Providing services to INRA and the international agricultural community: ProdInra & the Integrated Information System.

LE HENAFF Diane 1, AVENTURIER Pascal 2

1. DV-IST, INRA, France, lehenaff@versailles.inra.fr
2. ERIST, INRA, France, pascal.aventurier@avignon.inra.fr

Abstract

The main benefit of open archives is to optimize the visibility of research results. Universities & research institutes have been setting up OAI compliant repositories with this aim in mind.
INRA, a French organization and also the largest agricultural research institute in Europe has set up such a repository in 2005.
INRA needs to identify and track its production because it must evaluate individuals, the labs, the departments, etc. and also to produce strategic indicators about the trends of our research activity.
These indicators help INRA to direct funding and access the most advanced projects.
In 2005, INRA planned a 5-years project to create an Integrated Information System (IIS) that combines different internal applications in order to exchange data and create services for the INRA research community. Directories of INRA people, activities and research laboratories are available in the internet.
ProdInra aims to give a large audience to INRA products. As it becomes an applicative tool of this Integrated Information System (IIS), it can also provide services to the agricultural international community based on data coming from the IIS.
This communication will present the Integrated Information System as well as the different services provided.

Introduction

The French National Institute for Agricultural Research (INRA) is the world's second-largest research institute in the field of agriculture, food and nutrition, and the environment.
Its international visibility and recognition are a result of the widespread dissemination of its research work.
In 2005, INRA demonstrated its support for open access by signing the Berlin declaration and launching an institutional open archive called ProdInra.

Since that year, INRA has been overhauling its information system in order to set up indicators for steering agricultural research, but also to provide multiple services to INRA personnel as well as the agricultural scientific community at large.
This article sets out to demonstrate the usefulness of an open institutional archive that can provide a multitude of services to the research institute and its partners while meeting the goal of highlighting the value of work produced by INRA scientists.

First, we will describe the various data sources from which information is aggregated, as well as how the data is reintegrated into the IS. We will then focus on ProdInra's role as an open archive, and discuss

INRA's documentary policy, service, and goal of optimizing ProdInra data to enhance international visibility.

ProdInra: the INRA Information System's document block

Data sources

To address the growing needs of indicators, consolidate information, and eliminate redundant entries, INRA decided to overhaul and modernize its information system.
Initially, business applications such as human resources management and accounting were modernized. Shared data repositories (research activities and structures such as laboratories, research departments, etc.) were then created, forming the initial scope of an institutional information system.

Meanwhile, the assessment of research bodies, laboratories and scientists is based on scientific productivity. ProdInra, introduced in 2006, has thus become a strategic application in its own right. Consequently, the open archive was revamped so it could be integrated into INRA's information system. For the first time, a management tool for scientific publications was included within the scope of this institutional information system.

Data is exchanged in XML format between these application blocks via an Enterprise Service Bus (ESB) using webservices. Each application is thus potentially a data provider and a client, operating according to the diagram in Figure 1.

