RePEc and Socionet as partners in a changing digital library environment, 1997 to 2004 and beyond

Sergei Parinov, http://rvles.ieie.nsc.ru/~parinov/cv_par.htm
Institute of Economics and Industrial Engineering, Novosibirsk, Russia

Thomas Krichel, http://openlib.org/home/krichel/cv.html
Palmer School of Library and Information Science, Brookville, NY, USA

Abstract: This paper examines the theoretical foundation and practical development of the the RePEc and Socionet.ru digital libraries. RePEc is a UK-founded but internationally operating digital library for the economics discipline. Socionet is a Russia-based, but multi-disciplinary digital library for the wider social sciences. In 1997, Socionet copied the business model of RePEc and much of its technical infrastructure. As the Socionet library has matured, it has diverged from the RePEc model. Currently it emerges as a model and platform to build the Russian national level scientific and educational digital information space.

Acknowledgements: The authors are grateful to the Ford Foundation, the Joint Information Systems Committee (JISC) of the UK Higher Education Funding Councils, the Open Society Institute, and the Russian fund for the Humanities for supporting activities described in this paper.

1. Introduction

The precise nature of digital libraries remains difficult to define. In the research area, we can see two approaches to the study of digital libraries. In the first approach, dominated by computer scientists, researchers are interested in building computer systems that will make it easy for a user to access an information resource. Such work typically involves the collection and the mark-up of data, as well as its presentation in an interface to the user. The concern for actual usage is secondary. In the second approach, more typically associated with the library and information science community, digital libraries are seen as collection of information items, typically as they are available on a web site. Researchers in this area are interested primarily in the work that the user does with these resources. The concern for improving these resources is secondary.

Our point of view is distinct. As trained social scientists, we see the digital library mainly as a social tool. It organizes the transmission from the providers of information to the users of the information, and, in more limited circumstances, the other way around. At the heart of the digital library, as we see it, lies the business model of the collection and dissemination of contents. This main concern of this paper is digital libraries business models. The term business model should be interpreted in a general meaning as a simplified vision of something that keeps somebody busy, rather than about operating a digital library as a monetary revenue-generating device. We conjecture---though we have no empirical evidence to back up this claim---that digital libraries are held back more by the lack of business models than by technical issues. We admit this is somewhat of a truism. It should come as no surprise that with computer technology advancing at the rate of Moore's law and with network throughput advancing at an even faster rate, social reality has yet to catch up with the opportunities that are offered by this new technology.

From that point of view, this paper illustrate our view of digital libraries as, essentially, tools for communication and documentation between users. This vision is distinct from the

received wisdom that places a single user at the centre of the action and looks at the digital library as a service to the user, irrespectively of how that service can be sustained. We believe that successful digital libraries will be characterized by levying a community of users who retrieve information from the library as well as contribute to it.

Academic libraries provide an ideal example for this vision of a digital libraries. In recent work, Krichel and Koenig (2004) have set out a theory that we see as an overall conceptual framework for our work. We will introduce this theory of an open academic library in Section 2 of this paper. In Section 3 and 4, we look at our own work. We can see here how the concepts of open academic libraries are applied in practice. Section 5 has a comparative discussion between both systems. Section 6 concludes.

2. The open academic library

In a recent paper, Krichel and Koenig (2004) have put forward a theory of open academic libraries. They start from the well-established idea of open access to primary academic publications, i.e. the type of work that usually appear in scholarly journals, but then turn to the concept of open access to large descriptive aggregates. In the established jargon, such aggregates are referred to as abstracting and indexing databases. Such databases prepare the whole body of academic documents for information retrieval and build a record of such documents that is useful for performance evaluation. Nowadays, there is much some speculation, see Arms (2000) for an example, that the former functionality can be complete conducted by machine indexing.
The latter functionality of recording not be achieved by computerized methods alone. Human intervention is required at many stages of the project. To provide such databases freely, they advance a concept they call the open academic library. This is a virtual organization that provides such data in a decentralized way, using and XML format shared by its provides as well as such means to transport that data around, such as, for example the Open Archives Initiative protocol for public metadata harvesting.

