CRIS 2014

National Open Access Scientific Articles Registration System (NOSARS)

P. Mehrtash[a], O. Fatemi[b,a]*

a Iranian Research Institute for Information Science and Technology, Tehran, Iran
b University of Tehran, Tehran, Iran

Abstract

Research information is becoming increasingly important in the knowledge societies. In Islamic Republic of Iran, a national current research information system has been established to manage all research information. Scientific journal articles, as the outputs of the research, are one of the main entities in CRISes.

In this paper, we show the process of the establishment of our National Open-access Scientific Articles Registration System. We have proposed to employ registration process for collecting all related articles. Our experience shows that this method has achieved a great success in establishment of a complete, current and valid NOSARS.

Keywords: Open Access Repository, Registration System, Scientific Journal Article, National Repository

1 Introduction

In Islamic Republic of Iran, as Iran 1404, the Vision Document or 20-Year Document dictates the country should be the first in the region. (Iranian 1404 Perspective Document, 2006)

This has resulted in a national research movement and hence, the number of published scientific documents has been considerably grown in the recent years.

Every research is indisputably related and associated to the previous researches in the same field. To make the research output accessible for its followers, it is important to index it through a set of standard metadata in which research objects could be retrievable for the other researchers. We note that this is true for both our national publications and also for world publications; however the focus of this paper is on our own national publications in

---

* Corresponding author. Tel.: +98-21-61114921; fax: +98-21- 88633029.
  E-mail address: omid@fatemi.net

order to bring maximum visibility and accessibility of them for the researchers.

To achieve this purpose, national current research information system (CRIS), named as SEMAT, was designed to provide a collaborating environment across institutions by integrating national scientific data. (Khoshroo and Fatemi, 2010)

As explained there, SEMAT is only integrating metadata of research entities. It is designed to make the research information discoverable for researchers. However, the full texts of the research objects (such as journal articles) are required for scientific work. There is an ongoing effort in Iran, to make individual full text repositories for every research object. In this paper, we propose the methodology of making a national repository of full texts of scientific journal articles, as the peer reviewed outputs of the research.

There have been many attempts in the world to make government funded articles accessible through institutional open access repositories or an integrated portal of them. [A National Open Access Policy for Developing Countries, India (2006)] . Going through these national systems, we note that a number of them typically don't provide full text, a few of them consider embargo period for making full text accessible. The incomplete coverage of all institutes is another issue with these repositories.

To overcome these challenges, we designed a national open access registration system. This system collects all accepted scientific articles meta-data and full text directly from the publisher's online submission system. Scientific journal articles are provided to researchers as soon as the article is accepted.

In the next section, the value of the full texts of scientific articles will be discussed followed by the best practices of national repositories in the world. In section 4 the methodology of designing NOSARS will be presented. Section 5 shows the selection process of the institute in charge for implementing the repository, the establishment of NOSARS followed by statistics of the system. In section 6, the integration between NOSARS and SEMAT is presented followed by conclusions and references.

2   The value of scientific journal articles and open access to them

Research has numerous outputs and products such as books, book chapters, conference proceedings, dissertations and journal articles. However, scientific journal articles have a particular position among the various forms of research information. This position is due to the accuracy and improvement obtained by different views which should be satisfied in article acceptance process. Published journal articles receive independent and neutral experts' critical comments and suggestions in a peer review process. This leads to a great tendency among researchers having access to journal articles and the latest findings beyond restrictions.

This tendency led to a new approach in self-archiving a decade before emerging the term "open access". Later, profiting from the World Wide Web and electronic publishing, open access was born and developed by researchers as a solution to provide maximum access to journal articles. Providing maximum access to articles means increasing articles visibility and thus growing their impact factors.

Due to the expanding importance of open access to scientific journal articles, in the next section we will present the achievements of forming national repositories of journal articles around the world.

3   National repositories of journal articles

As mentioned in the previous section, researchers have always desired of having a free and complete access to the scientific peer reviewed literature. This tendency among the academic society led to the idea of mandating research grant recipients through their research institute to make their published, peer reviewed journal articles accessible via publishing them in open access repositories.

As of December 2013, open-access mandates have been adopted by over 240 universities and over 90 research fund agencies worldwide (ROARMAP, 2014). We now introduce a number of national open access repositories in the world.

3.1   Ireland

In 2007, Irish Universities Association had established open access institutional repositories in each Irish university and developed a federated harvesting and discovery service via a national portal with government funding. This project's first phase has launched in 2010.

Until March 2014, 16333 journal articles have been published through this repository by 7 member universities, which has been increased to 13 now.