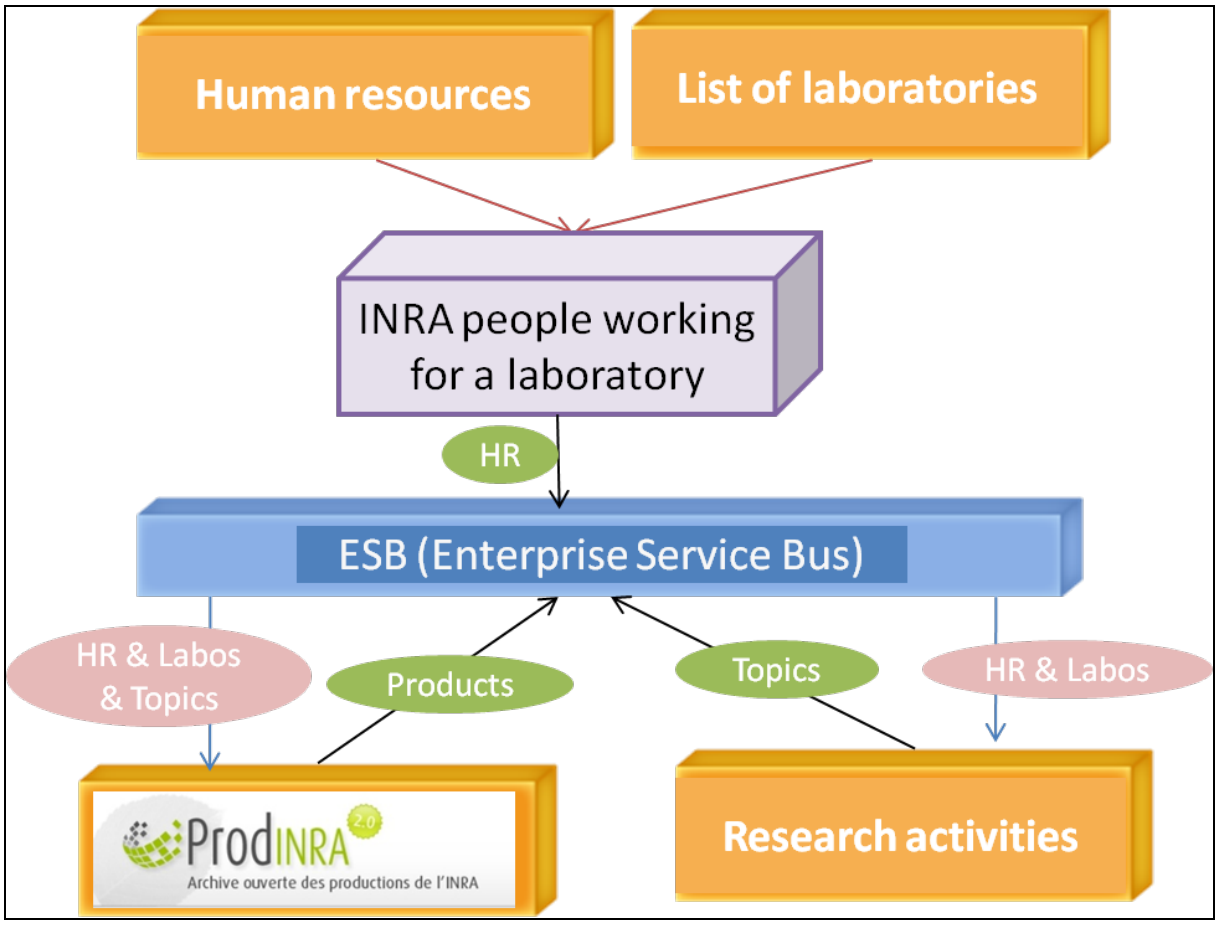

Fig. 1. Data exchange within the Information System

Forms of data restitution

The available data is first used to populate other IIS applications so that redundant data entry is avoided. Most of the data come from repositories, ensuring data consistency and offering the possibility of generating indicators and automated quality reports. This is made possible by the aggregation of data from the IIS.

Open services

The most useful function is the creation of directories. Directory entries for scientists can include a list of research activities and work. Directory of activities and laboratories are also available.
Since December 2009, all three directories can be consulted over the internet. This policy shift reflects our desire to promote INRA's research activities and skills, and make it easier for the international scientific community to find collaborators at INRA.
The diagram on figure 2 presents the service layer (green) resulting from the aggregation of data from application blocks (in orange).