The sustainability of such an organization rests on its ability to levy effort from its contributors. Since academic publishing is done to gain visibility, the participation in an open academic library is like an advertising venture. Once the open academic library has reached a critical mass of contributors and users, outsiders have a strong incentives to join. Insiders keep their work up to date because the recording function delivers a precise picture of how well they are disseminating their work. The performance evaluation gives immediate feedback on the advantages of participation. It therefore encourages continued participation. Thus once a critical mass is reached, the open academic library sustains itself.

To show that such a critical mass can actually be reached, it is sufficient to find an example. Krichel and Koenig examine the example of RePEc. to show that RePEc is indeed an open academic library.

3. The RePEc project

This project is the classic example of an open academic library. It operates in the economics discipline. Economics has a working papers are early accounts of recent research results. They are issued by academic economics departments and other institutions that do economics research, such as central banks, for example. In the print days, working papers were circulated by exchange arrangements between issuers. Electronic dissemination started in April 1993, when Thomas Krichel put the first every electronic working paper in economics out on a Gopher server. This was the start of collection of electronic working papers that he kept as a hobby project. For some of them, he had the full-text for, for others he linked to the full text. This small collection, named WoPEc, was complemented by a much larger collection of bibliographic

references to working papers provided by Fethy Mili at the University of Montreal. This collection was called BibEc. BibEc and WoPEc continued to be the largest access points for economics working papers in the Internet until 1997. In that year the RePEc project was founded whereby BibEc, WoPEc and a small number of other initiatives created a platform to exchange data. The data was encoded in a purpose-built format called ReDIF, see Krichel (2001), and exchanged using a purpose-built transport protocol called the Guildford protocol, see Krichel (1999). Both are still used in RePEc. At the time of writing, in June 2004, RePEc describes 270,000 items of interest to economists. There are over 175,000 of these available online. There are130,000 working papers, 139,000, journal articles, 1,000 software components and 750 book and chapter listings described. These data are provided by close to 400 RePEc archives that provide ReDIF data laid out according to the Guildford protocol. In addition, RePEc contain 5,000 author contact and publication listings, and 8,000 institutional contact listings. The latter two are the crucial components. The institutional contact listings are comparatively stable, and can therefore be compiled by one person, Christian Zimmermann, an economics professor at the University of Connecticut. The author contact and publication listings come from the RePEc author service. The RePEc author service replaces an earlier service called HoPEc, see Barrueco Cruz et al (2000). Authors provide the service with contact data, and then find the articles that they have written that are included in the database. At the time of writing, over 75,000 items have at least one identified author.

RePEc is a collaborative effort that unites data from over 370 archives. There are basically four types of archives. Direct document archives furnish data about publications that were issued by the providers of the archive. This is the vast majority of archives. Intermediated document archives furnish data that has been provided natively in format that is not compatible with the ReDIF format that is used by the RePEc. Intermediated archives rely on the cooperation of the providers of the original data. There are three archives that contain data that has been gathered by robots or by hand, or from existing bibliographical collections from data sources without the collaboration of the providers. Such collection was done intensively in the earlier years of the project, but has now completely stopped. Finally there are some special-purpose archives. These deliver personal and institutional data.

RePEc data is encoded is “attribute: value” templates similarly to email or http header. Users do not access the data in that form. Instead there is a plethora of user services to which use the data. Some of these user services are in fact contributor services. In particular, the RePEc author service is a service by which authors provide some personal information and then create associations with the document data in the RePEc database. The most frequent association is authorship. Thus, we are creating a relational database that provides an author record that relates to document records in a many-to-many relationship.

The RePEc author service is unique to the RePEc digital library. Despite its seemingly simple infrastructure, it is becoming the key piece. The data that it produces is not only used by user services to put together a list of an authors’ work. It is also used to aggregate usage figures per person. Such usage figures are compiled across user services by the LogEc project, see http://logec.hhs.se. These figures give a good picture of how well authors are exposed.

RePEc was built through the sustained enthusiasm of a few volunteers. Currently, it sustains itself exactly as the open academic library theory predicts. If authors and there institutional representatives do not participate, they are missing out on dissemination opportunities. They can guess how much they are missing out on by looking at the figures provided by LogEc. Thus, they have good incentives to collaborate. As shown in Section 2, this is a key aspect of the open academic library.