The main idea is to make Irish research material more freely accessible, by harvesting the contents of member institutional repositories through a portal. Increasing research profiles of individual researchers and their institutes

was another outcome of the national network of local repositories. We note that this system which is called RIAN, does not contain the full-text files of items, but provides a link back to the source repository where the full-text file can be downloaded. (Repositories Support Ireland, 2014)

3.2   Netherland

All Dutch research universities have one or more repositories. Central access for these repositories is provided by the portal of NARCIS (The gateway to scholarly information in Netherlands) which provides access to digital open access publications at 25 repositories.

At 9 universities the depositing of doctoral theses in the institutional repository is mandatory. The universities of applied sciences work together in collecting the materials deposited in their repositories and presenting theses through one portal, which provides open access to research outputs, educational materials. (Open access in Netherlands 2010)

3.3   Sweden

Almost all universities in Sweden have open access repositories, the majority of them are members of a consortium. Others have implemented open source software or created their own publishing platforms. Today most higher education institutions have integrated their open access repositories with their publication database. These are supposed to include meta-data from all the academic publications of the institution. The SUHF (Swedish Association of Higher Education) have been active in promoting open access among member institutes, but has no mandate to make decision on their behalf. (Open Access in Sweden, 2010)

3.4   Germany

There are a great number of institutional repositories in Germany, which are maintained mostly by universities and research institutes. The German Initiative for Network Information (DINI) supports a national repository infrastructure.

German Research Foundation (GFR) has tied open access to its funding policy: Recipients of DFG-Funding (German Research Foundation) are expected to make their research results published and available, where possible, digitally and on the Internet via open access.

Meanwhile in the key organizations like German Research Foundation (GFR), Helmholtz Association and Max Plank Society either open access publishing is not mandatory or depositing research material to open access depositories is required within 12 months of publication. (Global Open Access Portal, Germany, 2014)

3.5   Summary and review

The work of making scientific production accessible for researchers of the country mostly done through a national open access portal integrating individual institutional repositories and their services differ from providing metadata to full text. In the other hand publishing accepted articles in the open access repository might be mandatory with or without embargo on accepted articles.

As a result of this section, it is evident that very significant national experiences have been formed regarding open access publishing, but the lack of a comprehensive set which could fulfil an immediate full access to all submitted articles around the country is observed. In the next section we will present Iranian approach to address these factors which could lead to a complete and current national repository.

4   The Design of "NOSARS"

As we have seen in previous section, there are a number of difficulties and shortcomings in national repositories which need to be addressed. Basically, two methods have been employed to bring the data to the repository.

- Data push: The institute is responsible for publishing / submitting data to repository.
- Data pull: The repository is responsible for gathering and harvesting the data.

In 'pull method', the link between the institute and the repository might become broken which causes the failure of harvesting while in push method the institute is responsible to maintain the connectivity and tries to keep the link available. For this reason in the rest of this paper, push method is considered. We note that using push method causes better availability and reliability.

Even using push method, the previously presented national repositories suffer from problems of lacking the following attributes:

- The completeness of the repository. The repositories even when there is a mandate upon institutions to submit their data, suffer from incompleteness of data. Not all the journals and institutions take part in the submission process.

- The validity of the repository. For the repositories where there is an extra submission process, the correctness and validity of data need to be investigated.

- The extra cost of submission to the national repository. To setup the extra submission process brings more expenses to the institute which makes this process less attractive for the institute.

- The recency of the repository. The data which are being submitted to the repository should be recent, up to date and current. If an extra submission process is defined in the institute there would always be a delay in submitting the data to the repository.

- The homogeneity of repository. Every institute uses its own method to prepare the data and submits it to the repository. Hence, there are discrepancies in format and details of gathered data in the repository.

4.1   The Process

Since the NOSARS is planned to act as a complete national repository, we started the process for establishing NOSARS through SEMAT and CSTIS ("Commission of Science and Technology Information System"). SEMAT, Iranian national CRIS, which has been introduced in CRIS2010 [Khoshroo and Fatemi, 2010] is managing research information across all institutions in Iran. On the other hand, CSTIS, a policy making and standard establishing body, has been introduced in CRIS 2012 [Mehrtash and Fatemi, 2012]. CSTIS is developing standards for meta-data management of all research objects in SEMAT and also is responsible to assign national level projects to Iranian institutes.

Taking advantage of international experiences in building national open access repositories and considering local requirements, we designed a national open access repository considering maximum capacity expected, in which all the problems above are being addressed and every article along with its full text to be accessible.

4.2   Registration System

Registration system means the data is registered and managed during the work flow of every individual object.

