JRS Biodiversity Foundation

SHOW ME THE DATA:
An Evaluation of Data Access in the JRS Grant Portfolio

Summary

IN ORDER TO BE USEFUL for biodiversity conservation and sustainable development, biodiversity information must be not only available, but also accessible. The JRS Biodiversity Foundation contracted an independent study to assess the level to which data produced by JRS-funded projects were discoverable online, and the ease with which those data could be viewed or downloaded. Only about half of the expected data products were viewable or discoverable online, and in many cases data were not clearly connected with project results. Data accessibility was not dependent on the country in which the project was located, and hasn't changed over time. Interviews with grantees helped to identify challenges to and enabling conditions for creating sharable data products. JRS is actively responding to the findings of this study through new planning tools to support future grantees and a data sharing policy that explicitly supports open access to data and data publication to well-recognized and secure repositories. ∎

> *Promote data access in every way possible through multiple platforms with support of publications, web resources, and social media.*

Background

JRS Biodiversity Foundation's mission is to enhance the understanding of biological diversity for the benefit and sustainability of life on Earth. The foundation recognizes that sharing biodiversity information is vital to the success of its mission. JRS is committed to continually improving our practices and to learning from our experiences and the experiences of our grantees. This report shares the main findings and recommendations that resulted from a study on Data Access performed on 49 projects financed by JRS during the past seven years (approximately $11,000,000 in total investment). The independent study was carried out by Jante LLC, from January to May 2014. The purpose of this study was to assess the access to data generated by JRS grants to understand the obstacles and enabling conditions leading to successful data digitization, enhancement, and access (taken together to be the data sharing process) and to make recommendations.

ABOUT JRS

The mission of the JRS Biodiversity Foundation is to enhance knowledge and promote the understanding of biological diversity for the benefit and sustainability of life on earth. Founded in 2004, the JRS Biodiversity Foundation focuses upon supporting biodiversity data and knowledge tools that are used to preserve and to sustainably manage biodiversity in developing economies where biodiversity is most threatened. The foundation has awarded $13.5M in grants since 2007. For more information, please visit **jrsbiodiversity.org.**

Approach to the Study

The study was carried out in three search phases, each characterized by the type of information available to the searchers. The first attempt to find project data outputs was made by only using online resources and basic project information such as title, project director and organization. Second, proposals and progress reports to JRS were incorporated to the body of knowledge. Finally, consultations were done with each grantee to fill in the picture of discoverable and available data from their projects. During the study, all activities within projects were assigned to one of nine categories:

1. Technology Development;

2. Engagement: Training;

3. Research: Collection & Identification;

4. Research: Molecular Analysis;

5. Data Sharing: Digitization (including digitization and curation of data);

6. Data Sharing: Access (including development of data portals and/or their enhancement);

7. Data Sharing: Data Enhancement (geo-referencing, data-cleaning, etc.);

8. Downstream Uses (including decision tools, red lists, etc.);

9. Engagement: Social Involvement.

Assessing data access within this categorization scheme helped us to find relationships and dependencies between different kinds of activities and allowed us to identify the most vulnerable steps, those in which most problems arise, and why. The following is a summary of our findings.

Findings

IN GENERAL

• Data access success is independent of the JRS granting year and of geographic origin of the granted institution, showing there are challenges that cannot be easily overcome even with the experience and resources within some regions.

• It can be very difficult to determine whether a given product is a result of activities of a grant without "inside" knowledge, that is, without knowing exactly the aims of the proposal.

DATA ACCESSIBILITY BY ACTIVITY TYPE
(Percent of Projects)

Legend:
- Projects per Category
- Evidence Discovered
- Data Online
- Data Downloadable
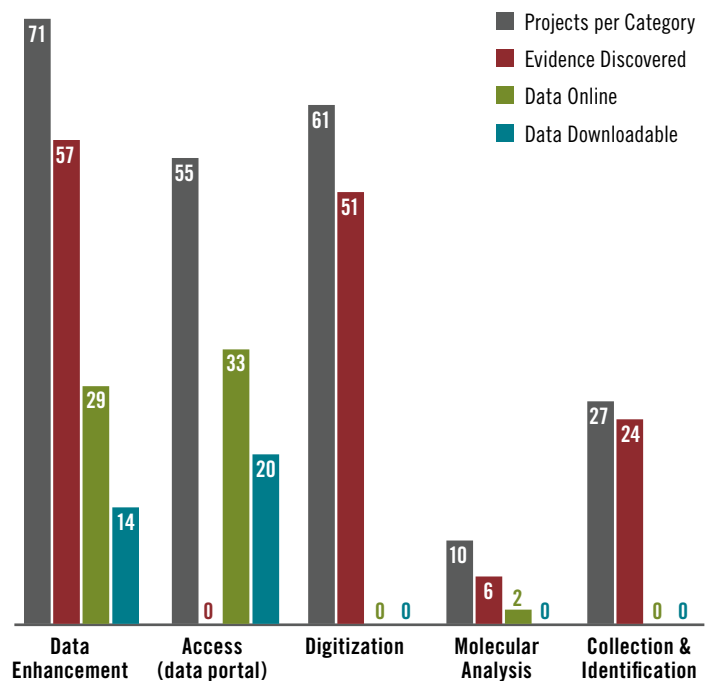
Results of the study of 49 grants in the JRS grant portfolio: Activities in projects were assigned to one of five categories that are part of the biodiversity data sharing process (horizontal axis). For each data sharing activity category, the vertical columns show the percent of projects per category, the percentage of total projects for which evidence can be discovered, and the percentage of total projects with online data or downloadable data. We find that evidence of the activity can be discovered in most cases, but a small fraction of activities produced data that is online and a smaller fraction of projects yielded downloadable data.

