ADEPT

Erik Sandewall

Agenda for the Experimental Electronic Press

Purpose and Topic Areas

The publisher of scientific journals is the traditional hub of the scientific communication process, and is likely to remain essential to that process during coming years. Other actors are also appearing, however, including local initiatives such as institutional repositories for research articles, and Internet-based actors such as Google and Wikipedia. One must therefore expect that the character of publishers and of publishing will change greatly, both because of the increased competition and because new technology is becoming available at a rapid rate.

The scientific library is a second, traditional hub for the preservation and the exchange of scientific information. Its mission is also changing rapidly: the preservation aspect is becoming less important, in the view of many, whereas the active dissemination aspect is being strengthened both within its own institution and beyond, in particular when the university library becomes the host of the institutional archive.

In fact, the borderline between the publisher's role and the library's role is becoming less and less clear. One should therefore not take for granted that traditional categories such as "publisher" and "library" are to be extrapolated into the future; maybe the future organization of scientific communication will have an altogether different structure.

The *Experimental Electronic Press* (XPEP) is a research project whose purpose is to explore these new possibilities by the design, construction and small-scale use of a prototypical infrastructure for scientific communication. As its name suggests, it incorporates the perspectives of an Internet publisher and an institutional repository, but it extends its range of activities in several directions. Our vision for what the XPEP should eventually contain consists of two parts: a conventional section that is oriented towards documents in the standard sense, and an exploratory section that brings in additional aspect. The conventional section includes:

- An *electronic publisher* facility whose purpose is to make research articles and associated research data persistently available on the Internet, for the duration of several decades at least

- Resources for *authoring support* that facilitate for an author or a group of authors to write the articles where their research is reported

- Facilities for maintaining the *personal library* of a user or the *group library* of a research group, consisting of electronically stored documents that the person or group has acquired, together with commentary information about them

- Resources for supporting the development of *research-group websites* which serve as a complementary way of disseminating the research results from a research group

- Resources for the *quality-based selection* of research articles and other research contributions and for several purposes, including both the traditional peer-review scenario and the traditional information-retrieval scenario

- Facilities for the *discussion and synthesis* of published research results, ranging from blog-like feedback on a particular article, to joint work towards survey and tutorial articles

The exploratory section contains the following:

- Sources for *raw information materials,* that is, collections of data that contain valuable information but which have not yet been processed and "cleaned up" as much as would be required for easy use

- Computational facilities for *understanding the contents* of documents and data sets that are stored in system. This includes thesauri and ontologies for representing that understanding, and analysis software for extracting some of the meaning of natural-language texts

- Facilities for *meta-level issues,* for example, assessment and discussion about the efficiency and reliability of various peer review schemes and of other methods for quality-based selection, and discussion of *research methodology* in the context of new technical possibilities for the communication of research results

- Resources for *technical support* of the above, for example, *client-side software* for use by authors wishing to make use of XPEP services, and information about the *conditions and restrictions for parallell publication* that are prescribed by various traditional publishers.

Most of these items but not all of them are already represented in the current prototype for XPEP which is located at `http://www.xpep.org` It is the *synergies between these facilities* that is the topic of investigation in the Experimental Electronic Press, including both design issues for a system where these facilities are to be integrated as seamlessly as possible, and assessment of the usefulness, convenience, and practical effects of having an integrated system of this kind.

A Note on Terminology

One may ask why we have chosen the term "press" for the XPEP - would it not have been better to call it a "publisher" or a "library", or to call it something entirely different like an "infrastructure" ? The reasons are very practical. We wanted to reserve the terms "publisher" and "library" for specific subsystems of the XPEP, as described above, and at the same time we wanted to use a name that gives a sense of continuity from the traditional context that is dominated by publishers and libraries. The term "press" then seemed like the best choice.

A second reason was that the name "Experimental Electronic Press" marked the continuity from its predecessor, the *Linköping University Electronic Press* which was an experimental enterprise when it started in 1996, but which of course proceeded fairly quickly to becoming a production entity within its university.

Software and Information Structure

The XPEP will be organized in terms of the following types of software constructs:

- *Hyperpages* for representing structured medium-sized texts. A hyperpage is a collection of webpages that address a specific topic, and is

often organized using a menu from which one can reach the various component pages of the hyperpage.

- *Software agents* that contain information for some of the purposes shown above, and that are able to communicate both with users, with each other, and with other information resources outside the XPEP. The use of software agents is in the spirit of the semantic web.

- *Knowledge modules* containing structured information, for example ontologies, scripts, and expressions in formal logic. Knowledge modules are publishable in the same way as documents are, they are intended to be produced by users of the XPEP and to be used by them, but they are also actively used in the operation of the XPEP itself.

More specifically, hyperpages are used both for conventional text, for hypertext containing links to other information, and for tables of contents containing links to documents or other information. Each software agent contains a web server, and some of the agents in the XPEP are used for publishing hyperpages. The webservers in the agents are also used for agent-to-agent communication.

Conventional documents such as research articles and reports are arguably a fourth type of software construct, but their presence of use is obvious in this context.

Initial contents

We have a number of currently operational systems that have been implemented using a common software base, and that can fairly easily be combined into a first version of the XPEP. The most important of these are as follows, using the same items as were shown above.

- An *electronic publisher* facility: the CAISOR archive at Linköping university and the PIEX archive at KTH

- Resources for *authoring support* the MADMAN facility that is in daily use and that has e.g. been used for producing the present report

- Facilities for maintaining the *personal library* of a user or the *group library* of a research group: not yet availabale

- Resources for supporting the development of *research-group websites* the MADMAN system provides this service as well

- Resources for the *quality-based selection* of research articles: the JARSS system

- Facilities for the *discussion and synthesis* of research articles: not available at present

- General: The CAPPA testbench for the scenario of article preparation, reviewing and publishing

- Sources for *raw information materials* the Common Knowledge Library (CKL) that is in operation at the CKL and its associated development software

- Computational facilities for *understanding the contents* of documents and data sets: not yet available

- Facilities for *meta-level issues:* the COAR hyperpage for information about open-access publishing issues, and the MORADOR hyperpage for issues in Methodology Of Research And the Dissemination Of its Results

- Resources for *technical support* the Leonardo software system that provides the agent technology for the XPEP, and the PARPUB knowledgebase for information about parallell publication conditions

Technological Strategy

All the resources mentioned above except the JARSS system use the KRF Knowledge Representation Format [1] and the Leonardo software system [2] which is uses KRF both for application data and for its own internal purposes. Facilities for document preparation, for message passing between agents and for operating web servers, for validation of structured data, for defining and using scripting languages, and for management of knowledge modules are already well supported in Leonardo, which makes it well suited for the needs of the Experimental Electronic Press. This software approach has been successful in the earlier projects and it will continue to be used in the works that follows.

Ideally we should re-implement the most important parts of JARSS on the basis of Leonardo, but this is an undertaking of nontrivial size. Furthermore, if and when JARSS is reimplemented it would make sense to organize it as a special use of some more general facilities, in particular with respect to how message exchange between persons is to be organized in the overall XPEP system.

Additional aspects of the technological strategy for XPEP are to follow in separate memos.

First Steps

The first steps will be:

- Combine the systems described above under a common umbrella for the purpose of presentation and visibility

- Review the existing and potential interfaces between these existing systems in order to improve their integration

- Review and extend the documentation of these systems in order to make them more easily accessible

- Pursue existing activities within some of these systems that we are already committed to through on-going projects

---

[1] http://piex.publ.kth.se/krf/
[2] http://www.ida.liu.se/ext/leonardo/

Next Steps

To be discussed.