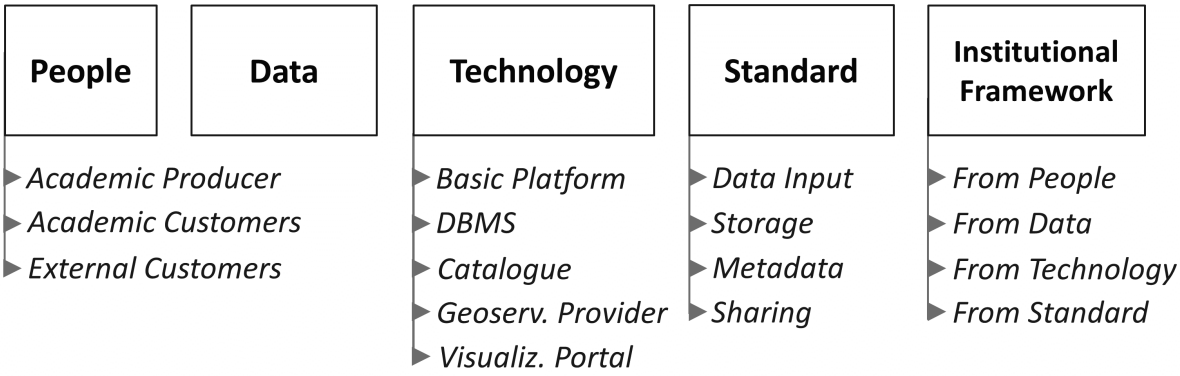

Figure 1 – UERJ-V-SDI components.

You will find below brief explanations about each component, considering the space constraints of this paper. The *People component* is divided into producers and customers. The customers can be diverse, that is, sets of users with different needs. This SDI may focus on academic and governmental users. The data producers are any university member that generates geographic data upon activities of teaching, research or extension. This way, it is necessary to identify potential data holdings and archives inside of the university's departments to seek an incremental integration with Sirius Network's databases. On the other hand, students, professors, directors and other members must be aware of the benefits and values behind their volunteer participation on SDI, since volunteers will produce new databases with the necessary procedures and documentation as metadata, for instance.

The *Data component* can be very diverse, not following a unique and rigid conceptual model, considering that countless themes and methodologies can be used. Furthermore, standards for character set, reference system, field names, data format, among others, need to be defined, and quality control encouraged.

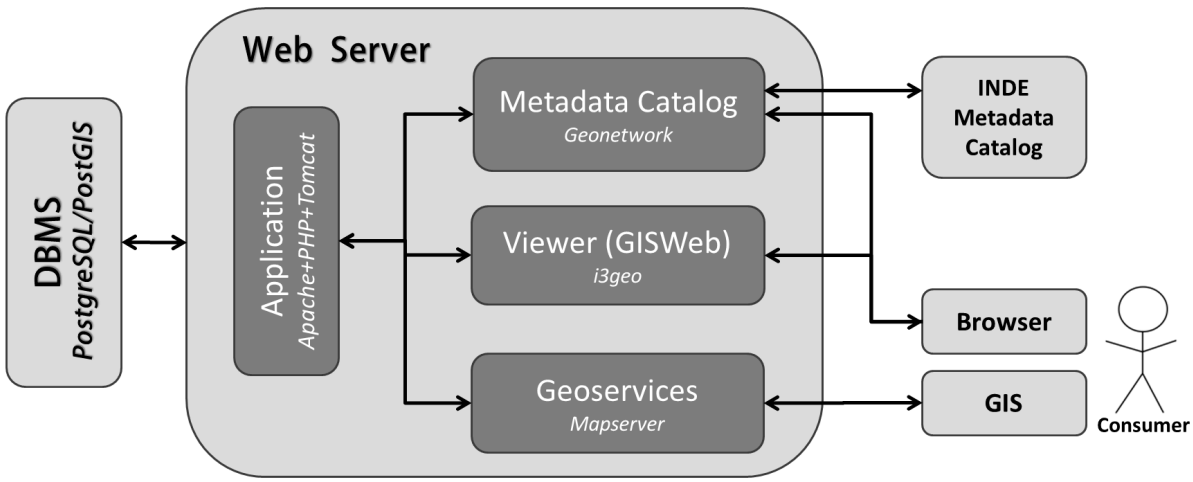

Figure 2 – The technological platform and its main communications.

The *Technology component* defines the network solutions that implement the digital means for storing, sharing and consuming data – see Figure 2. Note that this platform is fully open source, and the user can access the data using browsers or a GIS which implements the OGC standards. The user can discover the database through a metadata catalog, thus visualizing and analyzing the relevant data using the GISWeb or geoservices by interoperable means. In addition, the CSW (Catalogue Service for the Web) protocol is applied to share the metadata database with the INDE metadata catalog, thus enabling searches for the INDE catalog to return results in UERJ-V-SDI database. It is noteworthy that Figure 2 omits some communications between adopted solutions, showing the main links.

