Open Access to scientific knowledge: a methodological model for scientific information and knowledge management at the Brazilian Agricultural Research Corporation (Embrapa)

Fernando César Lima LEITE
University of Brasilia – Information Science Department; Embrapa Technological Information, Parque Estação Biológica - PqEB - Av. W3 Norte (final)
Brasilia, DF, CEP 70770-901, Brazil

Patrícia BERTIN
Embrapa Technological Information, Parque Estação Biológica - PqEB - Av. W3 Norte (final)
Brasilia, DF, CEP 70770-901, Brazil

and

Fernando PEREIRA
Embrapa Technological Information, Parque Estação Biológica - PqEB - Av. W3 Norte (final)
Brasilia, DF, CEP 70770-901, Brazil

ABSTRACT

This paper presents a methodological model for the establishment of Open Access to scientific information at Embrapa, as a strategy for scientific information and knowledge management. The model consists of elements that speed up scientific communication processes and allow for the research output management. The aim is to provide the necessary mechanisms to capture, store, organize, preserve and widely disseminate the scientific knowledge produced by Embrapa and by the scientific community involved in agricultural research, through the implementation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). It is our contention that effective information management improves institutional scientific communication, which contributes for the betterment of scientific research related processes.

Keywords: Open Access, Brazilian Agricultural Research Corporation, Embrapa, Scientific Communication and Knowledge Management.

1. INTRODUCTION

The objective of this work is to propose a methodological model for scientific information and knowledge management at the Brazilian Agricultural Research Corporation (Embrapa), considering the Open Access policies, statements and definitions for scientific communication. Scientific communication, i.e. writing scientific papers, reading and critiquing other papers, is an essential process for science development and ensures knowledge sharing and dissemination.

Shaughnessy defines science communication as a social phenomenon according to which intellectual and creative activity is transmitted among scientists [1]. Thus, scientific communication is responsible both for the sharing of registered scientific knowledge – the scientific information – and for the sharing of scientific knowledge related to the production, the experiences and the skills of the scientists, informally shared.

By their turn, Kaplan and Storer pointed out that scientific communication refers to the exchange of information and ideas among scientists in their roles as scientists [2]. According to the authors, there are scientific ideas that can not be disseminated as scientific information, i.e. there is much of knowledge that can not be expressed in scientific publications. Menzel (apud Kaplan and Storer, 1968, p.112) defines scientific communication as the summing of publications, facilities, occasions, institutional arrangements and costumes that affect directly or indirectly the transmission of scientific messages among scientists. The highlighted definitions of scientific communication emphasize concepts and aspects of interest for information and knowledge management such as, for example, sharing of creative and intellectual activity, exchange of information and ideas, publications, facilities, occasions. All these aspects are of interest for creating the methodological method of scientific information and knowledge management.

The functions of scientific communication

Scientific communication has specific functions, which, obviously, are transformed or modified through time, mainly by influence of the development of information and communication technologies. Kaplan and Storer [2] described the seven main functions of scientific communication, which are: i) to provide the scientists with the best answers to specific questions; ii) to contribute to the scientists awareness of new developments in their field of knowledge; iii) to stimulate scientists to search new knowledge beyond their field of interest; iv) to divulge the main tendencies of emerging areas, emphasizing the relevance of the scientists work; v) to attest the reliance of new knowledge through peers testimonies and verification; vi) to redirect or extend the scientists interests; and vii) to provide feedback for the improvement of the researchers production. Such functions of scientific communication could be considered the major objectives of scientific knowledge management.

According to Roosendaal and Geurts, the four most relevant functions of scientific knowledge communication would be: registration – the record of authorship, that ensures public acknowledgement and property rights related to the scientific finding; certification – which guarantees the quality control and provides legitimacy through peer-review processes; awareness – diffusing of research results to scientific community; and archive – the storage and preservation of scientific knowledge records [3]. These authors opened up a new line of thought,

relating the "scientist's desire of one unified collection of research achievements which then is distributed over many subcompartments, and results in a number of organizational conditions". That would be one of the first definitions of the digital repositories currently known to store and retrieve information, a question that strongly recovers scientific information and knowledge management.

Open Access to scientific knowledge

The last years of the 20th century have known a reaction to the restriction of the traditional system of scientific communication. The increasing specialization within all disciplines caused an exponential growth of information and brought about further expansion of new journal titles until the flood of new periodical literature begun to spur the notion of 'information overload'. Many investigators and other actors have criticized the delay between submission and publication of works, which somehow revealed the inability of the traditional system to attend the increasing capacity demand for recent scientific progresses dissemination, as a result of the global expansion of research and development.

