Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2007 Proceedings

European Conference on Information Systems (ECIS)

2007

A Peer-to-Peer Application System for Scholarly Communication

Lutz Seidenfaden
University of Goettingen, lseiden@uni-goettingen.de

B Ortelbach
University of Göttingen, bortelb2@uni-goettingen.de

Matthias Schumann
Universität Göttingen Chair of Information Systems and E-Business, mschuma1@uni-goettingen.de

Follow this and additional works at: http://aisel.aisnet.org/ecis2007

Recommended Citation
Seidenfaden, Lutz; Ortelbach, B; and Schumann, Matthias, "A Peer-to-Peer Application System for Scholarly Communication" (2007). *ECIS 2007 Proceedings*. 180.
http://aisel.aisnet.org/ecis2007/180

This material is brought to you by the European Conference on Information Systems (ECIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A PEER-TO-PEER APPLICATION SYSTEM

FOR THE SCHOLARLY COMMUNICATION

Lutz Seidenfaden, Institute for Information Systems, Dep. II, University of Göttingen,
Platz der Göttinger Sieben 5, 37073 Göttingen, Germany, lseiden@uni-goettingen.de
Björn Ortelbach, Institute for Information Systems, Dep. II, University of Göttingen,
Platz der Göttinger Sieben 5, 37073 Göttingen, Germany, bortelb2@uni-goettingen.de
Matthias Schumann, Institute for Information Systems, Dep. II, University of Göttingen,
Platz der Göttinger Sieben 5, 37073 Göttingen, Germany, mschuma1@uni-goettingen.de

ABSTRACT

The number of scientific journals and thereby the number of published articles grew with an enormous rate in the last century (e.g. Price 1986; Henderson 2002). In the second half of the 20th century the system seemed to abut against its boundaries, because in relation to research budgets, library budgets did not grow fast enough to cover all the scientific output produced. Price increases well above the inflation rate set by commercial publishers that bundle disproportionately high market power – especially for journals in the Science-Technical-Medicine-Sector in the last thirty years – intensified the situation even further. This situation is known as the serial crisis. New Information and Communication Technology (ICT) driven publication models are established and seem to be a promising way out of the crisis because they reduce distribution costs significantly. Especially the open access (OA) movement that advocates free electronic access to scientific output is subject to a fierce public debate. In this paper we will detail problems associated with OA and suggest a Peer-to-Peer (P2P) system that supports electronic scholarly communication as a tool to address the economic problems mentioned above.

Keywords

Peer-to-peer, decentralized systems, scholarly communication, knowledge sharing, open access

1 INTRODUCTION

The number of scientific journals and thereby the number of published articles increased enormously in the last century (e.g. Price 1986; Henderson 2002). In the second half of the 20th century the system seemed to reach upon its limits. The two main reasons for the growth are the enormous investments in research projects during the cold war that produced a lot of scientific output (Goellner 2002) and the so called "publish-or-perish"-phenomenon: While in the beginning the publication of a scientific article was intrinsically motivated, nowadays the publication in relevant journals is essential for the career advancement of scientists. Both factors lead to an enormous growth in the number of scientific journals and the number of articles that are published. Proportionately to the research budgets, the budgets for libraries did not grow fast enough to cover all the scientific output that was produced. As a result, the area-wide adequate supply with scientific literature could not be maintained. Price increases well above the inflation rate set by commercial publishers in an almost monopolistic market -

especially for journals in the Science-Technical-Medicine-Sector in the last thirty years – intensified the situation even further. According to The Association of Research Libraries (ARL) there was an average price increase of 188% per scientific journal from 1986 to 2004 (Association of Research Libraries 2004; Bergstrom 2001; Orsdel/Born 2003). In the literature this situation is called *serial crisis* (e.g. Woodward/Pilling 1993). It is regarded as one of the main forces that leads to changes in the system of scholarly communication. There are three main reasons for this development:

Firstly, further specialisation and differentiation of scientific disciplines lead to smaller reader segments for scientific journals. The problem here is the cost structure of media products. The fixed first-copy-costs (i.e. the costs that are independent from the circulation) are rather high while the variable costs per copy are low. As a result, the fixed costs need to be covered by smaller numbers of copies that leads to higher prices per copy.