4. Socionet

How does a project like RePEc adapt to the Russian context? In principle, Russia should be a fertile ground for scholarly communication innovation. During ten years of economic crises, there was a permanent problem with accessing new research publications and materials. As a result, researchers looked for alternative ways to get access to modern research information. One benefit of RePEc is better access to publications. In 1997 the Russian Virtual Laboratory for Economists and Sociologists (RVLES), with financial aid of Russian Fund of Research in the Humanities, created a mirror of RePEc. Before the end of 1999 the mirror was based completely on the RePEc business model. The RVLES mirror of RePEc had only one principal difference from other RePEc services: all instructions were in the Russian language.

In 2000, the Ford Foundation started to support a project to develop new services based on RePEc. The second Russian reincarnation of RePEc was called the Socionet, because the main idea of the project was a transformation of RePEc network of documents collections into a social network. The Socionet project works on building of personal set of tools, which should allow individual researchers and different units of academic organizations fulfil their activity within a professional information space established by Socionet. This project fits in well with the idea of the open academic library as advanced in this paper. But it does not work like RePEc. Socionet should be more flexible to structure of information flows disseminated in Russian social science community. As a result Socionet is multidisciplinary. Thus it require additional data types. Introducing these data types means loosing full compatibility with RePEc database. Partial compatibility still remains.

Socionet divides its interfaces into two sets: a) navigation and search for publications and datasets; b) personal tools for Russian speaking users to publish/manage their research materials and other types of academic information. Some part of the contributed data keeps its compatibility with RePEc.

The set of personal tools includes two separate systems: 1) the so-called “Personal Zone”, which gives users an online interface to maintain their materials at Socionet server; 2) the SocioArm software, which is Windows OS program to operate with research materials right on user’s local computer. The SocioArm has the ability to synchronize local collections with its remote Socionet version, but in general its functionally is narrow compared with the Socionet Personal Zone working online.

Technically, the implementation of Socionet fits well with the open academic library theory. It is using an XML syntax, available to multiple user services via protocols such as the Open Archives Initiative (OAI) Protocol for Public Metadata Harvesting and/or plain HTTP. Different types of users (see below) can freely contribute to the Socionet open library. Its data is exposed to OAI harvesting services.

Socionet has its own approach to solve the main problem of open academic library, that constant human intervention required to maintain it is very expensive. It is achieved by creating a complex digital environment for all members of its social science academic community. In this environment, members can occupy different roles, but they are using the same set of tools; and they act within the same information space. The environment produces a “single window” effect. This allows users to save the cost of learning to use different interfaces. Currently Socionet roles include “author”, “organisation”, “developer” and “reader”. Below we give descriptions of service models for each role.

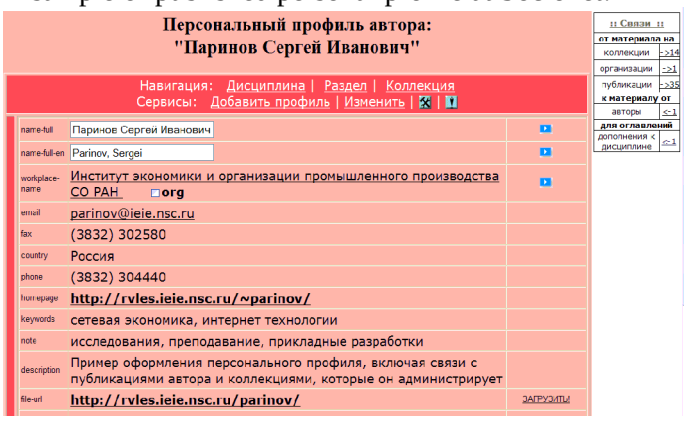
"Authors" are people who make their scientific results accessible to colleagues. The main author activity is to publish his/her research papers and other types of materials (like academic news, comments to published papers, conference articles, personal data, and etc.). Authors can create/manage whole personal collections of the same types of materials.

