Open Geospatial Science - Calling for a Digital Earth Laboratory in Europe

Sven Schade
European Commission – DG Joint Research Centre
Institute for Environment and Sustainability
Via E. Fermi, 2749 – 21027 Ispra (VA), Italy
sven.schade@jrc.ec.europa.eu

ABSTRACT

Nowadays, geospatial data management and handling facilities are supposed to break existing silos, address new audiences, and meet challenges of increased data variety, velocity and volume. A next generation of (digital) science and supporting e-infrastructures has to emerge in order to provide the required capabilities and capacities, including service interoperability, transparency, repeatability of experiments and reproducibility of scientific findings. This paper calls for a 'Digital Earth Laboratory' as a means to exploit the major challenges, share observations, experiments and their results, and facilitate collaboration in an open environment. We particularly (i) argue for a European focus, building on initiatives such as INSPIRE, GEOSS and Copernicus; (ii) sketch the early development status; and (iii) post major questions which will have to be addressed for a stepwise realization of our proposal.

Keywords

Digital Science, Open Science, Digital Earth, Geospatial Informatics, Experimentation, Laboratory, Engagement

1. INTRODUCTION

Especially in the European context, geospatial data organization and processing has come a long way. Driven by initiatives including the Infrastructure for Spatial Information in the European Communities (INSPIRE), the Global Earth Observation System of Systems (GEOSS) and the Copernicus programme, data and services get more available, discoverable and accessible. At the same time, sustainable development goals – which are defined on the global level – increasingly requests cross- and multi-disciplinary research [1, 2]. This implies that geospatial data management and handling facilities should not be offered in closed silos. New audiences – especially including individual citizen and civic associations – have to be addressed, and challenges of increased data variety, velocity and volume (aka Big Data [3]) need to be met. A next generation of science and supporting research infrastructures has to emerge in order to provide the required capabilities, including interoperability, transparency, repeatability of experiments and reproducibility of scientific findings, as well as a suitable learning environment.

Apart from detailing the arising issues, in this paper, we call for a 'Digital Earth Laboratory' as a means to exploit the major challenges, share observations, experiments and their findings, and facilitate learning and collaboration in an open environment. We particularly argue for a European focus, building on initiatives such as INSPIRE, GEOSS and Copernicus. At this initial realization phase, underlying principles – such as reducing repetition of existing open solutions or avoiding interference with existing community management – are presented, early developments are sketched, and core questions are posted.

Notably, we interpret 'open' in two ways. On the one hand, we tap into open source (geospatial) software, open data and open science. One the other hand, we investigate the implications to open the geospatial sector and scientific communities to other sectors (non-spatial) and stakeholders (non-researchers). We hope to simulate discussion so that the vision of the Digital Earth Laboratory can be further shaped, future developments can be aligned to ongoing activities and collaborations can be initiated.

Whereas the next section of this work in progress paper details the overall context, Section 3 presents our interpretation of open (geospatial) science and some of the implication this might have on future research and development. Section 4 then presents the concept of the Digital Earth Laboratory as one possible way to move ahead, just before we list major questions in Section 5.

2. BACKGROUND

The process around the INSPIRE directive – which by now completed all the legal grounds and entered the maintenance and implementation phase, partially even beyond the 28 Member States of the European Union – accelerated the establishment of spatial data infrastructures (SDIs), not only at national level. Information regarding the 34 themes of the directive become increasingly available for discovery, view and download. Activities for the evolution of INSPIRE into 'foreign' policy areas have been initiated – beginning with the air quality pilot, which was presented last year [4] and initiated the establishment of an INSPIRE download service for observations and measurements [5]. Simultaneously, GEOSS could increase its data sharing capacities enormously by changing its common infrastructure from a centralized and federated to brokering approach [6]. Just a few weeks ago, Sentinal-1A – the first satellite collecting Earth Observation data for Copernicus – was successfully launched [7]. It will soon help to deliver massive amounts of weather day-and-night imagery of Europe in near-real-time. These three initiatives push Europe into a leading position considering geospatial data acquisition, storage and processing.