Secondly, the market structure of the scientific information market shows some properties that may lead to high journal prices:

a. Scientific information products cannot be substituted due to their specific bundling of content. Therefore, there are no alternative journals neither for authors nor for readers, allowing commercial publishers to set prices as a monopolist on the market.

b. In addition to the content monopoly of scientific journals on the market, the industry structure shows a high concentration rate caused by mergers and acquisitions (McCabe 2002). As a result, the market structure is in favour of the intermediary publisher who strengthens his position towards scientists and libraries.

c. Scientist as authors normally do not get financial remuneration from the publisher. However, costs occur for the publisher, e.g. for organizing the review process of submitted manuscripts that has to be compensated by selling the information products (e.g. journals) that are normally sold to two kinds of subscribers: a) personal subscriptions by scientists and b) libraries as institutional subscribers. Personal subscriptions are price sensitive i.e. scientist cancel their personal subscriptions when the subscription becomes too expensive while libraries cannot cancel their subscriptions due to their public task of information supply. Therefore, less personal subscriptions lead to higher prices because the publisher is forced to cover his fixed costs out of the remaining (institutional) subscribers.

Thirdly, there is an increase of fixed cost in the publishing industry caused by the increasing effort that is necessary to organize the review process for the increased contributions and to establish electronic publishing. Most publishers provide both an online and a print issue which leads to high costs. Furthermore, publishers need to account for archiving of digital issues and for the development of capable search tools (Haank et al. 2004).

In this context new ICT seems to bring new dynamics to the system e.g. the open access movement (see BOAI 2006; Bethesda 2006) that advocates free electronic access to scientific literature in the established system of scholarly communication. As the technological enabler they build the foundation for new publishing and business models which are necessary to handle the dissemination of scientific information more efficiently than the current system does and are likely to have impact on the traditional value chain of scholarly communication (Roosendaal 2001).

2 FUNDAMENTALS

2.1 Functions of scholarly communication

Scientific publications are the formal means for the communication between scientists. In general, a scientific publication fulfils four main functions (e.g. Kircz/Rosendaal 1996) that are viable for the scholarly communication:

(1) the *registration* function that relates research results to a particular scientist who claims priority for them,

(2) the *certification* that concerns the validation of research,

(3) the *awareness* function that leads to disclosure and search needs and

(4) the *archiving* function that concerns the storage and accessibility of research results.

Every kind of scholarly publishing does fulfil these (abstract) functions in some way.

2.2 Value chain of scholarly communication

In the existing system the value chain consists of several main steps. The process is started by the scientist who writes an article and submits it to a publisher of his choice. The publisher organizes the review process to ensure the scientific quality of the submitted papers. After the article is accepted for publication the publisher processes the manuscript and puts it in a publishable layout. The finished information product is then distributed via private or institutional subscriptions (i.e. libraries) to the scientists who use it for their research. Unlike other markets, a) the scientist usually work for free as authors and referees and b) the scientists are authors and users and thereby stand at the beginning and the end of the value chain.

