

Full Proposal Submission

Section 1: General Project Information

Project Title: The impact of Brazil's Virtual Herbarium in e-Science

Duration of Project: 24 months

Countries included in this project: Brazil

Regions included in this project: Latin America

Research Themes: Motivation and outcomes due to open and collaborative science through e-infrastructures

Justification of Research Themes: The increase of knowledge on Brazilian biodiversity, associated with scientific advances to understand the evolutionary processes that generate and maintain this diversity, are fundamental to the sustainable use of this natural capital. Samples and associated information on plants and fungi collected in Brazil in the last three centuries are stored in national and international herbaria. The INCT – Virtual Herbarium of Flora and Fungi (<http://inct.florabrasil.net>) was established in 2009 to document, store, disseminate, and increase the knowledge base on the diversity of plants and fungi of Brazil. The Virtual Herbarium currently integrates over 100 collections, providing free and open online access to 5 million specimen records and 900 thousand associated images. At its first phase (2009-2014), 100 Brazilian herbaria joined the network, catalyzing improvements in data curation and quality management. This initiative also provided conditions to further develop tools for data analysis and visualization, and indicators to assess the progress of network development. The *Lacunas* systems allows the evaluation of data quality and completeness, and the identification of taxonomic and geographic knowledge gaps. BioGeo enables spontaneous voluntary work in producing and publishing online species distribution models, free to all. The success of the INCT Virtual Herbarium is the baseline to the ambitious goal to consolidate this e-infrastructure as an e-science platform to boost frontier developments in taxonomy, ecology, biogeography, and biodiversity informatics.

Total Budget Cost (CAD): 80,000

Section 3: Proposed Study Information

Research Project Abstract

WORD LIMIT: 250.

For the last 13 years, CRIA has worked on the development of the *speciesLink* network, an e-infrastructure that provides the means for biological collections to share their data within a structured system responsible for data storage, search, retrieval, and visualization. During the last 6 years, its focus has been on the INCT - Virtual Herbarium of Plants and Fungi, one of Brazil's National Institutes of Science and Technology that became a thematic network within the *speciesLink* network. Great focus was given on increasing online data availability by digitizing and integrating new herbaria to the network, improving data quality, developing new tools and providing training. The results are excellent. Besides making 5 million data records and 900,000 images openly available, new visualization tools have been developed to produce maps and graphs on-the-fly and to compare images, thereby enabling cyber taxonomy. However, not much attention was given on analyzing the usage aspects, to evaluate the effect of the e-infrastructure in graduate programs or to identify possible outcomes due to the participation of the herbaria in the e-infrastructure. This project is an opportunity to study these aspects of the network and to evaluate whether this e-infrastructure is promoting collaboration and is being used to solve local problems.

Research Problem, Significant and Justification

WORD LIMIT: 1,000. Please provide a brief overview of relevant literature and highlight the knowledge gaps that this project will address. Indicate the size and scope of the problem, as well as how the problem relates to the purpose and goals of OCSDNet; broader national development priorities, and the research and capacity needs of the countries involved.

Large investments have been made in developing cyber infrastructures to support research (Barjak et al., 2013). Examples from Brazil include Brazil's National Education and Research Network (RNP) and the National Centers for high performance processing (Cenapad). However, engineering breakthroughs alone will not be enough to achieve the outcomes envisaged for the undertaking of e-Science and other global collaborative activities supported by the 'cyber infrastructure'. If it is to be achieved, it will more likely be the result of a nexus of interrelated social, legal and technical transformations (David, 2005; Tenopir et al., 2011).

Regarding the *speciesLink* network, about 70% of the data is from herbaria, 25% of the remaining records are of animals, 0.2% of microorganisms, 0.1% of fossils and 4.7% of broad scope collections, such as OBIS-Br (Ocean Biogeographic Information System – Brazil). Besides integrating national collections, *speciesLink* also repatriates data of samples collected in Brazil that are held in collections around the world. Currently data from 12 biological collections from abroad are available in *speciesLink*, including 8 herbaria.

Together, they constitute about 10% of the total online records. *speciesLink* as a whole had over 200,000 visits in 2013, and these visits resulted in visualization or downloading of approximately 1.4 million records per day, the majority of which is data from herbaria. By *visualization*, we mean viewing of individual records, plotting records on maps or graphs, and viewing images.