By that time, the debate was still about the:

- obligation of depending on the editors for achieving diffusion of research results to the largest scale [4];

- requirement of subscription or 'pay-per-view';

- strictness of the peer review system;

- high manuscript rejection rates;

- tendency of system to focus on the quantity of output, favor sanctioned institutions and renamed authors, suppress new ideas and endorse traditional research methods [5,6]; and

- price increases of journals causing cancellations by libraries, which in turn are followed by new price increases [7].

The concern about developing new forms of formal and informal communication among scientists and the appearing of innovative approaches for accessing research and development results emerges from this scenery, together with the new information technologies' emergence, particularly Internet. It became clear that the traditional model of scientific communication limits, more than expands the availability and readership of most scientific research, at the time that it obscures institutional origins [8].

The scientific community dissatisfaction with the traditional model of scientific communication, together with the development of information and communication technologies resulted in a suitable environment for a significant transformation in the way information is shared inside scientific communities [9]. The new worldwide paradigm of *Open Access* to scientific knowledge appeared as a suitable alternative for scientific information management and has even been seen as a new communication philosophy [10].

The *Open Archives Initiative* (OAI), which has its roots in the Open Access and institutional repository movements, has performed an important and without precedent role, setting interoperability standards and facilitating the efficient dissemination of digital contents. The world scientific community has largely adopted such standards, initially defined with the institution of recommendations and mechanisms to assist digital archives cooperation and of value information systems. Among the general conditions required to establish interoperable digital archives, it can be detached: mechanisms of submission and self-archiving by authors – which means the deposit of authors' refereed journal articles in open digital archives; a long term system for storing and preserving information, with open interface that permits digital search of archives metadata; and management policies concerning submission of documents and preservation. In the perspective of *Open Archives* two main actors exist: the data and the service providers. Data providers manage and maintain digital repositories, which offer submission mechanisms, store and preserve documents, and expose its metadata. Institutional repositories (RepositóriUM – <http://repositorium.sdum.uminho.pt/>), subject repositories (E-LIS – <http://eprints.rclis.org/>), electronic scientific periodicals (such as the Brazilian Journal of Agricultural Research, edited by Embrapa – <http://seer.sct.embrapa.br/index.php/pab/index/>) that use the OAI-PMH protocol (Protocol for Metadata Harvesting from Open Archives Initiative) and metadata scheme compatible exemplify data providers. Service providers, on their turn, are responsible for collecting metadata exposed by several digital repositories and for creating information value systems (OASIS.Br – <http://oasisbr.ibict.br/>; OAISTER – <www.oaister.org>). That means to say that service providers collect metadata from data providers, store them in a databank and offer in a single interface, search mechanisms that provide access to all content of all digital repositories that had their metadata exposed and collected.

Besides of breaking out a new model of scientific communication, it is necessary to mention that Open Access consistently confers social, political, cultural and technological appropriate infrastructures to scientific information and knowledge management. Indeed, Open Access maximizes scientific results visibility and increases researches progresses, impact, productivity and rewards [11]. Regarding the impact of electronic journals, another author registered that online articles are cited 4.5 times more often than offline articles [12], when considering articles within each year, and averaging across all years from 1990 to 2000.

This work aims to present a methodological model for scientific knowledge management at the Brazilian Agricultural Research Corporation (Embrapa), to compare this model with the existing initiative of the Food and Agriculture Organization of the United Nations (FAO), and to report the current stage of the proposal's implementation at Embrapa.

2. THE METHODOLOGICAL MODEL

The methodological model was designed from the convergence and conceptual exploration of i) the functions attributed to scientific communication present in the models of Kaplan and Storer and of Roosendaal and Geurts [2,3]; and ii) the social, cultural and technological fundaments entrenched in the movement of Open Access to scientific knowledge. The theory embedded in the theoretical referential considers effective and efficient processes of scientific communication the central objectives of scientific information and knowledge management. Consequently, if sub-processes of scientific information and knowledge are appropriately administrated – i.e., taking into consideration the nature of information and its production, the information behavior of researchers and its communication patterns, and the institutional context of a research corporation – scientific communication will be substantially improved.

The impact of information and knowledge management practices based on Open Access over the scientific community can be understood in two perspectives: i) the internal environment should become more appropriate to learning, since knowledge will be explicit, socialized, organized and internalized; ii) the acquisition of external information and knowledge would result in a betterment of the internal processes related to research and development, what by its turn, favor innovation. In the external environment, scientific communication strategies applying the methodological model maximize impact of research results and of the institution itself.