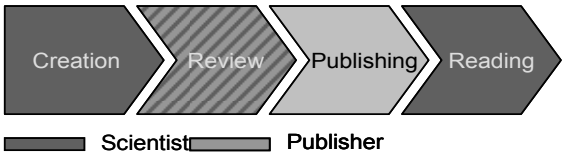


Figure 1: Traditional scientific value chain

2.3 Open access

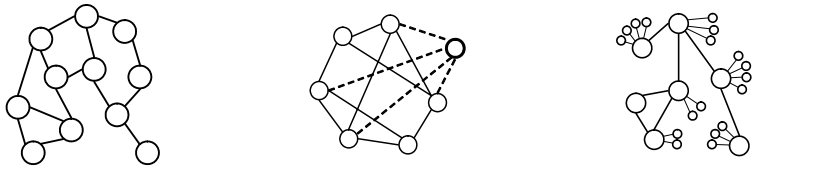
The situation on the market of scholarly communication as described in the introduction shows a misbalance of market power (i.e. an almost monopolistic market) that leads to increasing journal prices and thereby insufficient supply of scientific literature (e.g. McCabe 2000, 2002, 2004; Phillips/Phillips 2002; Doh-Shin/Menicucci 2003; Fay/Mackie-Mason 1999). A possible solution to the problem can be seen in the open access movement that pursues the aim to provide free electronic access to scientific publications. Open Access comes in two flavours (BOAI 2006; Guédon 2004; Harnad et al. 2004; Bolman 2003): (1) The "gold road", in which the authors publish their work by (for or not-for-profit) open access publishers that charge an author fee to cover publication costs but make the content freely available to users. (2) The "green road", in which the authors themselves archive an electronic copy of an article previously published in a traditional (i.e. subscription based) journal, in a repository (Beier/Velden 2004; Crow 2002) or on their own homepage freely accessible to the public. Although this paper focuses primarily on these two idealistic types of OA, today a variety of OA publishing models can be found that combine different types of financing models e.g. institutional and author fees. Other journals, e.g. the peer reviewed journal First Monday, function like Open Source Software Development and have established a collaborative publishing process in which scientists volunteer to fulfil tasks. These new possibilities of OA were only made possible by the introduction of electronic publishing because electronic publishing reduces the distribution costs significantly (Dryburgh 2003). Therefore, the underlying ICT can be seen as the enabler to the open access movement.

2.4 P2P Basics

In this section theoretical P2P network architectures are described and categorized. Miller (2001) characterized P2P networks by five key properties:

- The network facilitates real-time transmissions of data between the peers.
- Peers can function as a client and a server.
- The primary content of the network is provided by peers.
- The network gives control and autonomy to the peers.
- The network accommodates peers that are not continuously connected.

P2P networks are not structured the same way, in fact a lot of degrees of freedom exist while constructing such a network. A classification of existing systems (Hong 2001) should be followed that differentiates three classes (see figure 2):



Pure P2P architecture Brokered P2P architecture Hybrid P2P architecture

Figure 2: P2P-Architectures

In a pure there is no central unit for coordination purposes which leads to unreliable search behaviour and performance issues. To circumnavigate the issues regarding performance and scalability, the brokered architecture is coordinated by a central server. This ensures a faster discovery of peers and content. However, the server does not provide resources such as content or disk space; it only provides coordination mechanisms. Pure or brokered architectures do not mark alternative concepts. It is possible and often reasonable to combine both within a hybrid architecture in order to bring the advantages of complete decentralisation and a central unit together. Independent of its architecture, a P2P network can be organized in a structured or unstructured manner. Unstructured networks e.g. Gnutella, while not centrally planned in structure, grow according to a simple self-organizing process (Adamic et al. 2002). In contrast, in structured network protocols e.g. chord maintain a certain logical structure ("overlay") regardless of the size and the type of the (underlying) network (Stoica et al. 2001, Dabek et al. 2001) which improves the information retrieval.

3 MOTIVATION AND RELATED WORK

For a couple of reasons, the success of the open access movement can be questioned. The issues can be identified on different levels:

(I1) From a media economics perspective, the market power is simply shifted to other players (i.e. open access publishers), leading to increasing author or membership fees in the "gold road" model instead of subscription prices in the traditional model (Frank et al. 2004). Therefore it is questionable that open access really is cheaper than the traditional model since payment streams are simply redirected but the costs still occur (Bolman 2003; McCabe/Snyder 2004).

