

Case Study

Case Study: Brazilian Virtual Herbarium

Cameron Neylon ‡

‡ Curtin University, Perth, Australia

Corresponding author: Cameron Neylon (cn@cameronneylon.net)

Reviewable v1

Received: 17 Oct 2017 | Published: 19 Oct 2017

Citation: Neylon C (2017) Case Study: Brazilian Virtual Herbarium. Research Ideas and Outcomes 3: e21701.
<https://doi.org/10.3897/rio.3.e21701>

Abstract

The Brazilian Virtual Herbarium (BVH) is a project of the Brazilian Centro de Referência em Informação Ambiental (CRIA) that has been running since 2009. The Virtual Herbarium provides an infrastructure that gathers digital records of plant specimens from primary source, mainly in Brazil, and makes them available through a central web portal. The source herbaria have complete control over what data is made through the portal and the data collected by BVH is made fully available.

BVH in common with many data infrastructures, faces challenges in retaining funding. Most funding sources are project based and as has been noted elsewhere this creates problems for the sustaining of infrastructures. BVH therefore has an interest in demonstrating the use of the data resources it hosts. Through the OCSDNet project it has strengthened its capacity in this area to develop tools showing its wide usage.

Overall the BVH hosts over eight million records (as of October 2017) and received 70 billion data requests in October 2017. Its users are mainly in Brazil but there is also substantial global usage. The primary uses are for research and education. There are a broad range of educational users, including universities but also schools. Through providing a central aggregation and access point BHV provides a data infrastructure that is greater – and more useful – than the sum of its parts.

Keywords

research data, data management, data sharing, botany, Brazil, infrastructure, data infrastructure

Main Findings

The main distinguishing characteristic of the BVH (Maia et al. 2017) from the other case studies is that it is an infrastructure and not a project. Data is collected elsewhere and made available through the BVH system. Over its lifetime BVH has developed a policy of encouraging but not requiring openness of data. The source herbaria are regarded as owners of the data.

The fact that infrastructures are not well served by project-focussed funding models is well established (Canhos et al. 2015). This leads to organizational failures as infrastructures are pushed to develop new features rather than to strengthen resilience and other infrastructural characteristics. What is less obvious is that the tools and approaches for research data management are not well suited to data infrastructures in subtle and unexpected ways. While the broad thrust of questions in most data management plans are useful in addressing data planning for infrastructures, the structure and supporting documentation is built around an assumption of discrete projects.

- Data infrastructures like the BVH can successfully leverage dispersed data resources to create additional value through aggregation and services. The success of BVH emphasizes how discoverability and services are critical in maximizing the value of data sharing.
- The case study reinforces the findings and commentary from others that project-based funding models do not support infrastructures in focusing on their key infrastructural characteristics such as technical resilience, lowering costs and enhancing discoverability.
- Research Data Management support tools and guidance are similar built around an assumption of discrete projects and are unhelpful for infrastructures in subtle ways that are not immediately obvious. The focus of RDM policy on using grant-making as a point of leverage likely exacerbates this issue.

Awareness and pre-existing capacity for managing and examining data

The BVH project was recruited from the [Open and Collaborative Science in Development Network](#) (OCSDNet, Chan et al. 2015). This project was therefore already engaged with Open Science issues, in this case from the perspective of better understanding the downstream use of the service and platform.

The project has extensive experience of the technical aspects of data management and technical platform provision. The structure of the system means that 'data' is seen primarily as the materials flowing from the upstream herbaria with less focus on the objects generated by the platform itself. In the data audit reference is made to the products or processing and visualisation for the web platform, but for instance, usage data is not mentioned at this stage. A strong conception of data and an existing management framework meant in this case that objects outside that scope were not obvious concerns.

The development of data management plans

The development of the data management plan (Canhos 2017) and the planning process was a challenge. The existing generic templates are project driven and not well suited for ongoing infrastructure projects. The distinction between projects *about* the BVH that generate new data, development projects that create new capabilities, and therefore generate new internal data, and the data being sourced from upstream herbaria complicates this.

The BVH team provided extensive comments in their response to the Pilot Project interim report which discussed these issues at some length. Large scale data management/production projects which receive substantial funding generally develop a bespoke management plan for managing data at scale. Small scale research projects are adequately served by generic templates in many cases. However platforms that sit in the middle, particularly those that are infrastructures that survive based on project funding are not well served by the existing templates.

Nonetheless the team was very supportive of the concept of DMPs and did find the process of some value. As noted in the response to the interim report (see the data package under Interim Report, Neylon 2017):

All this said, a data management plan (DMP) at the project level continues to be essential. If the data is to be indexed by an existing e-infrastructure or deposited in an institutional repository it probably must use accepted standards and protocols. A DMP is also necessary to ascertain that project data needs and outputs are attended.