In parallel, underlying technologies are changing rapidly and major governmental decisions have been taken in the last years. Both phenomena strongly influence the way we have to advance organizational structures, the available content, and supporting infrastructures. Novel collaborative environments, the realization of the Internet of Things (IoT), and improved use of mobile devices became top research issues – alongside renewable technologies and food science [8]. All of these areas are complemented by governmental requests for open data and sustainable development. The G8 Charter on Open Data [9], the rapid growth of the Open Government Partnership [10], the Digital Agenda for Europe [11] and the establishment of the Research Data Alliance (RDA) [12] provide clear directions on future accessibility and the treatment of intellectual property rights. Concurrently, sustainable development goals have been reaffirmed by the United Nations in 2012 [1], and initiatives such as Future Earth [2] or Global System Science [13] begin to address the arising challenges.

All in all, major achievements have been made and we continue to progress on the sustainability agenda. However, the issue (or hype) of Big Data, as well as the co-design, co-production and co-delivery paradigm, still have to be better understood, and used to evolve – or in some cases revolutionize – existing scientific and political processes. In the next sections, we investigate how open geospatial science could contribute to this bigger picture, and we propose an approach for addressing the most eminent issues. Our argumentation directly matches with the concepts of e-Infrastructures [14] and Digital Science [15], which are both promoted under the before mentioned Digital Agenda for Europe.

3. OPEN GEOSPATIAL SCIENCE

Ultimately, we have to optimize knowledge transfer. This should allow cross- and multi-disciplinary research, i.e. we have to open the existing knowledge silos to enable the use of geospatial science in other disciplines, as much as we have to allow for the integration of 'foreign' data, models, software, etc. inside the geospatial and environmental sector. Regardless of the direction, this has to include community-targeted training materials.

At the same time, scientists have to leave their ivory towers. Increasing requests for innovation and user engagement have to be met by extending the audience from purely scientific communities to policy makers, industry and public citizen. This spared of (active or passive) interactions calls for new forms of coordination, including changing communication, information flows and integration needs – within and across domains.

Thirdly, and in addition to increasing data velocity and volume within our discipline, both of the above contribute to the growing variety of data offerings. Historical data centers of the environmental and earth sciences will have to be connected to social sciences' repositories, hubs for accessing information from the humanities, as well as citizen engagement adds novel data structures into the this common space. Analysis and visualization capabilities have to be provided for this new level of size and complexity.

Many of the requested capabilities are already available today. After decades of successful societal engagement in domains such as biodiversity, citizen science nowadays sees record numbers of projects and participants. The European Citizen Science Association (ECSA) just was established and keep forming their agenda [16]. With global platforms, including Zooniverse [17] and Crowdcrafting [18], interested layman can now contribute to many scientific areas. Five Citizens' Observatories are funded to provide user contributed content to GEOSS, too [19]. Simultaneously, the European Observation Network for Territorial Development and Cohesion (ESBON) programs contributes significant to the territorial dimension of sustainable development in Europe [20], while large scale digital research infrastructures are established worldwide – see, for example, European Strategy Forum on Research Infrastructures (ESFRI) [21], EarthCube in the US [22], and the Australian Urban Research Infrastructure Network (AURIN) [23]. A Reusable INSPIRE Reference Platform (ARE3NA) supports access to common reusable software and other components for spatial data in European e-government [24]. Initiatives such as the Mozilla Science Lab [25] address issues of open publishing and learning across disciplines, and more specialized networks – much notable the Geo4All initiative that is powered by OSGeo and the International Cartographic Association (ICA) – want to make education and opportunities accessible to the widest possible audience [26].