(I2) From a media management perspective, the business models of open access publication forms are of interest. The different variants of author-pays-models are not tested towards their sustainability and several voices question that the fees charged so far are sufficient to cover publication costs (Frank et al. 2004; Dryburgh 2003; Cozzarelli et al. 2004; Bolman 2003). The not-for-profit OA-Publisher Public Library of Science recently increased its author fees by 66%. In addition, the break even point i.e. the financial success of an author-pays journal heavily relies on the rejection rate i.e. scientific

quality, because processing cost for the publisher increase linear with the number of articles reviewed but rejected and therefore not published. Thereby, lowering the rejection rate allows charging lower author fees which means that an economic factor is intertwined with the scientific aim of the journal that may lead to lower scientific quality (Bolman 2003; McCabe/Snyder 2005). Furthermore, an author-pays model may put financial burdens on research institutions with two possible results: (1) only scientists belonging to wealthy institutions can publish and (2) institutions that generate high research output have to face disproportionate financial burdens (Cozzarelli et al. 2004). In addition, authors are generally not willing to pay high publication fees (Bolman 2003), which leads to a lack of acceptance of open access.

(I3) From an information systems perspective, the lack of standardisation of protocols and combined development efforts and therefore a variety of incompatible of application systems in the self-archiving model ("green road") is a problem for implementing open access because it reduces the awareness for newly published works by restricted searchability. Furthermore, archiving and retrieval of electronic copies cannot be ensured.

Research regarding p2p systems in connection with the scholarly communication has only received modest attention (Lagoze 2004). The decentralized application systems developed so far do not support sufficiently all functions of scholarly communication (see table 1). The LOCKSS-system focuses on the decentralised long-term archiving of digital objects (Reich/Rosenthal 2001). The open source project Edutella provides an RDF-metadata-scheme and an OAI-PMH-compliant search functionality and a replication service for scholarly literature (Nejdl/Wolf/Qu 2002). The project Annotea which was initiated by the W3C supports the annotation of digital documents in decentralized network structures (Kahan et al. 2001). Lagoze (2004) suggests the development of the OAI-PMH-protocol towards a peer-to-peer-framework. Lionshare supports the exchange of digital objects and the collaboration between students and faculty (Lionshare 2003).

	LOCKSS	Edutella	Annotea	Lionshare	Lagoze
Registration					
Archiving	X	X		X	
Awareness	X	X	X	X	X
Certification					

Table 1: P2P-systems and scholarly communication functions

The state of the art shows that research in this area is in the very beginning and still concentrates on the development of basic p2p-technologies (e.g. distributed hash tables) and frameworks. The development of services has only recently begun as the small number of available services (most of them being prototypes) illustrates. Besides the early stage, the most important lack of research seem to be the functional restrictions of all projects (see table 1). The projects mentioned above only focus on specific functions of the scholarly communication, but no project is able to support all four necessary functions. Some functions are not considered at all in the research e.g. registration and certification. As a result none of the projects is able to fully support the scholarly communication prozess. For this reason in this paper we present a p2p system that is able to support the scholary communication from registration of new material through to its archiving within the system.

4 REQUIREMENTS TOWARDS A SCIENTIFIC INFORMATION SYSTEM

In order to support the scholarly communication system, it is necessary to identify the need of the scientists for the design of a proper application system. In this section the requirements of scientists as authors and readers are identified[1].

4.1 Authors

In order to define the requirements authors have towards a scholarly communication system, it is necessary to shed light on two aspects: (a) the motives that lead authors to publication and (b) the criteria that play a role in the decision where to publish.

Studies regarding the motivation of authors to publish research results in scholarly communication media suggest that the three most important motives are (1) career advancement through publication, (2) the possibility to make a scientific contribution to its own field of research and (3) publication for posterity (see Tenopir/King 2001).

Although, the selection criteria of publication media vary from discipline to discipline, some general results can be extracted from relevant studies (Schauder 1994; Swan/Brown 1999; Swan/Brown 2003; Rowlands 2004; DFG 2005). The most important criterion is the reputation of the medium (4), followed by its ability to target the right audience (5) and its circulation (6). Furthermore, the coverage of the medium by abstracting and indexing (A&I) services (7) as well as the speed of publication (8) were mentioned as important selection criteria.