Thus, one may not find data by searching for a project and discovered data may not lead to related project results.

• The types of information most difficult to discover as resulting from project activities were publications, presentations, and information on workshops and other kinds of training activities.

• Most projects (90%) in the JRS portfolio focus on major aspects of the data sharing process workflow, that is, on digitization, enhancement and access to data. This shows an enormous effort being made towards achieving a broader reach and availability of the biodiversity information.

- Among the most common activities on grants are data enhancement (74%) and data digitization (61%). There are many common challenges for these categories, especially the significant underestimation of effort required.

- Training is the most common component of grants (80%), but tends to be on specific project-related topics rather than on general biodiversity informatics.

ON CREATING ACCESS

- Data are accessible and downloadable less than JRS had hoped and less than was implied by grant proposals. Even though much work has been done towards data access, data are not available through the Internet. In some cases this is due to timing issues (grants in process), while in other projects that are already complete this is due to lack of sustaining resources and motivation for data sharing.

- Challenges were diverse and could be characterized as internal or external. General challenges such as weak planning at the project level negatively impact the data sharing process as well.

- Within a project, the success of later steps in the data sharing process is dependent on the success of the earlier steps (namely, data access is dependent upon enhancement, which is dependent upon on digitization).

- Projects frequently encountered delays and challenges in the later steps of technical processes, and stages in data access processes generally had not been tested at smaller scale.

- Inadequate technical infrastructure and technical expertise constitute the major challenges in accomplishing objectives.

Recommendations

Based upon our findings and analysis, we offer a set of recommendations to help in the planning and development of biodiversity data-sharing projects. Recommendations are framed and driven by the ideas of free and open publication, which we understand might not be available and/or appropriate for every step of a project or for every taxon and geographical region involved. Hence, we encourage project leaders and their collaborators to decide carefully upon the best use of these recommendations within a project's goals.

IN GENERAL

- Be specific, succinct, and transparent about all aspects of the project from proposal through final report, particularly with respect to key partners in the data sharing process.

- Work to engage partner institutions in formal agreements for making data freely available before and during the project.

- Establish collaborations formally before submitting a project proposal. Binding written agreements are suggested when possible.

CALCULATING EFFORT BEFORE AND THROUGHOUT

- Develop and document reliable metrics of digitization and data enhancement rates and costs in collaboration with someone who has previous experience.

- Include risk assessment of potential problems along the entire process of data generation and access creation, and use the assessment to develop risk mitigation plans.

- Develop and test data sharing workflows (any combination of steps in the data sharing process) end-to-end rather than trying to do the steps completely separately and sequentially. Wherever possible, make sure that the entire workflow is functional and produces results at reasonable rates on a smaller scale before entering production mode. Avoid relying on one step to be completely finished before beginning the next step, as errors in an early step could require the process to be at least partially redone.

ABOUT DATA

- Make primary data available as broadly as possible under a Creative Commons (CCO) Universal Public Domain Designation by publishing it in highly visible venues such as the Global Biodiversity Information Facility in the case of species occurrence data.

- Promote data access in every way possible through multiple publication platforms with support of publications, web resources, and social media.

- Publish data records even if some part of them (such as location) must be withheld to protect sensitive species. Use metadata to say what other information may be available for bona fide research.

Collaborative planning, training regarding data standards, and setting open data access norms within projects appear to foster the eventual access to and use of the resulting biodiversity data. (Photo courtesy of Gavin Withers Photography, Cape Town).

MAKE IT WORK AND KEEP IT GOING

- Use the right tools and software in workflows. Ask for advice from experts before committing to using particular tools. Assess how much effort it would take to train personnel in the use of these technologies and consider the cost of their long-term maintenance.

- If network infrastructure is unreliable, collaborate with a hosting institution or include a budget for cloud computing and hosting. Cloud-based solutions are advised unless data infrastructure is proven to be reliable and available at reasonable cost.

- Be aware of sustainability of the projects you propose, choosing technologies appropriate to the expected project landscape and lifetime. When possible, choose platforms and software that do not require continuous financial support and that are maintained and updated by the biodiversity and information technology communities.

How JRS is Responding to Our Findings

The study of data access in our portfolio was a very important moment of learning for the foundation and we recognize the opportunity to better support our grantees and to pursue our mission. The results of this data access study in our grant portfolio were sobering for JRS and a signal call that we must improve. Only about half of the expected data products were discoverable on-line and accessible on-line. That means that information and research that could be benefiting science, conservation, and society is out of reach of potential users.

In response, JRS has developed a new data access policy which states that projects funded by JRS will adhere to these principles:

1. Provide unrestricted access to and use of all project products;

2. Provide timely access during the grant period or, maximally, within one year of its termination;

3. Employ, as a default, the Creative Commons Zero (CC0) license that permits copy and reuse;

4. Employ well-documented software, community standards, and repositories; and

5. Acknowledge data sources and abide by their terms.

Our grant applications, grantee selection process, grant agreements and progress reports will all align with these five principles. We will work with grantees to ensure that plans are in place for data sharing and for the evaluation of data sharing. We are developing tools to help with workflow planning and risk assessment. The JRS Biodiversity Foundation is proud join a growing number of international organizations and donors in taking a strong position in support of open access. Open data access will foster data use for research, policy development and implementation, and biodiversity conservation. ■

JRS Biodiversity Foundation

936 N 34th Street, Suite 400
Seattle, WA 98103
+1 (206) 454-7915

SHOW ME THE DATA: An Evaluation of Data Access in the JRS Grant Portfolio

JRS Biodiversity Foundation

jrsbiodiversity.org