Having an e-infrastructure focused on making data and tools openly and freely available online, Brazil's Virtual Herbarium of Plants and Fungi provides an ideal case study to help understand the drivers for collaboration and participation in this field.. The network includes herbaria that have been sharing data since 2003 and those that have joined the network in 2014, besides those that are not associated to the INCT - Virtual Herbarium, but participate in the *speciesLink* network. In addition, of the 25 graduate courses related to botany and taxonomy in Brazilian federal and state universities, 24 have herbaria associated to the INCT – Virtual Herbarium.

Currently, 94.6% of flowering plants have at least one record online. As *speciesLink* has reached critical mass for plant data, tools such as Lacunas and Biogeo were designed and implemented to extract and analyze information from raw primary data. Lacunas was developed to help identify data or knowledge gaps of plant species' distribution in Brazil and to help evaluate the conservation status of endangered species. It facilitates the identification of under-sampled species or geographic areas that have been neglected by specialists, or the need to support digitization and georeferencing of data from important and representative collections (Canhos et al., 2013). Biogeo is a public web interface that facilitates the generation of distribution models of Brazilian plants and fungi species. It aims at helping to increase and refine the biogeographical knowledge of Brazilian flora, elucidating species' environmental needs, as well as indicating species that lack occurrence points, information that can be used by the specialist to guide new fieldwork. Biogeo is based on a pre-defined workflow in which specialists can actively select the records to be used, and then evaluate the quality of the model. The model generation is an automatic process, based on a series of parameters according to De Giovanni et al. (2012).

Although all these interconnected cyber infrastructures are in place and being used by a broad range of actors and access is measured, no proper attention has been given to qualify this usage, to evaluate the effect of the e-infrastructure in the graduate programs or to identify possible outcomes due to the participation of the herbaria in the e-infrastructure. It is important to know who is using the e-infrastructure and for what purpose. Through this project we intend to analyze data providers to identify possible outcomes from data sharing and who are the data users of the system. We hope to identify how diverse the user community is – diversity of fields, institutions, countries, purposes, etc. The impact (negative and positive) and possible outcomes through online data sharing will be identified, as well as the reasons for not participating and for limiting access to part of the data. Furthermore, as 24 of the 25 graduate programs in botany have herbaria that are openly sharing data in the e-infrastructure, we would like to investigate whether this has promoted any change in the programs and research activities. We also intend to identify policy, legal, institutional, and technological impediments or drivers for collaboration and participation.

Towards the end of the project, the aim will be to understand the drivers of participation and collaboration to build the virtual herbarium. The main uses of the e-infrastructure and its users and verify whether the data, information and knowledge are being used to solve local

problems.

Outputs of this project shall contribute to the following OCSDNet thematic research areas:

- Theme 1 (T1) - Motivations: the project shall identify drivers that motivate herbaria as data providers to share their data through the e-infrastructure;
- Theme 2 (T2) - Infrastructure and technologies: The e-infrastructure and online tools available, local connectivity and IT support will be evaluated to indicate possible barriers to full participation;
- Theme 3 (T3) - Communities of Practice in Open and Collaborative Science: herbaria will be evaluated as to institutional policy and legal impediments concerning open data sharing; and
- Theme 4 (T4) - Potential impacts (positive and negative) of open and collaborative science: outcomes from participation to data providers and the diversity of uses and users of the e-infrastructure shall indicate potential or real impacts of open and collaborative science.

Research Questions and Objectives

WORD LIMIT: 500. Outline your project's central research question(s), sub-questions, and objectives. There must be congruency between the questions, objectives, research design and methods. You should highlight how the study's questions and objectives will contribute to the research themes of the OCSDNet.

The project's objective is to study both, the drivers for collaboration and the impact of the Brazilian Virtual Herbarium (BVH) in e-Science.

Currently BVH integrates close to 5 million textual data records and about 900,000 images from 100 herbaria across the country and 8 from abroad, making it freely and openly available online for anyone interested.