4.2 Readers

The requirements of scientists as users of scholarly literature can be identified by analysing their reading and searching behaviour. Studies have shown that use of scholarly journals is by far the most important source of information for scientists (see King/Tenopir 1999) and that the proportion read by electronic sources is continuously rising (Tenopir 2001; Rusch-Feja/Siebeky 1999). Together with the intensified usage of electronic sources some changes in behaviour can be observed: (8) scientists that use electronic sources read from a broader variety of journals, (9) they rely more on online search tools to find articles (browsing) and (10) read copies of individual articles rather than whole issues of journals. Furthermore, scientists seem to have no doubts about the quality of electronic sources in comparison with print sources (11) (Speier/Palmer/Wren/Hahn 1999). Not surprisingly, a study found that the most important factors for the usage of electronic sources are (12) easy access, convenience/desktop access and (13) searchability (Superjournal 2006).

To conclude this section the following requirements are derived from the author and user behaviour mentioned above:

Easy Access: The system needs to provide convenient access to its contents. This follows from (12).

Standardization: From (1), (6), (7), (9) and (13) follows that the system must support (metadata) standards (e.g. OAI-PMH) that allows search engines or A&I services to discover the publication and therewith increase the awareness i.e. the visibility of articles published.

Topic specific communities: The need to define communities of scientists with similar research interests in the system can be derived from (2), (5) and (10).

[1] The requirements vary between disciplines. Therefore, only the most important requirements derived from surveys across multiple disciplines are used here.

Quality control: Furthermore, the necessity to establish efficient certification mechanisms that ensure the scientific quality of the works published is a consequence of (4), (8) and (11).

Long term archiving: This can be derived from (3).

The requirements are used in the following section to determine the functionality of the system.

5 P2P – APPLICATION SYSTEM

5.1 Why P2P?

The reasons to propose a P2P system are detailed below. A P2P system seems to be appropriate to address the issues on the market for scholarly communication and fulfil the functions of scholarly communication in the following way:

(I1) A possible solution to the problem of market power concentration seems to be disintermediation i.e. distributing the functions of intermediaries over all parties involved rather than on one player. Distributing the functions of scientific communication to the users of the system will eliminate publishers or other intermediaries from the value chain. The client application that runs on the scientists' computer, will share its resources (e.g. papers) with other members of the scientific community, no publication fees will be charged for publication. This would not change very much as scientists already fulfil the roles of authors and reviewers without financial remuneration. (I2) Thereby, the question who pays for the intermediary (being it subscriptions or author fees) and therefore the question for a sustainable business model for it becomes redundant. (I3) On technical level, the problems regarding the heterogeneous forms of self archiving on the "green road" can be addressed by developing an application that is compatible to various standards (e.g. OAI-PMH, Dublin Core) and a stable archiving structure. This would lead to the development of complementary products (personalized search tools, linking services) that increase the usefulness of the p2p network.

With regards to the fundamental user requirement R1, it seems to be advantageous to combine the publishing and searching in one client software i.e. providing single point of access to the system rather than having different tools for publishing and searching electronic scholarly content.

For these reasons (and to fulfil the other user requirements aforementioned), a p2p application system will be suggested that allows easy publishing and self archiving for authors in a standardised application system (*registration*) and ensures the *awareness* by obeying common meta-data standards for the system wide search functionality. By saving redundant copies of works on a large number of clients, the *archiving* can be fulfilled by p2p applications (Reich/Rosenthal 2001; Gehrke/ Seidenfaden/Baule 2005). *Certification* mechanisms need to be integrated in a decentralised system in order to ensure the quality of the content. This can be done by establishing groups of peers that share the same interest (or research topic and therefore have the knowledge necessary for review). Furthermore, functionality for annotations (i.e. open peer review) need to be provided. Thereby, no economic factor such as author fees will play a role during the review process.

5.2 Prerequisites

Scholarly communication functions were introduced earlier. In order to fulfil these functions in a p2p system we assume the following prerequisites to be satisfied: every peer has a pair of keys (i.e. a private and a public key) for encrypting, signing and checking the integrity of the information exchanged and every peer posses a certificate that insures its identity and holds the public key of a peer. Furthermore, the certificate contains the roles of the user in order to control access. Both, the key pair and the certificate need to be issued by a trustworthy third party.