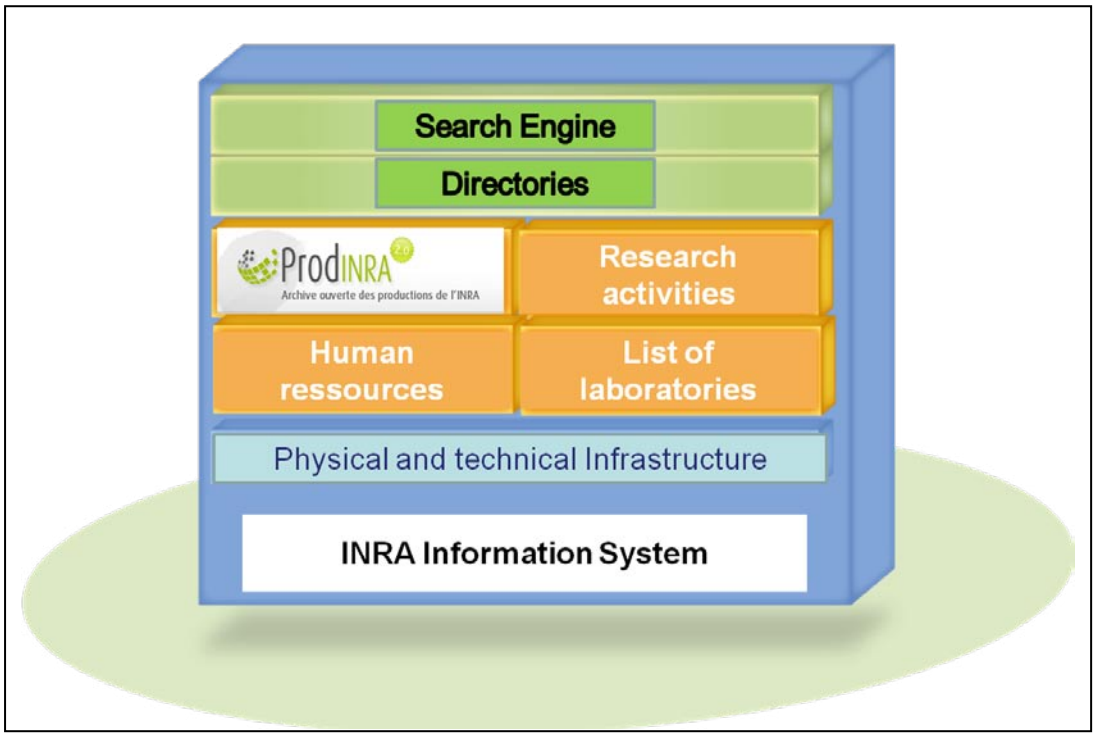

Fig. 2. INRA IIS & open services

Creating indicators

Leadership in agricultural research is determined on the basis of indicators: the number of academic publications, number of patents filed, etc. The ability to measure ratios—the number of

publications/number of scientists, or the number of publications co-written with non-INRA authors/number of publications—makes it possible to measure INRA's productivity and its openness to collaborations.
With data from human resources, funding, research activities and production, it is theoretically possible to generate detailed reports that can be used to steer INRA's strategy and orientations.

In practice, application administrators are often faced with the problem of the poor quality of repository data or incomplete data sets.
This problem does not spare the INRA open archive which, with its 124,000 references, is far from achieving a satisfactory level of referencing.
The advent of ProdInra 2, the new document application, represents a chance to mobilize the parties involved (submitting researchers and documentalist-editors) to support a policy that strongly encourages users to systematically deposit their work in ProdInra. This aspect was highlighted in an INRA report on open archives (L'hostis and Aventurier, 2006).

ProdInra : the INRA open archive

Documentary policy

As stated earlier, the ProdInra document database is incomplete in terms of referencing scientific production; it also contains little in the way of full texts (see table 1). Using this as a starting point, the Scientific Information team working on ProdInra came up with a new documentary policy.

This policy is centered on the systematic deposit of productions by the authors themselves. In practice, documentalists found themselves filling this role and were confronted by the problem of tracking and accessing the document. Document deposits will now be carried out by researchers, who will then be able to retrieve their list of publications in the requested format in researcher and research unit evaluation dossiers.
Meanwhile, the ProdInra administration team will work on identifying productions in Scopus/WoS so that a full survey of the best academic publications will be possible using ProdInra.
Finally, the consolidation of full-text articles in ProdInra will be conducted in two ways:

- Encouraging the researcher to deposit the document at the same time as its bibliographic information, in accordance with the publisher's policy. A new submitter's chart is being drafted.
- Negotiation between Scientific Information heads and major scientific publishers to allow the automatic archival of published articles after an embargo.