In the last 13 years, with the development of the speciesLink network and the last 6 years with the development of Brazil's Virtual Herbarium, focus has been on convincing data providers of the benefit of sharing data and thus increasing their visibility and importance. Some questions related to possible outcomes to data providers are:

- Has data sharing through BVH lead to more recognition and support for data providers?
- Are official evaluating mechanisms considering data sharing as an important element and do they result in incentives to collect, organize, qualify and share data?
- Is data sharing being affected by the way scientific production is evaluated and has this lead to inter-institutional, multi-discipline projects?

The BVH e-infrastructure was developed with a focus on researchers (taxonomists, ecologists) and policy makers as its main data users. Questions to be addressed in this project are:

- Who are the data users?

- For what purpose do they use the data and tools?

Little attention has been given to an ever-increasing exchange of data and expertise motivated by open sharing of data and developments in information and communication technology. We would like to know how openness is practiced and will analyze why some data fields are blocked. Examples include geographic coordinates. Tools to block data fields were developed to allow institutions to control the release of data according to their own data policy and criteria for sensitive data. It is time to evaluate why these data fields are blocked.

Our hypothesis is that data providers are withholding data that have not been published and geographic coordinates of data referring to commercial and endangered species. Are these assumptions correct? Are there other reasons? The network also developed a feedback mechanism that allows users to comment specific data fields of specific records. We would like to evaluate whether users are helping to qualify data through the annotation and feedback mechanisms in place. Another system is BioGeo (<http://biogeo.inct.florabrasil.net>) that depends on voluntary crowd sourcing to produce and publish ecological niche models for Brazilian plants and fungi. We would like to study:

- What motivates crowd sourcing within BioGeo?
- Are there mechanisms that could be used to increase participation?

Another line of research focusses on the e-infrastructure developed to make data sharing relatively easy, to give due attribution, increase data usefulness and usability, to innovate by providing a number of tools and applications for both data providers and end users. Here technology plays an important role and we would like to evaluate how technology influences data sharing and collaboration.

These are some of the questions that we expect to answer through this project. Answers to these questions will contribute to all four thematic research areas of OCSDNet:

- Theme 1: Motivations. We propose to study existing evaluation and reward mechanisms to see if data sharing is being considered.
- Theme 2. Infrastructures and Technologies. IT is the basis of e-infrastructure and directly affects governance and expertise sharing.
- Theme 3. Community of practice in Open and Collaborative Science. Identifying who are the data users and what the e-infrastructure is used for may help understand how to better communicate and share data and knowledge.
- Theme 4. Potential impacts of open collaborative science. Again, a better knowledge on usage can help identify existing and potential impacts and recommend strategies to enhance open collaborative science.

Stakeholders

WORD LIMIT: 250. Identify and briefly describe your project's stakeholders. How will your project respond to their needs and interests?

Stakeholders include data providers (herbaria curators and taxonomists) and data users (curators, taxonomists, ecologists, graduate students, teachers, policy and decision makers, and others). The project aims at identifying outcomes for herbaria due to their participation in the network. By identifying outcomes and understanding how they occurred, it might be possible to enhance or replicate these to other data providers.

By identifying who the users are and what/how the e-infrastructure is used, one might get a better understanding of the dynamics or needs of specific tools to enhance collaboration. When developing the speciesLink network, a strong assumption is that local e-infrastructures are fundamental as they promote cooperation and collaboration of local communities, understand local specificities, address local problems, and increase the likelihood of appropriation and use of data, information and knowledge locally. The project will help understand who and how the e-infrastructure is used and if these assumptions are true.

Research Design & Methods

WORD LIMIT: 1,000. In this section, applicants should clearly indicate and justify the proposed study design. You should discuss how you intend to collect the data that you will need to achieve the study's objectives and answer the project's research questions. You should clearly outline how each data collection activity will contribute to the study objectives.

There are three target groups that will be studied within the scope of this project: data providers, individual contributors, and data users. Contributors include individuals that are contributing with their knowledge by using available annotation tools and are producing and publishing ecological niche models online (see biogeo.inct.florabrasil.net).

Data Providers

1. Classifying the herbaria (month 1-2)

Data providers of Brazil's Virtual Herbarium consist of a more homogeneous group as all are herbaria. But they vary in size, geographic location and nature of the institution (public, private; university, research institution). They also vary as to when they began sharing data through the e-infrastructure. All participating herbaria will be classified in these different categories, based on the metadata available online at <http://splink.cria.org.br/manager>.