As a special type of publishing activity the "author" can create supplements for different materials in the database. There are two types of supplements, "comment" and "news". Supplements (if they do not violate the Socionet service terms) are publishing in special series of the Socionet open archives, provided that they do not violate the Socionet terms of service. Browsing as a "reader" through the database a user will get information about total number of supplements for the current (see the table "Relations" at the right side of the screen). Readers can open a list of supplements and read it. Supplements can be connected with following materials or database objects: a) a publication description (e.g. a review of the publication); b) a series content c) a section content d) a discipline content. There is a nested hierarchy to supplements to different objects. That means, supplements to a publication will be displayed in lists of supplements for the appropriate series, section and discipline. In the same way supplements to a series content will be nested into lists of supplements for its section and discipline, and so on. This service can be used for the following purposes: 1) To provide publications with comments, reviews or critical materials in which their content is analysed or discussed (only those supplements that conform to the scientific ethics requirements are allowed). 2) To supplement the publications with different kind of news relevant to the appropriate publication. For example, information on publishing of revised versions of the given material or on appearance of publications close to the original, on arrangement of events related to the original publication, etc.

The author can to manage his/her personal profile including links to own publications, series and organisation profile. Such personal profiles can be published by authors in a special series of Socionet. The Socionet database includes series of personal profiles for each discipline of Social Sciences.

A sample of online form at SPZ to maintain personal profiles:

A sample of published personal profile at Socionet:



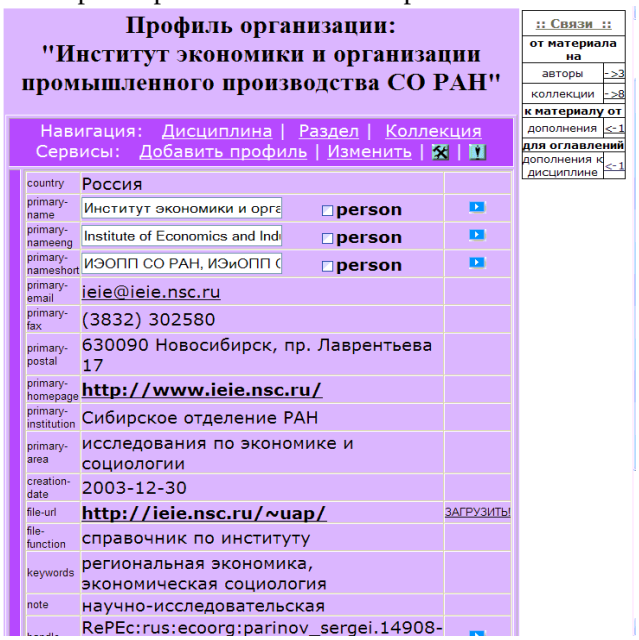
Notes: the right column includes links to different related materials from the Socionet database (like author's series and publications, his organisation(s), existed links to the profile from other materials, and etc.)

"Organization" is the second role that a contributor can have. In that case, the contributor is a representatives of academic organizations willing to publish their institutional data at the Socionet. A organization user has the same abilities as an author user, but additionally one can create/maintain institution's profile. This profile may contain links to series of materials published at Socionet to authors who work for this organization and have personal

profiles at Socionet and to related materials. Linking institution profile with related materials the theoretical principle - institutions should be identified rather than simply named - can be also implemented in the Socionet.

A sample of online form at SPZ to maintain an institution profile:

A sample of published institution profile at Socionet:



Notes: the right column includes links to different related materials from the Socionet database (like authors, series, related materials, and etc.)

Such institutions profiles can be published by "organisation" in special series of the Socionet Open Archives.

"Developer" is the third role. It is assumed by people who create information resources and online services addressed to research and education community in the Social Sciences. They can profit from the open Socionet platform in constructing their new online services. The platform includes tools enable development of information resources (series) integrated into the RePEc/Socionet database and/or other online OAI-based harvesting services; and also tools to build personal sites and services used parts/contents of the Socionet. After creating a Personal Zone in Socionet, a developer can design his/her personal site at Socionet server. In addition to textual content, the site can contain standard services configured by the developer to allow the site visitors: 1) to view the content of the thematic collection of series created by the developer on the basis of the RePEc/Socionet database and to subscribe to notifications of new additions to the thematic collection; 2) to participate in online polls, to propose a new poll subject, to view the statistics chart for the previous voting and to browse the archive of votings.