The methodological model is constituted of the following elements (Figure 1):

- Data providers for internal scientific information: composed by scientific digital periodicals edited by the institution and the institutional repository (essential element whose functions are to store, organize, preserve, retrieve and largely disseminate the institution' intellectual production).

- Data providers for external scientific information: canalizes all scientific production concerning the institutional interest areas that are available in an Open Access environment and uses the OAI-PMH protocol.

- Institutional service provider: collects metadata that describe all the contents stored in the data providers. Providing access to the institutional intellectual production and to external information sources comprehend the objective of the institutional service provider.

- Institutional information policy for compulsory self-archiving, according to international recommendations.

3. DISCUSSION

To discuss the appropriateness of the proposed methodological model for scientific information and knowledge management at Embrapa, it is extremely relevant to present a little background about this research corporation and other existing Open Access initiatives, particularly those initiated by the Food and Agriculture Organization of the United Nations (FAO).

The Embrapa in few words

Embrapa's mission is to provide feasible solutions for the sustainable development of Brazilian agribusiness through knowledge and technology generation and transfer.

From the very beginning, on April 26, 1973, Embrapa has generated and recommended more than nine thousand technologies for Brazilian agriculture, reduced production costs and helped Brazil to increase the offer of food while, at the same time, conserving natural resources and the environment and diminishing external dependence on technologies, basic products and genetic materials.

Networking through 38 Research Centers, 3 Service Centers and 13 Central Divisions, Embrapa is present in almost all the states of the Union, each with its own ecological conditions. There are approximately 8,500 employees in Embrapa, of which 25% have master's degrees and 66% doctoral degrees. Embrapa coordinates the National Agricultural Research System, which includes most public and private entities involved in agricultural research in the country.

Embrapa maintains projects in International Cooperation in order to perfect knowledge of technical and scientific activities or to share knowledge and technology with other countries. Embrapa makes part of more than 70 bilateral agreements with almost 40 countries worldwide and 20 multilateral agreements with international organizations which involves principally research advances or technology transfers.

At Embrapa, the division that is responsible for the organization, production and availability of technical and scientific information produced is called the Embrapa Technological Information. The unit uses up-to-date information and communication technology in its activities and elaborates information products such as video-tapes, books, and softwares, among others, formatted in the language and appropriate vehicle targeted to its publics. It supports the transfer of technology by organizing and coordinating technical events and creating technical, scientific and socio-economic databases.

However, Embrapa has still a long way to procced so that scientific knowledge that is generated by the corporation will be organized and easily recoverable through specific information systems and through the web. It is still necessary to integrate the electronic publishing and the bibliographic cataloguing requirements into interoperable systems. Besides contributing to the management of Embrapa's research outputs, the proposed strategy will benefit research itself through the access to externally produced scientific knowledge. An initial survey revealed 261 data providers of interest to Embrapa, of which 230 are scientific journals (84 of them are from Brazil and 104 participates of the Scientific Electronic Library Online - SciELO) and 31 are digital repositories (conferences, subject and institutional repositories).

The FAO Open Archive Initiative

The FAO has more than 50 years of experience in the production and the dissemination of information related to nutrition, food, fisheries, forestry and agriculture [13].

The FAO created a Corporate Document Repository (CDR, http://www.fao.org/documents/) still in 1998, which houses FAO documents and publications, as well as selected non-FAO publications, in electronic format. Recognizing the value of indexing metadata and keywords to enhance efficient and quick location of information on the web, FAO created the Electronic Information Management System (EIMS), which maintains the electronic publishing workflow of FAO Corporate Document Repository. FAO departments use EIMS to insert records of publications, meeting documents and other materials such as photos for immediate publication on www.fao.org or to begin the process of converting written documents into HTML followed by publication. The EIMS-CDR holds more than 38.500 documents and is the gateway to FAO's publications [13].

The EIMS-CDR coexists with FAODOC, an online catalogue that indexes both electronic and printed documents while the EIMS-CDR manages full text documents and a minimal set of metadata. EIMS-CDR and FAODOC are currently in a process of merging to compose an unique FAO Open Archive based on the integration of the electronic publishing and the bibliographic cataloguing requirements. The challenge is mainly to solve the duplication of efforts in creating and managing metadata; and the lack of integration of electronic publishing and cataloguing.

The FAO Open Archive will guarantee efficient electronic publishing and metadata management, the effective dissemination of FAO information resources and the preservation of the Organization's institutional memory [13].