We propose to employ registration system in order to overcome the mentioned problems. The idea is to register and collect related data of the articles during the typical workflow of publishing an article in a journal which begins by submitting a manuscript by author as illustrated in Figure 1.
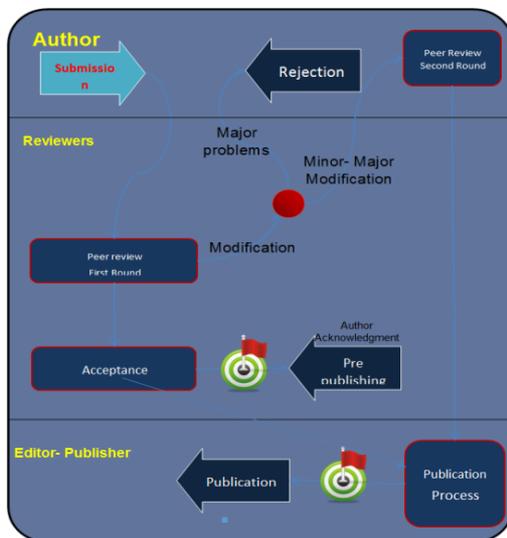
After the submission of an article by the author, the editor assigns the reviewers for the article for peer review process. This might lead to rejection, acceptance or minor or major revision. Author might be asked to revise and resubmit the article once or more until receiving the final acceptance from reviewers. In NOSARS, this point is considered the first mandatory point for registering the article and the state of the article becomes "pre publishing". All academic publishers are required by CSTIS to setup registration services for every accepted article. Actually "NOSARS" registration system communicates with publisher's online submitting system, once article receives acceptance.

The article is then being edited and prepared to be published in a volume in the journal. This state is publishing state. The minimum requirement for NOSRAS to accept registration of the articles is: "pre publishing" and "publishing" phases. In these two points, which are shown by flags in Figure 1, metadata and full text are being registered in NOSARS and thus the scientific information becomes retrievable for researchers.

Figure 1 - The typical process of article submission

4.3   The result of registration system

At the beginning of this section we listed problems of other national repositories. We note that using registration system will eliminate all of the problems.

- The completeness

Since every individual article is going through the publisher workflow, we can be sure about the completeness of the repository. Every article is passing by the flags shown in Figure 1 and therefore it will be registered in the system.

- The validity

Every individual article is going through a review process by the reviewers assigned by the editor. This peer review process and the final judgment by the editor are the best validation proofs.

- The recency

As soon as the article is being accepted by the editor, it will be registered and therefore will be available in the repository. This process assures the recency of the repository and hence, all the current information are being collected in the system.

- The homogeneity

To register the article of the data in the system, a web service is designed and implemented. The web service enforces one single format of the data being entered in the system.

- The extra cost of submission

Assuming the web service described above is being implemented and designed, we note that there is no extra activity is required and hence no extra cost is being applied on the institute. In other words, just by having a system set up for publication process, the system works smoothly without any intervention and any additional cost to put the date into the repository.

5   Selection Process and Establishment of NOSARS

After the design of NOSARS (which is explained in previous section), the next step was to assign the task to a public information center. In this section the steps for this selection is presented.

5.1   Call for Proposal

To implement the national open access repository our next step was to monitor information centers and research institutes around the country for their executive tasks and performances to determine the most appropriate as the organization in charge for steering the national open access registration system based on the issued policy and instruction. The main information centers in the country received call for proposal. The applicant organizations assumed capabilities were assessed based on criteria below:

5.1.1   Selection Criteria:

- **Functionality** – Ability to fulfil NOSARS instruction and requirements
- **Extensibility** – Ability to integrate external tools and software (various publishers online submission system with NOSARS)
- **Interoperability** – Ability to interoperate with SEMAT and other repositories.
- **System security** – Ability of the system to meet NOSARS security requirements.
- **System performance** – Overall performance and response time (accomplished via load testing). System availability internally and externally.
- **Platform support** – Operating system and database requirements. Staff expertise to deal with required infrastructure.

5.2   ISC

After evaluating received proposal, Islamic World Science Citation center (ISC) was qualified and assigned as organization in charge for implementing and steering NOSARS.

Islamic World Science Citation Database (ISC) is a citation index established by the Iranian Ministry of Science, Research and Technology after it was approved by the Organization of the Islamic Conference. It is managed by the Islamic World Science Citation Center, located in Shiraz.

5.3   Establishing NOSARS

Islamic World Science Citation center (ISC) received government founding to develop and equip a dedicated repository and coordinate registration system as "NOSARS".
The Ministry of Science, Research and Technology obliged all government funded scientific journals publishers to start registering their accepted articles in NOSARS and to publish the full text of articles through it.
"NOSARS" was launched in 31 August 2009 and it is up and running. It can be found at http://www.isc.gov.ir.

5.4   Statistics

 The system contains the information of 664121 Persian articles and 93008 English article, also 497 Persian and 8 English scientific journals. The number shows that the registration system has been successful.