We are dealing with a rich and yet growing research area, in which amazing progress is made in Europe and across the globe. Fully acknowledging the contributions of all the above (and many others), we can only try to imagine how powerful their combinations might become for advancing open geospatial science. With the mostly overlapping interests and common objectives across various ongoing initiatives, the required connection points already exist, but environments for establishing cross-walks and joining available forces are yet to be provided. Notably, this does not propose to integrate all the existing capacities, but request a facility to easily connect between, jointly learn from, and collaborate with existing entities – if desired.

Focusing again on the European context, this might be directly embedded into the Digital Science policy and the evolution of e-Infrastructures. *Digital Science* considers the transformation of science through Information and Communication Technologies (ICT) tools, networks and media, to make research more open, global, collaborative, creative and closer to society [15]. Hence, it provides a frame for the desired changes in the geospatial domain. The exact fitting still has to be detailed, including the establishment of (i) transparent, (ii) repeatable, and (iii) reproducible scientific findings in the geospatial domain.

ICT-based infrastructures (aka *e-infrastructures*) provide common technologies to a broad range of research areas [14]. Here, we imagine fundamental enablers for geospatial and environmental data management, processing and visualization. These might, for example, include 2D and 3D mapping applications, harmonized access to geo-sensor networks, components handling geo-statistics, spatial interpolation and forecasting, as well as geographic reference data and spatial indexing – two essential assets for data integration. Multi-sensor networks and scientific methods, should be considered as much as distributed storage and computing resources. Most eminently, service interoperability has to be achieved beyond the geospatial domain, i.e. somewhat 'closed' SDIs have to be able to interact with mainstream web services, and specific solutions of other scientific areas. The growing number of data and service inter-connections raise the importance of sustainable access, and – above all – persistent identifiers and resolvers. An open learning environment has to be developed in parallel, so that future contributors (geospatial experts and novices) can acquire the necessary skills.

4. DIGITAL EARTH LABORATORY

In order to build on, join and evolve the rich – yet distributed – European expertise on central issues of open geospatial science for sustainable development, we suggest to establish a 'Digital Earth Laboratory' as a means to exploit the major challenges, share observations, experiments and their findings, and to facilitate learning and collaboration. We envisage such a lab to be an open place in which individuals or already established communities could meet, get an integrated view on their joint assets, and are supported to work together. The facilitating services should be available to all parties that are interested to join forces in the spirit of open geospatial science. This might cover any temporal scale, i.e. from a few of days up to decades.

Considering the local level of sustainable development, usage scenario could, for example, include (a) the synchronized organization, execution and reporting on a 24-hour BioBlitz [27] in a network of smart cities; (b) a medium-term collaboration between citizen science initiatives and smart city networks to exploit new indicators for quality of life in urban areas; or (c) the long-term curation of the outcomes of small- and medium-scale citizen science projects for future re-use.

We see three essential pillars for realization (Figure 1), which basically represent the central parts of any scientific process – observation, experimentation and dissemination:

- An *observatory*, in which data and services from gathering parties could be made discoverable and accessible for common visualization and integrated analysis;
- A *laboratory*, in which joint experiments can be carried out and reported upon, and where existing physical laboratories could become virtually connected; and
- A *showroom*, in which findings can be shared and any kind of scientific material can be curated for long-term – adopted to address multiple audiences.

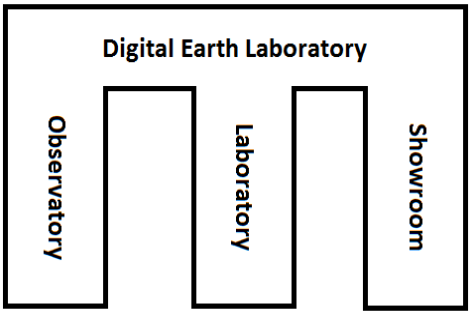

Figure 1: Fundamental pillars of a digital earth laboratory.