5.3 Layer model

In this section the layer model that was developed for the realization of the prototype is explained. Within the layers, several modules encapsulate the functionality that is derived from the users' requirements. We identified three different layers that fulfil specific types of services (see Figure 3).

The *Basic Services* layer comprises services that are supported by virtually any other p2p network (e.g. file sharing networks). The **communication service** fulfils communication task and is already in place by using the internets TCP/IP protocol stack. This layer needs no further description. On top of it lies the **security service** that adds secure communication mechanisms (e.g. secure socket layer, SSL) and fulfils encryption and decryption functions for other layers. Furthermore, it handles operations regarding the certificates and public keys of peers. The two services mentioned above are essential for any p2p system and therefore do not fulfil specific requirements for scholarly communication. For simplicity they are not detailed below as this paper focuses on the layers that are specific to scholarly communication.

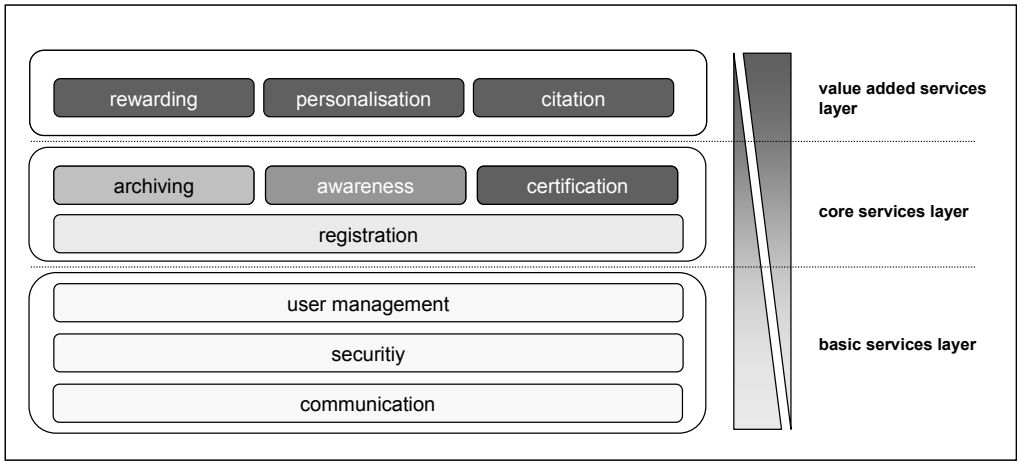


Figure 3: Layer model of the proposed p2p system

Services for the management of users and groups are provided by the **user management service**. In our context it manages the different communities of scientists. The community concept is essential as one of the scientists' requirements is to address the right kind of readership (R3). A community bundles readers with same research focus. Community information is stored on superpeers to ensure its availability. Furthermore, this layer controls the different levels of membership:

User/Participant: this is the most restrictive level. No registration is necessary for this membership level. Neither a key pair nor a certificate will be issued because user level only allows downloading of content. Submission of content or review of papers is not permitted. Users do not belong to any community.

Author: This level includes the user level. A registration is necessary and therefore key pair and certificate are issued by the trust centre. An author is entitled to submit papers under his own name to communities he belongs to.

Referee: Includes the two levels mentioned before. Referees obtain a key pair and a certificate during registration since these are necessary to fulfil their role. This is mainly the review of newly submitted papers in the communities in which the referee is a member.

The different levels of membership are necessary (1) to avoid the over flooding of the system with non-scientific material, (2) to establish a functioning review process that ensures the scientific quality

of content and (3) to guarantee the integrity of the content exchanged. For the coordination, a peer provides functions for the administration of communities and users (e.g. create, register, delete).

Services in the *Core Services* layer provide functionality necessary to fulfil the four basic functions of the scholarly communication.