Table 1. ProdInra content (+ 3 years and + 4 months)

Date	Number of references	Number of FT records	Part of Public FT records
Oct 06	101 214	0	0
Sept 09	120 953	4 498	3 591
Jan 10	123 515	5 693	4 729

Data flows

Inputs

Submission is made tedious by the large number of bibliographic entries that need to be recorded. We therefore sought to develop a means of importing notices once such notices have already been entered somewhere. For example, using the PubMed Identifier (PMID) automatically retrieves information from the PubMed reference in Prodinra.
This feature is even more eagerly awaited owing to the fact that part of INRA research involves the biomedical field.
It is also possible to import notices listed in EndNote and RIS formats used by bibliographic reference management software. Researchers using such tools for their work only need to import this list in ProdInra and consolidate it by adding the full text.

Finally, INRA, like many other French institutes and universities, pools resources with its partners in joint research structures. Some scientists from these laboratories submit their production directly into the HAL open archive. HAL, or *Hyper Article en Ligne*, is the French research open archive. We have therefore developed gateways to exchange data between ProdInra and HAL

Outputs

Research is the most obvious function of an open archive. Nonetheless, we have developed associated services as well as exports so that ProdInra data can be used, and scientists receive services in exchange for their efforts to signal their production.
Several exports are available and we have planned to offer very soon exports of lists of scientific productions in the format used for evaluations.
These services include also features tied to new uses of the web: RSS feeds, social bookmarking, etc., as shown in Figure 4.