Open Access Impact To Brazilian Agricultural Research

Meadows emphasizes that Communication lies at the heart of research. It is as vital for the research as the actual investigation itself, for the research cannot properly claim that name until it has been scrutinized and accepted by colleagues. This necessarily requires that it be communicated. According to the author, whatever the angle by which it is examined, efficient and effective communication is an essential part of the scientific investigation process [14]. Differently from the health sector in Brazil, where information systems and scientific communication are relatively well organized and structured, thanks, mainly, to the efforts undertaken by the Latin American and Caribbean Center on Health Sciences Information (BIREME, http://www.bireme.br/php/index.php), information in agricultural research does not have a well developed infra-structure, which affects, directly, the production of knowledge. Brazilian agriculture, nonetheless, is considered internationally competitive.

In spite of the establishment of the National System of Agricultural Research (SNPA), in 1992, which is composed of Embrapa and other organizations direct or indirectly linked to agricultural research, there does not exist, in Brazil, an articulated information system in agricultural research that provides input for the activities of SNPA and of Embrapa itself. Considering that Embrapa is responsible for a substantial part of agricultural research in Brazil and that it is the coordinator of SNPA as well, the initiative of building a scientific information net founded in the philosophy of open access and interoperability standards will offer concrete alternatives for the speeding up of scientific communication processes and for information management in agriculture. The expected impacts of open access on Brazilian agricultural research are:

- Alternative information flows in agricultural research to subsidize scientific activities;
- Integration of local research to the global knowledge collection;
- To promote Brazilian agricultural research capability;
- Low access costs and scientific knowledge dissemination;
- Improvement of invisible colleges.

4. CONCLUSIONS

The methodological model for scientific information and knowledge management exposed, supported on Open Access rules, contributes to effective acceleration and improvement of internal and external scientific communication since it constitutes mechanisms to:

- Promote the scientific production, researchers and institution's visibilities, with great potential to enhance impact of results from research at Embrapa;

- Provide the scientific information management methodology, directed to the internal and external knowledge management (identification, capture, storing, organization, retrieval and largely dissemination);

- Associate and preserve, by using specific techniques, the scientific memory and the institutional intellectual production;

- Give unified access to all institution scientific production in digital format and integrally;

- Offer tangible indicators for evaluating institutional scientific production, to demonstrate public value as well as scientific, social and economical relevance of its activities;

Finally, the proposed model constitutes a target component for the elaboration of the scientific information policy at Embrapa, especially through the creation and institutionalization of a self-archiving policy.

5. REFERENCES

[1] T.W. Shaughnessy. Scholarly communication: The need for an agenda for action - A symposium. **The Journal of Academic Librarianship**, v.15, n.2, 1989, p.68.

[2] N. Kaplan; N.W. Storer. **Scientific communication**. In: SILLS, D. L. International encyclopedia of the social sciences. New York: Macmillan, v.14, pp.112-117, 1968.

[3] H.E. Roosendaal; P.A.T.M. Geurts. **Forces and functions in scientific communication: an analysis of their interplay**. First International Workshop of Cooperative Research Information Systems in Physics (CRISP97). Ondenburg, Germany. 1997. Available from: <http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>. Access: May 20th, 2008.

[4] S. Bachrach; R.S. Berry; M. Blume; T. von Foerster; A. Fowler; P. Ginsparg; S. Heller; N. Kestner; A. Odlyzko; A. Okerson; R. Wigington; A. Moffat. Intellectual property: who should own scientific papers? **Science**, v.281, n.5382, 1998, pp.1459-1460.

[5] S. Harnad. Learned inquiry and the net: the role of peer review, peer commentary and copyright. **Learned Publishing**, v.11, n.4, 1998, pp.283-292.

[6] S. Harnad. The invisible hand of peer review. **Exploit Interative**, v.5, April, 2000. Available from: <http://www.exploit-lib.org/issue5/peer-review>. Access: May 20th, 2008.

[7] M.M. Case. Igniting Change in Scholarly Communication: SPARC, Its Past, Present, and Future. **Advances in Librarianship**, v.26. Available from: <http://www.arl.org/sparc/bm~doc/advances_case2002.pdf>. Access: May 25th, 2008.

[8] R.K. Johnson. Partnering with faculty to enhance scholarly communication. **D-Lib Magazine**, v.8, n.11, 2002. Available from: <http://www.dlib.org/dlib/november02/johnson/11johnson.html>. Accessed: May 24th, 2008.