Tools and systems: Experience of use in developing world context

The BVH team used the Portage DMPAssistant tool successfully and did not report any substantial technical problems. Brazilian network access is reasonably robust and an online service is appropriate. The team works in English so language was not a specific barrier, although questions were raised about the meaning of the questions in common with other contributing projects.

CRIA has a substantial IT infrastructure provided through the Brazilian National Research and Education Network, which is dedicated to providing web-based and data management services. Technical provision is therefore not limited, although the funding stability for CRIA services is a concern for the longer term.

Challenges of implementation and data sharing

The key challenge for data sharing in the context of BVH is the mode of control built up to enable access. The success of BVH is largely built on the control that the source herbaria have over the use of “their” data. This emphasis on control and ownership limits the ability of BVH to directly enact change. Nonetheless BVH is an extraordinarily successful example of enhancing data sharing within a specific context.

A specific challenge in the context of BVH is the provision of geographical data on endangered species. Again, views differ amongst the data providers as to what is appropriate. Again, quoting from the response to the interim report (Neylon 2017):

One of the studies I am carrying out in our OCSDNet project is in finding out what data is being blocked and why. Reasons vary, such as not publicizing geographic coordinates of species in red lists or of species of commercial value, or blocking data that has not been published. At the same time we have data providers that want to publicize geographic coordinates of endangered species so that there can be social control at those sites. There is no consensus, but there is freedom in following one's own convictions. We even have a case of a curator who did not know the data were blocked. Some curator in the past blocked the data for whatever reason and no one unblocked it.

This illustrates the strengths and weaknesses of a federated approach. Giving full agency to data providers allows them to develop their own comfort level with sharing, and in the experience of BVH, provides a framework in which they gradually move towards greater sharing. At the same time differences in practice, particularly when it comes to the response to issues of ethical concern such as endangered species can lead to inconsistent practice which may be harmful in the long term.

Changing culture and the role of policy

The BVH was built out of culture of and interest in data sharing and availability. The team embodies a culture focussed on ensuring the use of a diverse and valuable data sources by a range of user communities. They are engaged in a long term effort to promote a cultural change within the upstream herbaria driven by evidence of the increased usage that comes from a shared data access platform (Maia et al. 2017).

The challenges of funding infrastructure through piecemeal projects means that policy imposition by individual funders at the project level can easily be counterproductive. Unless policy across all relevant funders is highly consistent the problems of reporting for differing policies will create substantial administrative overheads (Canhos et al. 2015). The combination of funding from multiple sources and the way in which project-based funding already distorts the infrastructural nature of the project means that policy design needs to take careful consideration of the effects that it has on platforms like BVH.

Grant title

Exploring the opportunities and challenges of implementing open research strategies within development institutions (Neylon and Chan 2016).

References

- Canhos DAL (2017) Data Management Plan: Brazil's Virtual Herbarium. Research Ideas and Outcomes 3: e14675. <https://doi.org/10.3897/rio.3.e14675>
- Canhos DL, Sousa-Baena M, Souza Sd, Maia L, Stehmann J, Canhos V, Giovanni RD, Bonacelli MM, Los W, Peterson AT (2015) The Importance of Biodiversity E-infrastructures for Megadiverse Countries. PLOS Biology 13 (7): e1002204. <https://doi.org/10.1371/journal.pbio.1002204>
- Chan L, Okune A, Sambuli N (2015) What is open and collaborative science and what roles could it play in development? In: Albagli S, Maciel ML, Abdo AH (Eds) Open Science, open issues. IBICT, Rio de Janeiro. URL: <http://livroaberto.ibict.br/handle/1/1061> [ISBN 978-85-7013-111-9].
- Maia LC, Vieira AOS, Peixoto AL, Canhos DAL, Stehmann JR, Barbosa MRdV, Menezes M (2017) Construindo redes para promover o conhecimento da biodiversidade brasileira: a experiência do INCT - Herbário Virtual. [Building networks to promote knowledge of Brazil's: the experience of the INCT - Virtual Herbarium]. 1. Instituto Nacional de Ciência e Tecnologia (INCT) - Herbário Virtual da Flora e dos Fungos, Campinas, 111 pp. [In Portuguese, English]. URL: http://inct.florabrasil.net/wp-content/uploads/2017/08/inct_livro_10_2_ed-online.pdf [ISBN 978-85-415-0901-5]
- Neylon C (2017) Dataset For Idr Project: Exploring The Opportunities And Challenges Of Implementing Open Research Strategies Within Development Institutions. International Development Research Center. Zenodo <https://doi.org/10.5281/ZENODO.844393>
- Neylon C, Chan L (2016) Exploring the opportunities and challenges of implementing open research strategies within development institutions. Research Ideas and Outcomes 2: e8880. <https://doi.org/10.3897/rio.2.e8880>