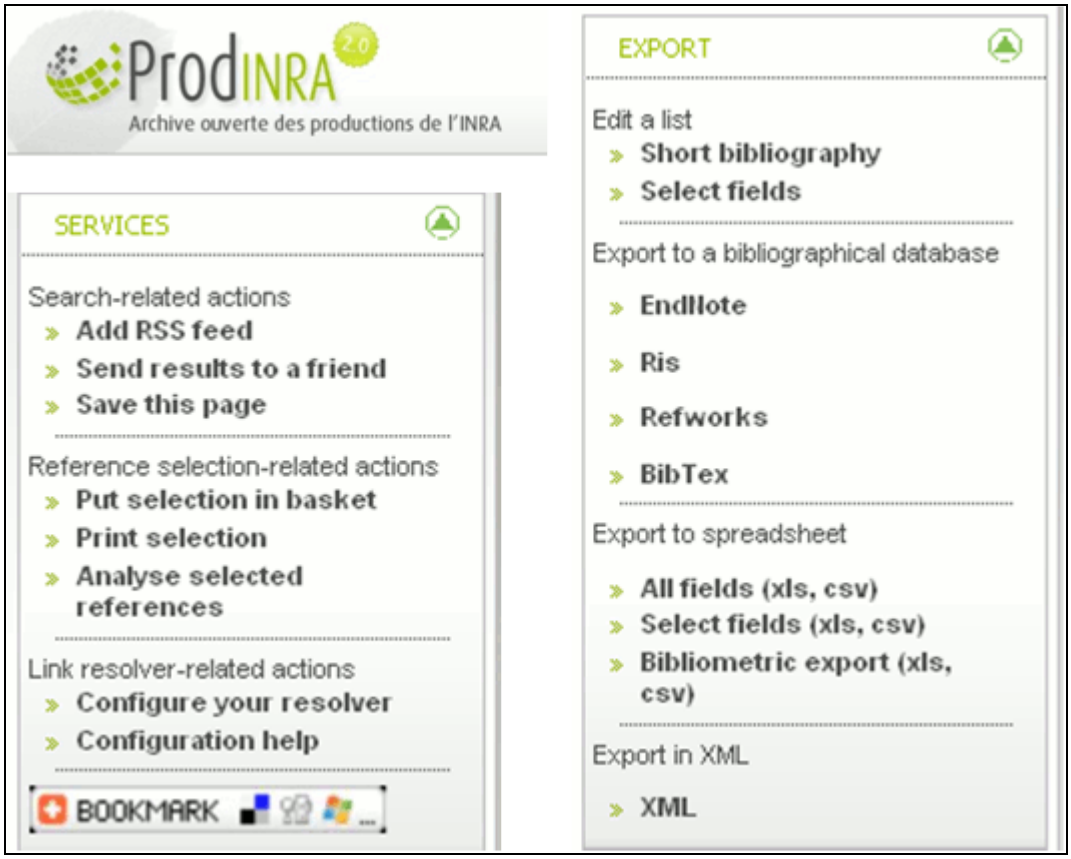

Fig. 4. ProdInra Services & Exports

Highlighting the work of INRA

In its four-year strategic contract, INRA restated its desire to develop its international activities. Promoting INRA research and highlighting its scientific and technical production internationally serve to generate collaborations and contribute to INRA's image as a major player in international agricultural research.

The ProdInra staff has therefore rolled out collaboration strategies in order to promote INRA's scientific production by having it referenced in numerous major document databases.

In addition to data exports to HAL, ProdInra will make its data available in Agris AP format so that the FAO's Agris database can perform OAI harvesting.
We are also exploring the possibility of exporting our production from research in economics and social sciences to rePEc, the international thematic archive for economics.
Recently, we have made contact with the TGE (large equipment) SHS (Social and Human Sciences) project of the CNRS (French National Centre for Scientific Research), known as TGE Adonis, in order to promote the sociological aspect and social relevance of INRA research.

Fig. 5. Referencing of INRA production in partner applications

To increase the dissemination and optimization of INRA research, the ProdInra staff has also opted to expose its production in OAI-MODS format. Once data is in MODS, a format created and maintained by the Library of Congress, we hope to form several other partnerships on the basis of lower costs for disseminating ProdInra data.

Finally, INRA is one of 14 partners in a European consortium involving 11 countries, for the VOA3R project (Virtual Open Access Agriculture & Aquaculture Repository) aimed at disseminating scientific research in agriculture, food and nutrition and the environment using a thematic technological & semantic platform.

The project began on 1st April 2010 and will run for 3 years, with funding of 2,000,000 € from the European Commission as part of the 7th FP.
By participating in this European project, the ProdInra team falls in line with INRA's international ambitions and hopes to attract an international audience to the work of its researchers.

Conclusion

INRA has chosen to invest in an open archive that forms part of the institute's information system, rather than an independent application based on a software program commonly used by universities and research bodies worldwide, such as DSpace or Eprints.
The additional costs stemming from this integration will be justified, it is hoped, by the benefits in terms of indicators and services.
The first part of this article described such benefits but also underscored the need for quality data and the completeness of the corpus being used.
This implies the introduction of a demanding documentary policy in terms of goals and resources.
The second part of this paper describes the ProdInra documentary policy, services and exports, but also lists ongoing partnerships aimed at promoting the work of INRA‘s scientists. These partnerships, whether concluded, ongoing or upcoming, demonstrate IST’s desire to be a full participant in INRA's international ambitions.

References

Report

L’hostis D. and Aventurier P., 2007. Open archives : Towards a policy of mandatory deposit ? A summary report on current developments, researcher practices and the role of institutions. 45p, http://www.prodinra.inra.fr/prodinra/pinra/doc.xsp?id=PROD20062cdda66f

WebSites

INRA Directories, http://www.international.inra.fr/research/directories

INRA Open Archive : ProdInra, http://www.prodinra.inra.fr

HAL, Hyper Article en Ligne, http://hal.archives-ouvertes.fr/index.php?langue=en