It is necessary to adopt standards for each step of the data lineage of the SDI, namely: input, storage, catalogue, access and sharing. Thus, the *Standard component* defines directives which tell how some activities of the SDI must be performed in a coordinated way. So far, the main standards adopted are: OCG standards as WMS, WFS, WCS, KML, and CSW to Web services, the core metadata of ISO 19115:2003, and format standards to store the data using the same reference system and character set, among others.

At last, the *Institutional component* can be defined as the driver force which maintains and guides the SDI and its components towards the current goal. In other words, the targets and policies must be frequently updated to reflect the current needs of the SDI. As of today, this component recognizes challenges to: (a) identify spatial data pools in the university; (b) identify and recognize data producers as end-of-course students or key professors who could volunteer; (c) promote the debate about the organization of spatial data production inside the university environment and the SDI values (as this paper); (d) put together an initial and valuable database; and (e) promote the university as an important data producer.

4. Conclusion

Data infrastructures are technological platforms to ease data retrieval, whereas SDI initiatives show the way governments and others groups enforce cyber-infrastructures which create interoperable environments to access and share the geographic information. When the university keeps spatial data produced by its members in data holdings or data archives, part of its intellectual production becomes unavailable or lost for the internal or external community. Therefore, the public university and several government agencies in Brazil must recognize this academic database as more relevant for the INDE objectives – it is necessary to promote a larger debate.

When the university shares its spatial data more efficiently as open data, it spreads its production deeper in society, fulfilling its social role. However, SDI proposals in Brazilian universities might face the same problems of several SDIs around the world, which are not of a technological nature, but rather related to the awareness of institutional parties. Because of that, this paper is sometimes more about context than technology.

This proposal is at its beginning, and needs to be discussed further, where the volunteer approach is a strategy to disseminate the SDI idea among the university environment and in external communities.

References

BRASIL – Ministério do Planejamento, Orçamento e Gestão 2010, *Plano de Ação para implantação da INDE: Infraestrutura nacional de dados espaciais*, Comissão Nacional de Cartografia (CONCAR), Rio de Janeiro.

BUDHATHOKI, NR, & NEDOVIC-BUDIC Z 2008, 'Reconceptualizing the role of the user of spatial data infrastructure', *GeoJournal*, v. 72, n. 3-4, pp. 149-160.

CYBERINFRASTRUCTURE COUNCIL 2007, *Cyberinfrastructure vision for 21st century discovery*, National Science Foundation. Available from < www.nsf.gov/pubs/2007/nsf0728> [16 March 2015].

DATAUERJ, *Anuário Estatístico 2014*. UERJ – Núcleo de Informação e Estudos de Conjuntura – UERJ/NIESC/VR, Rio de Janeiro. Available from <http://www2.datauerj.uerj.br/> [16 March 2015].

DESSERS, E 2012. *Spatial Data Infrastructures at work. A comparative case study on the spatial enablement of public sector processes*. PhD Thesis, Leuven University.

GOODCHILD, MF 2011, 'Challenges in geographical information science, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Science*, The Royal Society A 467(2133), pp.2431–2443.

KITCHIN, R 2014, *The data revolution: Big data, open data, data infrastructures and their consequences*, Sage, London.

LAURIAULT, TP, CRAIG, BL, TAYLOR, DR, & PULSIFER, PL 2008. *Today's data are part of tomorrow's research: Archival issues in the sciences*, Archivaria, n.64.

LIU, SB & PALEN, L 2010, 'The new cartographers: Crisis map mashups and the emergence of neogeographic practice', *Cartography and Geographic Information Science*, 37(1), pp. 69-90.

O'CARROLL, A, COLLINS, S, GALLAGHER, D, TANG, J & WEBB, S 2013, *Caring for Digital Content: Mapping International Approaches,* NUI Maynooth/Trinity College Dublin/Royal Irish Academy and Digital Repository of Ireland, Dublin.

OPEN GOVERNMENT WORKING GROUP 2015, *Principles of open Government data*. Available from: <www.openGovData.org> [16 March 2015].

POLLOCK, R 2006, *The value of public domain*, IPPR. Available from <http://rufuspollock.org/papers/value_of_public_domain.ippr.pdf> [16 March 2015]

RAJABIFARD, A 2008, A Spatial Data Infrastructure for a Spatially Enabled Government and Society in *A Multi-View Framework to Assess Spatial Data Infrastructures*, eds CROMPVOETS, J, RAJABIFARD, A, VAN LOENEN, B & FERNÁNDEZ, TD, Wageningen University, RGI, pp.

WARNEST, M 2005, *A collaboration model for national spatial data infrastructure in federated countries*. Ph.D. Thesis, University of Melbourne.