2. Interviews and preparation of a questionnaire (months 2-4)

A sample of these herbaria will be interviewed to identify their motivation and difficulties to participate and outcomes that resulted from their participation, both positive and negative. Sample selection will take into consideration various factors like size, region, and type of institution, in order to make sure that a representative group is selected for the interviews. It

is noteworthy to mention that 10 herbaria that share their data through the *speciesLink* network do not participate in the National Institute of Science and Technology project. They will also be asked to participate. Based on these interviews, a questionnaire will be prepared and applied to all participating herbaria.

3. Application of the questionnaire (months 4-5)

In the questionnaire, herbaria will not be identified, nor will the institution or the person who is answering it. The questionnaire will have basic fields to help classify the herbaria in their specific groups (public or private, size, region, etc.) and to know the position and personal characteristics of the person responsible for the information, such as age and gender. An online form will be made available and will include all items identified in the interviews under *motivation, difficulties* and *positive and negative outcomes*.

An explanatory email will be sent to all curators indicating the objective of the project and requesting that they answer the online questionnaire. If necessary, a second email will be sent in the beginning of month 5.

In the questionnaire, the herbaria will be asked to indicate a numerical value for each element, based on the significance of that element for their herbaria. For example, if under the item *Positive Outcomes* there is an option *Institutional recognition*, the herbaria will indicate a value between 1 and 5 that means “*very weak*” recognition to “*very strong*”. A free textual field for item (motivation, difficulties, positive and negative outcomes) will also be available so that one can identify aspects that were not mentioned by those that were interviewed during step 2.

4. Tabulation of results and preparation of a report (Months 6-8)

The answers will be tabulated and by determining the frequency of each numerical value per item, one will be able to identify which items are more common or relevant and perhaps which are more relevant per group, such as size. All textual answers will be analyzed and evaluated to see whether some common characteristics can be identified.

A report with the results and analysis will be prepared.

5. Dissemination (months 9-12)

Once prepared, the report will be made available (during a period of one month) to all herbaria to receive comments, suggestions, and criticisms. It will then be presented at a side event during the 66th National Congress of Botany in October, 2015, in Santos.

An article will be submitted with the process and results of this work.

Contributors

There are basically two systems that allow and give visibility and credits to contributors: annotation (within the *speciesLink* network) and BioGeo. For both systems, CRIA has the necessary contact information.

1. Preparation and application of the questionnaire (months 4-7)

The objective of the questionnaire is to identify their motivation for collaboration and possible outcomes they may have had. The questionnaire will not identify the person who answers, but will identify his/her expertise, age, gender and type of institution.

In the case of the annotation system, all contributions are available in a database and will be evaluated to be used as the basis for the questionnaire. Emails will be sent out to all contributors explaining the objective and the use of the results.

2. Analysis and Report (month 8-9)

All answers will be analyzed and a report will be prepared.

3. Dissemination (9-12)

The results will be presented at the National Congress of Botany in 2015 and an article will be submitted.

Users

This target group is more diverse and, to some degree, unknown. The aim is to identify users, qualify usage and identify possible collaborative initiatives.

1. Analysis of usage of the years 2013 – 2014 (months 2-8)

The first step is analyze in detail the existing usage statistics. To analyze logs to its systems, CRIA uses AWStats (Advanced Web Statistics). Our statistics page presents the number of visitors, unique visitors, bandwidth and pages and top 10 countries and IPs. We will look more closely at IPs to identify the origin of users. If necessary, different reports will be produced to better visualize these statistics.

CRIA also has statistics for *speciesLink*'s search and retrieval interface that quantifies and identifies what tools and data are being accessed. This data will be analyzed to evaluate whether a usage profile can be drawn.

2. Prepare the project strategy for year 2 concerning Users (months 9-12)

The results of the analysis carried out will be studied and the strategy for a more detailed qualification of uses and users will be defined for project year 2.

3 Carry out the work defined in Year 1

4. Assessment of the role of the Virtual Herbarium for Graduate courses (Year 2)

A separate strategy will be directed at graduate courses. This will be discussed with the Virtual Herbarium's steering committee that have direct involvement with these courses.

Analysis & Synthesis

WORD LIMIT: 1,000. Describe how you intend to organize, examine and model data to arrive at conclusions and insights.