"Reader" is the fourth role. Readers are researchers, tutors and students in social sciences gathering information within their specialisation and tracing new publications in their field. They have full access to metadata and linked open full-text publications. They use the resources and services of the Socionet system as produced by "authors", "organisations" and "developers" under its instrumental platforms. Readers have following main abilities: a) to navigate through Socionet database and search by keywords; b) to accumulate favourite publication by setting up at SPZ personal collections of bibliographic descriptions (metadata) and add new records to them

while browsing Socionet database; c) to trace in automated mode all new additions to Socionet/RePEc database by own interest profile managed at SPZ. The last of these features is the most innovative. It is implemented using a service component called “Personal Information Robot” (I-Robot). It allows users a tracing of new additions to the RePEc/Socionet database following the assignments of its user (item "c" above). A user can create and run several independent and autonomous I-Robots to trace new additions by thematically different profiles. Each is configure to describe an interest. Then a list set the list of parameters related to periodicity and the form of reports, as well as the way of delivery of the reports.
A configuring of I-Robot includes following sections:

a) The "Customization" service deals with tracing the content of the selected subset of RePEc/Socionet series and identifying new additions. This service allows a user to choose the series that are of particular interest. The I-Robot keeps track of new publications in this group of selected series.

b.) The "Filtration" section provides control over the flow of new additions to the database. This service makes it possible to specify a group of parameters for automatic recognition of new and updated publications responding to user interests.

c.) Using the "I-Robot Report Manager", one can combine the reports produced by customization and filtration services in one report, select short or full form of the reports, specify periodical (e.g., daily, weekly, etc.) or quantitative (e.g., when more than 10 publications have been accumulated) parameters for report production, as well as the way of delivery (by e-mail) and/or saving in the user personal site.

The combination of abilities to publish materials into Socionet information space with its I-Robot tracing mechanism of new additions gives readers an implicit network of professional connections. A user can easy place his/her research material into professional database and at the same time he/she will be informed about the same actions of other members of this professional community.

Uptake of all these services has been moderate. After around three years of operation, in May 2004, we have the following figures. The total number of users registered at SPZ is 1081. About 16% of them passed registration, but never login to the SPZ. About 1.1% of users registered at SPZ use filtration service, 15% use customisation, and 20% of users created and manage personal series at their SPZ. During last 2 years they published only 104 papers at Socionet Open Archives, but they activated 884 open archives series in their SPZ. Most of these were created to publish one paper in the open archives. For this purpose one has to activate appropriate series in his/her SPZ. User created 504 personal series at their SPZ for different purposes and 271 of it were created with the type “secondary” for collecting descriptions of favourite publications found in Socionet database.

The services of Socionet are supplemented by the Socionet database. This database is developed much in the same way as RePEc is. Data is collected from a set of archives based at Russian partner institutions. These were developed with a range of Russian partners. It has been quite successful. For June 2004, we have the following figures.

6	social sciences disciplines	1,100	registered personal zones
8,000	working papers	75,000	journal articles
950	academic news items	48,000	book and chapter listings

This digital library is export to RePEc. However, few RePEc services have take in it up. The main reasons is that the documents are is Russian and therefore are of little interest to the average RePEc service user who does not understand that language.

5. Discussion

The Socionet platform is based on the following principle. A standardized set of data is assembled together in the single whole information space. Using a space-forming mechanism the data are put in online services. If the data at the output can also be standardized, a direct exchange of data between the services created by different developers is possible. In this case, the output of some services are the inputs of other services.

Socionet supposes that arrangements for interaction between different services and control over them can be integrated in the interface of the personal information robot, which is a part of the Socionet Personal Zone. As a result of configuring the chain of services by developers, a new service can be built. The assembled service can be included in the personal site of the developer in order to be accessible to all potential users.

The data exchange between the services through the information space would make it possible to establish: a) a "conveyor" of services; b) a set of interconnected services at different Internet servers, as well as distribution of tasks between different servers; c) implementation of the module principle of constructing of composite services, etc.

