Title: Andalusian Health Repository: promoting the scientific health output among professionals and providing citizens with quality health information.

Toro-Sánchez-Blanco, Pilar; Matamoros-Casas, Teresa; Juan-Quilis, Verónica
Biblioteca Virtual del Sistema Sanitario Público de Andalucía. Consejería de Salud de la Junta de Andalucía, Seville, Spain

Introduction

The Andalusian Public Health System (Sistema Sanitario Público de Andalucía-SSPA) Repository is the unique open environment where all the scientific output generated by the SSPA professionals, resulting from their medical care, research and administrative activities is comprehensively collected and managed. It was created with the aim of assuring its long-term preservation, promoting open access policies and disseminating knowledge among the scientific community and citizens.

The Andalusian Public Health System Virtual Library (Biblioteca Virtual del Sistema Sanitario Público de Andalucía-BV-SSPA), was entrusted with carrying out the repository development and maintenance.

The creation of the repository constitutes an Andalusian Regional Health Government commitment to using this key tool to manage knowledge at a regional level, to promote health research among professionals, who work particularly in primary healthcare centres, and to provide citizens with quality health information.

- This repository possesses special features which determined its development:
  - the SSPA consists of many different kinds of institutions and centres which carry out activities such as health care, research, administration and professional training. It has 45 hospitals, 1,520 primary healthcare centres and more than 20 research and teaching centres. They are dispersed all over the Andalusia region in the South of Spain and are run by 102,500 professionals for a population of more than 8 million inhabitants.
  - Apart from monographs, published by some SSPA entities, or articles published in scientific journals with international reach, which have been evaluated by an editorial committee or by a peer review system, the SSPA also generates a very specific set of document types such as plans and health programs, guides, healthcare protocols, research projects, materials to help patients to make decisions and citizen education handouts. An important proportion of these particular documents is produced within the centres and is not published externally.
The stakeholders of the project are the SSPA and its professionals, the authors of the scientific output, public funders, the national and international research community and citizens.

Objectives

With the setting up of an open and interoperable repository the SSPA wants to achieve the following goals:

- To collect and manage in one place research outcomes, scientific output and quality health information for patients and citizens. This information was previously scattered around many institutional websites and intranets of the different centres and institutions of the SSPA.
- To guarantee the long-term preservation and access to all the information generated by the system professionals.
- To facilitate the knowledge transfer among the health system professionals and turn scientific discoveries into clinical practice.
- To share and disseminate research outcomes, protocols and guides, among others, within the SSPA and the scientific community.
- To provide quality health information for patients and citizens.

Material

We decided to use DSpace for several reasons. The most important ones were that it is an open source software, which complied with the Andalusia Government standards and policies; it has a large community of users and developers; and it covered our main specifications, including open access policy requirements. We are using version 1.6.2 of DSpace.

The SSPA Virtual Library is responsible for the repository development, the integration of DSpace with other tools, such as its metasearch, and the subsequent management of the repository. In those SSPA centres which have their own library, the collection management and items ingest is carried out by that library. In those where there is no library the input of their items is made by batch ingest.

Methods:

Within this environment there are several steps that we had to accomplish in order to set up the repository and to run an initial ingest of documents.

Firstly, due to the set structure of DSpace around communities and collections and the specific organization of the SSPA, we decided to establish a hierarchy of communities and sub-communities taking into account the type of centres and their activities. As a result there are five communities at the top level: Consejería de Salud (Andalusian Regional Health Government); Servicio Andaluz de Salud- administration support services; hospitals; research and teaching centres; and primary healthcare centres.
Secondly, we established the collections based on the different kinds of documents and the particular submission policies to be complied with. Up to now, we have defined six collections, but it may be increased in the future. Furthermore, taking into account that it is an institutional health system repository it was essential to do a study of the type of documents that were going to be collected, in order to:

- define an exhaustive metadata scheme to describe and retrieve the documents. Therefore, we included specific qualifiers in the Qualified Dublin Core Metadata contained in DSpace.

- establish strict quality criteria to be fulfilled by every document prior to being submitted to the repository. There are three sets of criteria: Format, Content and Legal.

After carrying out the study of the different metadata which were needed for each type of document, we also personalized the document submission form, providing a simpler submission process. Therefore, at the beginning of the submission the user chooses the type of document which will condition the metadata present in each of the subsequent screens.

In order to maximize the performance of the repository and provide more flexibility when retrieving contents, the advanced search form was also customized, including some of the specific metadata introduced in the Qualified Dublin Core scheme.

In addition to these changes to the software we considered it necessary to use, as subject headings, the MeSH (Medical Subject Headings) Thesaurus created by the U.S National Library of Medicine (NLM) to describe the Health Sciences production. This thesaurus is composed of more than 26,000 terms and nowadays there is no repository that has integrated the MeSH descriptors as controlled vocabulary in DSpace. Assigning MeSH descriptors to each item during the submission process increases the possibilities of retrieving precise and relevant information. To achieve this goal it was necessary to develop a script to transform the MeSH terms file, which was downloaded from the NLM, into an XML file used by DSpace.

Along with the integration of the MeSH to classify the content of the repository, we developed an interface to search by MeSH terms.

Another challenging stage prior to launching the repository was to run a batch ingest of the scientific output of the SSPA published in scientific journals, which was indexed in international bibliographic databases from 2006 until now. This scientific output is estimated to be 18,000 documents. This task was done using an internal tool called Impactia, a piece of software developed by the Virtual Library, which processes the output from different databases, such as WoS (Web of Science, Institute for Scientific Information), Scopus from Elsevier/() or PubMed from/ (NLM), and extracts the bibliographic records from a specific set of centres.

Although the batch ingest was done through command line, a later review was needed from the administration web interface to complete the metadata, assign the MeSH terms, check the contents and upload the associated files.
Results

To sum up, the Virtual Library of SSPA has used DSpace 1.6.2 and adapted it to obtain a powerful tool for its Health Sciences Repository. In order to achieve this, there have been several key changes in the application:

- Adaptation of the Qualified Metadata Dublin Core scheme, adding specific qualifiers to some elements such as subject, identifier, type, relation, etc.
- Modification of the submission form according to the different types of documents generated by the SSPA, especially by hospitals and primary healthcare centres.
- Optimization of the advanced search, increasing the number of fields and offering an interface search by MeSH terms.
- Use of a mechanism to run batch ingests of metadata records.

Conclusions

The existence of the repository, as the only open and interoperable environment that collects and manages all the scientific output of the SSPA, yields great profits for the SSPA because the information can be accessed more widely by professionals and citizens. It especially enhances access to unpublished materials. Furthermore, the repository reduces duplication of work in drawing up healthcare protocols and guides; it facilitates the dissemination of knowledge and experiences among professionals within the SSPA; it meets the expectations of professionals related to the access and preservation of their unpublished materials; and it provides citizens with the same quality health information that adheres to the SSPA principle of fairness.