We believe that the answers and contributions made by data providers will have less "unknown" aspects to us, as this is a community we have been working with and for, for the last 12 years. For this reason we believe that it will be possible to submit a questionnaire where numerical values will be attributed to the answers in order to present the frequency of the answers and this way identify the most important or frequent drivers of motivation and both positive and negative outcomes. We will also consider the variety of institutional contexts to see what factors influence these drivers.

Based on this analysis, we believe that we will have important elements that may be included in public policies or strategies that will help ignite data sharing and collaboration. The same applies to those elements that represent barriers to data sharing and collaboration.

Usage is still very much unknown to us. We have general statistics on visits, bandwidth, and access to tools and data but this just quantifies usage on a numerical basis. It is important to qualify the use of the e-infrastructure, identifying who is using it (institution and expertise)

and for what purpose (commercial, research, conservation, recuperation, etc.). Answers to the online questionnaire directed at users will be analyzed weekly to improve the questionnaire and present options using a common vocabulary. Periodically, answers will be tabulated and the frequency of each option will be determined. With this we will identify the users (institutions and expertise) and what the data and possibly tools are being used for. As for the graduate courses, we believe that this is an important user group that must be made aware of the opportunities and benefits of e-Science. We need to know whether the e-infrastructure is reaching these groups and identify what is being used and for what purpose. We also need to know if this is affecting research and is helping to promote collaboration – in house, within the Virtual Herbarium Network or even with international partners. This will be an analytical study and the result will be shared with all involved – herbaria, graduate teachers and students. As to individual contributors, BioGeo today has a little over 100 voluntary registered contributors and the annotation system has 446 contributions. A general email will be prepared and sent to each contributor who will be asked to answer an online questionnaire and this way guarantee anonymity. The aim of the questionnaire is to identify motivations and research interest besides qualifying the contributor by age, gender, type of institution and expertise.

Outcomes & Outputs

WORD LIMIT: 700. Describe the major project outputs and intended outcomes. Your project outputs should creatively reflect the principles of open and collaborative science.

Outputs shall include the identification of drivers and barriers to open data sharing that may be technical, institutional, legal, and may include evaluation and funding policies. Other outputs may include drivers and barriers for participation and collaboration. Envisaged outcomes include new evaluation and reward mechanisms to open data sharing, and improved policies and strategies for collaborative research.

Knowledge Translation & Dissemination

WORD LIMIT: 700. Describe how you will disseminate your outputs. To ensure that the results of your study are applied to address development challenges, explain how you intend to package, disseminate and promote the application of your findings amongst relevant stakeholder groups.

Preliminary and final conclusions will be made available to the e-infrastructure's data providers and users for comments, suggestions and criticism. This work will also be presented at the National Congress of Botany in Santos, SP, Brazil in October 2015. Articles will be submitted for publishing and a description of the work will be posted on CRIA's blog (blog.cria.org.br).

Network Connections & Interactions

WORD LIMIT: 500. Illustrate how you will contribute to the overall OCSDNet framework and themes. Draw on other initiatives and approaches discussed at the OCSDNet workshop, if applicable.

This project connects with the planned development of the second phase of the Brazilian Virtual Herbarium (May 2015- April 2020) where the following specific objectives are being proposed:

(1) Expand INCT-Virtual Herbarium's e-infrastructure, with the integration of 170 Brazilian herbaria (95% of the country's active herbaria) and 12 selected international herbaria, implementing mechanisms to support the continuous qualification of the holdings of the distributed herbaria, strengthening each one as consistent nodes of the network. The aim is to make 6.5 million records freely and openly available online (both national and repatriated data), 1.4 million with associated images.

(2) Stimulate research in taxonomy of neotropical plants and fungi, enabling the development of studies of taxonomic groups that are bottlenecks for the characterization of floras and mycotas, filling out the gaps of less known taxonomic groups and/or geographic areas.

(3) Enhance research in biodiversity informatics, developing tools and products to support data analysis and usage. New developments include a geographic interface for data searching, retrieval, and visualization; tools for the dynamic integration, organization and visualization of species information; tools to improve the metrics of usage and quality assessment of data, and an application for Smartphones.

(4) Capacity building and training in biodiversity sciences in the era of digital taxonomy, linking institutions and research groups in an articulate and cooperative manner through the development of improved mechanisms for communication, dissemination and knowledge sharing, and delivering practical and theoretical courses, at academic and technical levels.