In their combination, these pillars address some of the major challenges of open geospatial science. Sustainable data analysis and visualization is supported by the shared data access via the observatory, together with the hosting of repeatable experiments in the laboratory. Societal engagement can be achieved by incorporating user contributed content into the observatory and by including citizen directly in the laboratory experiments. Science communication is the central aim of the showroom.

Some guiding principles should help future developments. First and for most, we have to escape interferences with existing community management. This also implies to pose minimal – in the best case no – change requests to existing systems and work flows. Furthermore, repetition of existing open solutions should be avoided as much as possible, i.e. neither should we replicate existing open (geospatial) software, nor should we attempt to establish networks and communities that are already existing elsewhere. In terms of development process, we shall begin with a few rather small experiments and expend the laboratory when it proves useful. (We will never know if we do not try!) Last but not least, the Digital Earth Laboratory has to be open by principle, i.e. results should be fully accessible and repeatable, used inputs and models should be made available and developed code, as well as other resources, should be provided. Apart from these obvious needs, the material that is required to understand and use the provided resources has to be offered, too. All in all, we will have to build an 'un-platform', in the sense that we require an environment that does not share all the characteristics of current platform approaches, as it has to avoid approaches that are specific to a certain domain or community.

While all of the above is independent of any location, we see particular benefits in providing a Digital Earth Laboratory in Europe. Apart from potential funding opportunities on this level, INSPIRE and Copernicus will continue to provide solid grounds in terms of geospatial data and services. In addition, the above mentioned ESBON and ESFRI, together with other large scale initiatives (see for example the European Network of Living Labs (ENoLL) [28]) offer an ideal context for expansions and inter-disciplinary cross-walks. Addressing the macro-regional level also provides the required bridge between global initiatives (such as GEOSS and Future Earth) and national and local activities (including the smart city movement).

But how should we approach this new scientific era? Which issues will we face on the organizational dimension, for example, concerning diverse responsibilities, mandates and trust-levels? And which new technologies will appear on the horizon? Clearly, while framing the surrounding environment, experiments with emerging technologies and foresight activities have to be carried out continuously. A dual approach of conceptual and strategic shaping, together with hands-on experiences for implementation is needed. As a start, we currently (1) investigate alternative storage possibilities with Hadoop [29] and MongoDB [30]; (2) use Kowalski [31] and other data mining tools for social media to complement data and processing capacities from GEOSS, INSPIRE and others; (3) set up a knowledge hub for curating the results of European citizen science projects; and (4) apply the overall approach for defining and investigating indicators for urban quality of life as a flagship application.

5. FINAL REMARKS AND QUESTIONS

We introduced our approach to address the increasing requests for open, transparent, and reproducible research in the earth and environmental sciences. We suggested to meet requirements such as sustainable data analysis and visualization, science communication, and societal engagement, using the notion of a Digital Earth Laboratory. We particularly argued for a European focus, building on initiatives such as INSPIRE, GEOSS and Copernicus. The initial state of development was briefly sketched.

With this, we aim at early feedback regarding this overall approach and constructive inputs for improvement. We hope to simulate discussion so that the vision of the Digital Earth Laboratory can be further shaped, future developments can be aligned to ongoing activities and collaborations can be initiated.

In order to progress, we should answer questions such as:

- Which are best practices to implement open science? And how can we embed the geospatial sector into the broader picture?
- Which are the most important experiments (either essential for advancing open geospatial science or quick-wins) to be carried out in order to advance the Digital Earth Laboratory?
- Which are the key communities to address as potential early adopters?
- Which potentially disruptive technologies are coming up within and outside the geospatial sciences?

Before closing, it should be noted that our presentation promotes cross- and multi-disciplinary collaborations. Still, the need for intra-disciplinary progress should not be forgotten. Future research and development has to continue to also account for advances within the geospatial and environmental sectors, but increasingly account for the opening of existing and novel processes.

ACKNOWLEDGMENTS

Our thanks to colleagues in the Digital Earth and Reference Data unit of the European Commission's Joint Research Centre (JRC) for the many discussions, which helped shaping this paper.