This should result in establishing a more efficient and competitive environment for constructing online services (as compared to that currently available in the Internet), since it enables the developers to benefit from the specialization and to save local resources at the expense of sharing results.

Such developments are still in their infancy. In the current stage, Socionet relies on essentially aggregating every aspect of its work in a single interface on a single machine. There is an obvious single window advantage to this setup. There are two problems with this setup. One is that the interface to the service can be highly complex. While the users who have learned the interface benefit from its single window, the users who come for the first time have a harder task understanding what is going on. This makes in more difficult to create a user community for the services. The other problem is that a complicated interface is more prone to bug and has higher costs.

The user services of the RePEc digital library have taken some steps towards distributed interoperable services. All use the same database, but each has its own brand name and marketing. In this environment, there is pressure on competing services to optimize their operations. The user can learn the interface of an individual service relatively quickly. But there are two problems. First, there is great confusion among users about the nature of the digital library. Users perceive a variety of brand names of each service. They may see that they have similar contents, and they may see that they refer to each other. But from all that we know about them, we understand that they do not grasp the concept of the open academic library underlying RePEc. The second problem is that it takes time for services to update. Once users understand that some form of common underlies several services, they become impatient for the updates that they make on one service to be implemented on other services immediately. Such an operation requires the ability of services to update incrementally, and it requires a signaling protocol from one service to the other. None are there at present. There is no coherent planing to RePEc.

On the contrary, setting up such a mechanism of data exchange between the services is envisaged in the plan for development of the Socionet platform for building-up of online services. In combination with other opportunities provided by Socionet, the platform will provide developers with both the means for integration of their services in the RePEc/Socionet database and the tools for designing individual modules and assembling composite services from available components.

The main next challenge for Socionet is to generally overcome the weakness of Russian system for dissemination of academic materials in digital form. We expect that Socionet online

personalised services to remain focused on supporting different roles of complex academic activity. They will produce favourable conditions to reduce "costs" for publishing academic information in digital form. The idea of future development is to promote the Socionet as a successful example and probably as a platform to create in cooperation with other organisations an open digital academic library infrastructure at a national level.

6. Conclusions

Building large complicated digital libraries is a very complicated business. In this paper, we have examined a reference model of open academic libraries, and two implementations, RePEc and Socionet.

Socionet has implemented many principles of open academic library concept. Technically, it is based on standard compliance, including the use of XML syntax and OAI/HTTP harvesting protocols. From a business point of view, it contain descriptions of almost all types of academic documents. And most crucially, it fulfils the open academic library's accounting function for authors and institutions.

Compared to the RePEc's implementation of this concept Socionet is not so popular as RePEc. But generally, it has been successful when it comes to usage of all its services. Overall it has better managerial control of its services. As a result, some of its tools are more advanced. Thus, while the RePEc concentrate on developing a single-discipline, multi-national model, Socionet focuses on a single country and a takes a multi-discipline approach. To date, RePEc has been more successful. But if, in the future, Russian authorities will implement a research assessment exercise like the one carried out in the United Kingdom, an infrastructure such as Socionet will come in as a very useful tool.

References:

Arms, William Y., (2000) "Automated Digital Libraries: How Effectively Can Computers Be Used for the Skilled Tasks of Professional Librarianship?", D-lib magazine, available at http://www.dlib.org/dlib/july00/arms/07arms.html

Barrueco Cruz, Jose Manuel, Markus Klink and Thomas Krichel, (2000) "Personal data in a large digital library", presented at ECDL2000, available at http://openlib.org/home/krichel/papers/phoenix.html

Krichel, Thomas, (1999) "Guildford protocol", available at ftp://netec.mcc.ac.uk/pub/NetEc/RePEc/all/root/docu/guilp.html

Krichel, Thomas (2001) "ReDIF version 1", available at ftp://netec.mcc.ac.uk/pub/NetEc/RePEc/all/root/docu/redif_1.html

Krichel, Thomas and Michael E.D. Koenig, (2004) "From Open Access to Open Libraries: Claims and Visions for Open Academic Libraries", presented at the second China Digital Libraries conference, Beijing, China, September 6 to 8.