(5) Develop e-science outreach activities, to support policy development, integrating the botanical community with other scientific communities, public administration and private initiatives through the e-infrastructure.

As indicated in the research problem, outputs of this project shall contribute to the following OCSDnet thematic research areas:

- Motivations: the project shall identify drivers that motivate herbaria as data providers to share their data through the e-infrastructure; and those that motivate individual contributions (annotations and BioGeo);
- Infrastructure and technologies: The e-infrastructure and online tools available, local connectivity and IT support will be evaluated to indicate possible barriers and enablers to full participation;
- Communities of Practice in Open and Collaborative Science: herbaria will be evaluated as to institutional policy and legal impediments; and
- Potential impacts (positive and negative) of open and collaborative science: outcomes from participation to data providers and the diversity of uses and users of the e-infrastructure shall indicate potential or real impacts of open and collaborative science.

Bibliography (APA style)

- Barjak, F., Eccles, K., Meyer, E. T., Robinson, S., & Schroeder, R. (2013). The Emerging Governance of E-Infrastructure. *Journal of Computer-Mediated Communication*, 18, 1-24.
- Canhos, D. A. L. Sistemas de informação em biodiversidade e a formulação de políticas públicas na era digital. PhD Thesis. Unicamp. 2013.
- Costello, M. J., Appeltans, W., Bailly, N., Berendsohn, W. G., de Jong, Y., Edwards, M., Froese, R., Huettmann, F., Los, W., Meesce, J., Segers, H. & Bisby, F. A. (2014). Strategies for the sustainability of online open-access biodiversity databases. *Biological Conservation*, 173, 155-165.
- David, P. A. The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: a Primer. Washington, p. 238. 2003. (ISBN: 0-309-52545-4).
<http://www.nap.edu/catalog/10785.html>.
- David, P. A. The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution. *Capitalism and Society*, 3, n. 2, 2008.<http://www.bepress.com/cas/vol3/iss2/art5>.
- David, P. A. (2004). Towards a cyberinfrastructure for enhanced scientific collaboration: Providing its 'soft' foundations may be the hardest part. In *Advancing knowledge and the knowledge economy* (chapter 24).
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. Sage.
- Lastres, H. M. M., Albagli, S. Introdução: Chaves para ao Terceiro Milênio na Era do Conhecimento. In: Lastres, H. M. M.; Albagli, S. Informação e globalização na era do conhecimento. Rio de Janeiro: Editora Campus Ltda, 1999. p. 7-26. ISBN 85-352-0489-X.
- Latham, R., Sassen, S. Digital Formations: Constructing an Object of Study. In: Sassen, R. L. A. S. Digital Formations: IT and new architectures in the global realm. Princeton: Princeton University Press, 2005. Cap. 1, p. 1-34. ISBN 0-691-11987-2.
- Leite, F. C. L., Costa, S. Repositórios institucionais como ferramentas de gestão do conhecimento científico no ambiente acadêmico. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 2, p. 206-219, mai./ago. 2006. Disponível em: <http://www.scielo.br/pdf/%0D/pci/v11n2/v11n2a05.pdf>
- Nowotny, H., Scott, P.; Gibbons, M. (2001) Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty. Cambridge: Polity Press. 278 p. ISBN 978-0-7456-2608-6.
- NSB. Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. National Science Foundation. Washington DC, p. 9-87. 2005.
- Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., Hokko, H., Jennions, M. D., Kruuk, L. E. B. (2014) Troubleshooting public data archiving: suggestions to increase participation. *PLoS biology*, 12, e1001779

Schroeder, R. (2007) E-Research Infrastructures and Open Science: Towards a New System of Knowledge Production. *Prometheus*, 25, 1-17.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011) Data sharing by scientists: practices and perceptions. *PloS one* 6, e21101.

The Royal Society Science Policy Center. Science as an open enterprise: open data for open science. The Royal Society. London, p. 1-104. 2012. (978-0-85403-962-3).

Wood, J., Anderson, T., Bachem, A., Best, C., Genova, F., Lopez, D., Los, W. et al. (2010) Riding the Wave - How Europe can gain from the rising tide of scientific data. European Union. p. 36.