6   The interoperability between NOSARS and SEMAT

As described before, SEMAT (Iranian national CRIS) is integrating metadata of all research entities. On the other hand, NOSARS is responsible to integrate all data related to scientific journal articles including full-texts and metadata. In addition to these two national systems, there are also the publication systems of every individual journal. The interoperability among these systems is explained below.

6.1   The Research Hub

There are lots of systems including research systems and non-research systems which involve in research information as shown in Figure 2. For example, human resource system as a non-research system of an institute is holding the information of researchers.
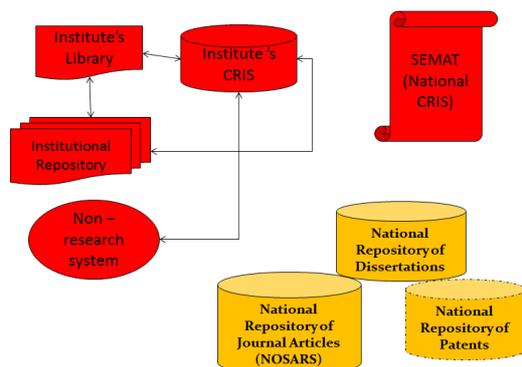

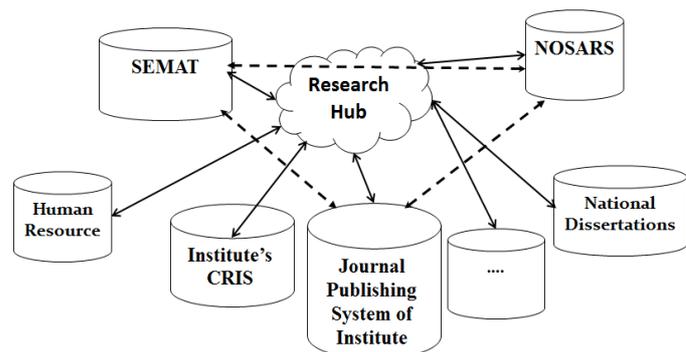Figure 2 - The variety of systems using research information   Figure 3 - National research hub

We have employed an enterprise bus system to act as the national research hub. Every system is physically connected to this bus through Internet as shown by solid lines in Figure 3.
By having the physical connection between every system and the hub, the logical connections between any two system is feasible as shown by dotted lines in Figure 3. Therefore any message generated in the journal publishing system of an institute is delivered to both NOSARS and SEMAT through this hub. The meta-data of the article is stored in SEMAT database while the full information including the full-text is stored in NOSARS.

6.2   Loosely Coupled System

Once the physical connections and therefore logical connections are established, all systems need to communicate with each other. Such a distributed environment introduces many issues like latency, synchronization and partial

failure of one system which need to be addressed. We have selected loose coupling using message passing mechanisms. Messaging protocols based on xml are defined. Each message is issued by an authorized issuing system which is the origin of the data. Each system sends the message to the hub where there is a message buffer. Whenever the connection between the target system and the hub is alive, the messages are sent to the target system.

7   Conclusions

We have presented the methodology for designing and establishing a National Open-Access Repository. Technical and non-technical requirements have been addressed. For technical side, we have implemented a loosely coupled system by message passing between all the actors in the system through a research hub. To make the repository complete, current and valid, we have proposed to employ registration system. All publishers are required to register the articles upon being accepted in the system. Our experience shows the design decisions made are satisfactory and the established repository outperforms similar national repositories around the world.

8   References

1. A National Open Access Policy for Developing Countries, India (2006) available at: http://users.ecs.soton.ac.uk/harnad/Temp/stronger-OApolicy.htm (accessed 8 February 2014)
2. Global Open Access Portal, Germany (2014) available at: http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/access-by-region/europe-and-north-america/germany/ (accessed 8 March 2014)
3. Iranian 1404 Perspective Document (2006), available at: http://vision1404.ir (accessed 15 April 2014)
4. Khoshroo, Mohammad Javad and Fatemi, Omid (2010), "SEMAT, National Current Research Information System for Iran", Paper Presented at Euro CRIS Conference, Aalborg, Denmark
5. Mehrtash, Pouneh and Fatemi, Omid (2012), "CSTIS, Policy Making Body for National Research Information System in Iran", Paper presented at Euro CRIS Conference, Prague, Czech Republique
6. Open Access in Sweden (2010), available at: https://www.openaire.eu/en/open-access/country-information/sweden (accessed 15 March 2014)
7. Open access in Netherlands (2010), available at: https://www.openaire.eu/en/open-access/country-information/netherlands (accessed 15 March 2014)
8. ROARMAP: Registry of Open Access Repositories Mandatory Archiving Policies (2014), available at: http://roarmap.eprints.org (accessed 20 February 2014)
9. Repositories Support Ireland (2014), available at: http://www.irel-open.ie/moodle/ (accessed 15 March 2014)