REFERENCES

[1] United Nations. 2012. *Future We Want - Outcome document*. A/RES/66/288.

[2] Official web page of the *Future Earth Initiative*. http://www.futureearth.info/.

[3] Douglas, L. 2001. *3d data management: Controlling data volume, velocity and variety*. Gartner.

[4] Kotsev, A. 2013. INSPIREd solutions for Air Quality problems. Atmosys final conference.

[5] Jirka, S. 2014. 52°North Supports the JRC in Developing an INSPIRE Download Service based on SOS. 52°North blog entry, 30 January 2014. http://blog.52north.org/2014/01/30/52north-supports-the-jrc-in-developing-an-inspire-download-service-based-on-sos/.

[6] Nativi S, Craglia M, Pearlman J. 2013. *Earth Science Infrastructures Interoperability: the Brokering Approach*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6 (3); p. 1118-1129. JRC75251.

[7] European Space Agency's official web page on *Copernicus Sentinal-1*. http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-1.

[8] Official web page of the Foresighting Europe Newsletter . http://ec.europa.eu/research/foresight/index_en.htm.

[9] G8. 2013. *G8 Open Data Charter and Technical Annex*. Cabinet Office UK, https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex.

[10] Official web page of the *Open Government Partnership*. http://www.opengovpartnership.org/.

[11] European Commission. 2010. *Digital Agenda for Europe*. COM/2010/0245 f/2.

[12] Official web page of the *Research Data Alliance (RDA)*. https://rd-alliance.org/.

[13] Jaeger, C., Jansson, P., van der Leeuw, S., Resch, M., Tàbara, D. and Dum, R. 2013. *Global Systems Science Orientation Paper*. GSS Web Portal.

[14] The official web page of the *European E-infrastructure Forum (EFF)*. http://www.einfrastructure-forum.eu/.

[15] The European Commission's official web page for *Digital Science*. http://ec.europa.eu/digital-agenda/en/science-and-technology/digital-science.

[16] Official web page of the *European Citizen Science Association (ECSA)*. http://ecsa.biodiv.naturkundemuseum-berlin.de/.

[17] Web page of *Zooniverse*. https://www.zooniverse .org/.

[18] Web page of *Crowdcrafting*. http://crowdcrafting .org/.

[19] Official web page of the *First Citizens' Observatories Projects Coordination Workshop*. http://ec.europa.eu/research/environment/index_en.cfm?pg=gepw-meeting-5&section=geo.

[20] Official web page of the *European Observation Network for Territorial Development and Cohesion ESBON)*. http://www.espon.eu/.

[21] Official web page of the *European Strategy Forum on Research Infrastructures (ESFRI)*. http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri.

[22] Official web page of *Earth Cube*. http://earthcube.org/.

[23] Official web page of *Australian Urban Research Infrastructure Network (AURIN)*. https://aurin.org.au/.

[24] Official web page of *a Reusable INSPIRE Reference Platform (ARE3NA)*. https://joinup.ec.europa.eu/community/are3na/home.

[25] Official web page of the *Mozilla Science Lab*. https://wiki.mozilla.org/ScienceLab.

[26] Official web page of the *Geo4All initiative*. http://www.geoforall.org/.

[27] Amateur Entomologists' Society. 2014. *BioBlitz - Definition of BioBlitz n the Entomologists' glossary*. Amateur Entomologists' Society (AES).

[28] Official web page of the European Network of Living Labs (ENoLL). http://www.openlivinglabs.eu/.

[29] Owens, J., Femiano, B and Lentz, L. 2013. *Hadoop Real World Solutions Cookbook*. Packt Publishing.

[30] Chodorow, K. and Dirolf, M. 2013. *MongoDB: The Definitive Guide, 2nd Edition*. O'Reilly Media.

[31] Romano, A. 2013. CitizenScience – "Twitter data for Italy and Kowalski". COST ENERGIC workshop.