[9] D. Prosser. Information revolution: can institutional repositories and open access transform scholarly communications? **The ELS Gazette**, v.15, July 2003.

[10] S.M.S. Costa; A.C.S. Moreira. **The diversity of trends, experiences and approaches in electronic publishing: evidences of a paradigm shift on communication**. In: S.M.S. Costa; J.A.C. Carvalho; A.A. BAPTISTA; A.C.S. Moreira. From information to knowledge: proceedings of the 7th ICCC/IFIP International Conference on Electronic Publishing. Guimarães: Universidade do Minho, 2003, pp.5-9.

[11] T. Brody; S. Harnad. **The research impact cycle**. Available from: <http://opcit.eprints.org/feb19oa/harnad-cycle.ppt>. Access: May 24th 2008.

[12] S. Lawrence. **Online or invisible**? Nature, v.411, n.6837, p.521, 2001. Available from: <http://citeseer.ist.psu.edu/online-nature01/>. Access: May 25th 2008.

[13] C. Nicolai; I. Subirats; S. Katz. **The FAO Open Archive**: Enhancing access to FAO publications using international standards and exchange protocols. Proceedings ELPUB2007 Conference on Electronic Publishing – Vienna, Austria – June 2007. pp.141-151. Available from: < ftp://ftp.fao.org/docrep/fao/010/ai166e/ai166e00.pdf >. Access: May 25th 2008.

[14] A.J. MEADOWS. A comunicação científica. Brasília: Briquet de Lemos, 1999. 268p.

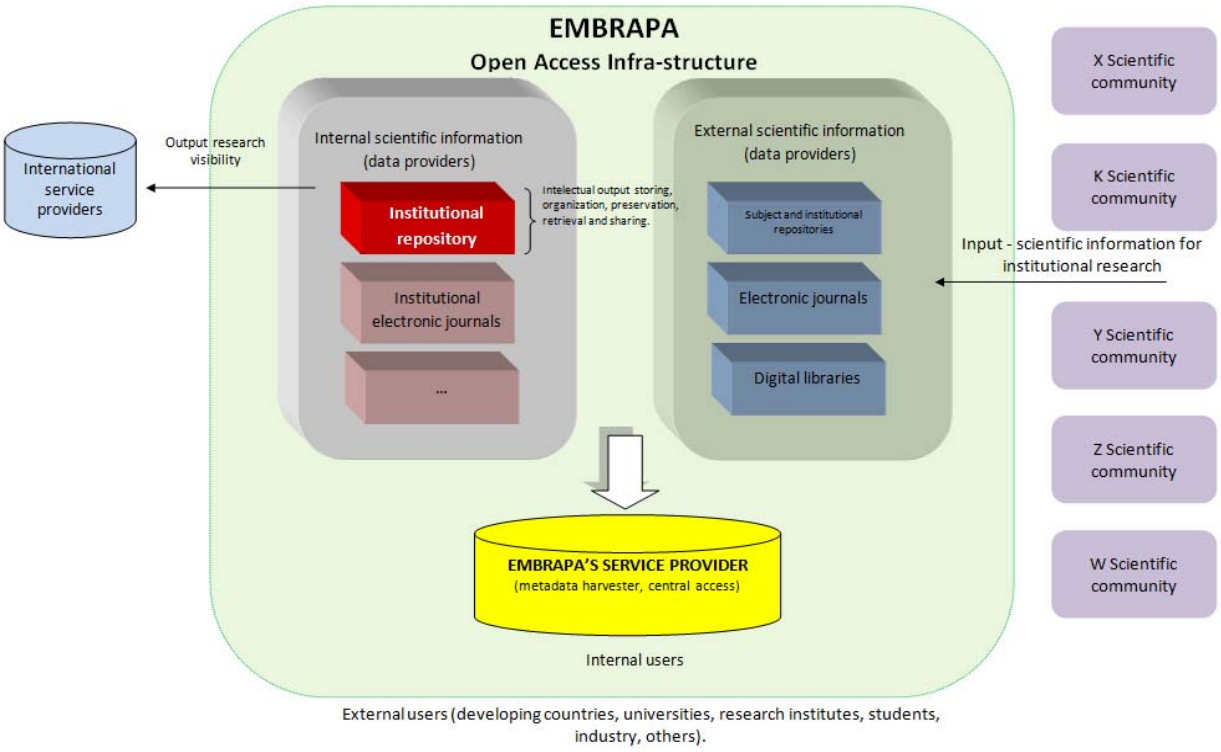

Figure 1**. Methodological model for scientific information and knowledge management at Embrapa.**