Registration is fulfilled on a basic level e.g. when content is put in the shared folder on a users machine. This is not suffiecient in the scholarly communication, the registration requires more functionality since the moment when a research result becomes publicly available determines which researcher can claim priority for it. Therefore, in the registration layer a timestamp need to be added to a document and it needs to be signed using the authors' private key. Furthermore, if necessary, the review process is to be started (R2).

On top of the registration layer the **archiving service** handles the decentralized archiving of submitted documents (R5). Therefore, an efficient distribution and retrieval mechanism for the documents stored is to be implemented here. We will not deal with issues of long-term archiving since this is not the focus of the paper (see e.g. Reich/Rosenthal 2001; Gehrke/Seidenfaden/Baule 2005). For simplicity we will not develop a concept for this layer but used (i.e. define appropriate interfaces for it) the approach described in (Gehrke/Seidenfaden/Baule 2005) as it shows an efficient way of archiving information in a p2p system.

So far we have described and implemented functionality that is somehow fulfilled by other p2p networks as well (e.g. file sharing networks). We will now focus on layers that add special functionality for the scholarly communication i.e. value added services.

The first special functionality is added by the **awareness service** that provides mechanisms for the notification of the relevant scientific community, i.e. a list of the most recently submitted papers is displayed in the client software and users that have subscribed to the email notification service are notified via email. Furthermore, besides rudimentary search mechanisms of the system the layer is able to allow searching on peers from external search engines (e.g. Google Scholar: http://scholar.google.com; Scirus: http://www.scirus.com) by providing OAI-compatible interfaces.

The **certification service** supports the review process, whose functionality in contrast to the services aforementioned cannot be realized solely by technical means because it involves human knowledge and judgement (R4). After a paper is submitted by an author, the layer forwards it to randomly chosen individuals that have previously registered as referees in the specific community. One individual of the referees is chosen to organize the review for that paper and to notify the author of its results.

The two layers described above are able to (electronically) fulfil the four main functions of scholarly communication. On top of them, it might be helpful to establish a *Value Added Services* layer that hosts services that may be helpful to users. Examples could include:

- rewarding services i.e. referees get some kind of gratification for reviewing papers,
- personalisation services i.e. search services that are personalised based on searches performed previously,
- citation services i.e. that show who is cited by whom in which paper or
- reputation mechanisms.

Because one can think of a variety of services that could add value to the system, none of them is described in detail. Instead, the underlying layers provide interfaces that can be used by value added services because all value added services use the functionalities provided by the core and basic layer.

5.4 Interaction of layers

To demonstrate how the different layers work together, the publication process beginning with the submission of a paper and ending with its review is detailed below. The process includes the following steps:

- An author submits a paper using the upload functionality of the client software and fills in the necessary metadata. Thereafter, the paper is registered using the registration service, which adds a timestamp to the metadata and signs it using the authors' private key from the security module.
- Once the paper is registered, the registration layer calls the appropriate function of the archiving service that saves the paper according to the process described in (Gehrke/Seidenfaden/Baule 2005).
- After the saving process is finished the paper can be found by searching the metadata provided in (1). For searching, the awareness service calls the search method.
- If the paper is to be reviewed, the registration service calls the appropriate method in the certification service which initiates the review process by notifying individuals that have qualified as referees in the community the paper was submitted to. Each review report will be digitally signed by the referee (with his private key) and added to the system (using the archiving module). The unique fingerprints of the review documents will be added to the metadata of the reviewed paper and thereby referenced for later usage.

6 THE PROTOTYPE

6.1 Implementation[2]

Upon the concept described above a prototype has been developed using the programming language Java to ensure platform independence.

The class diagram of the prototype is depicted in figure 4. Each of the layers described above is represented by a handler-class that provides the specific functionality of that layer. The underlying network structure for the proposed system is a hybrid network architecture (figure 2) that allows to combine the strength of both pure and brokered architectures. A structured overlay network for each layer has been realized with the JChord library. In general, the prototype needs low hardware resources, only the Java Runtime Environment 1.4x is required.

[2] The authors would like to thank Karl-Philipp Naegler for the technical implementation of the P2P prototype.

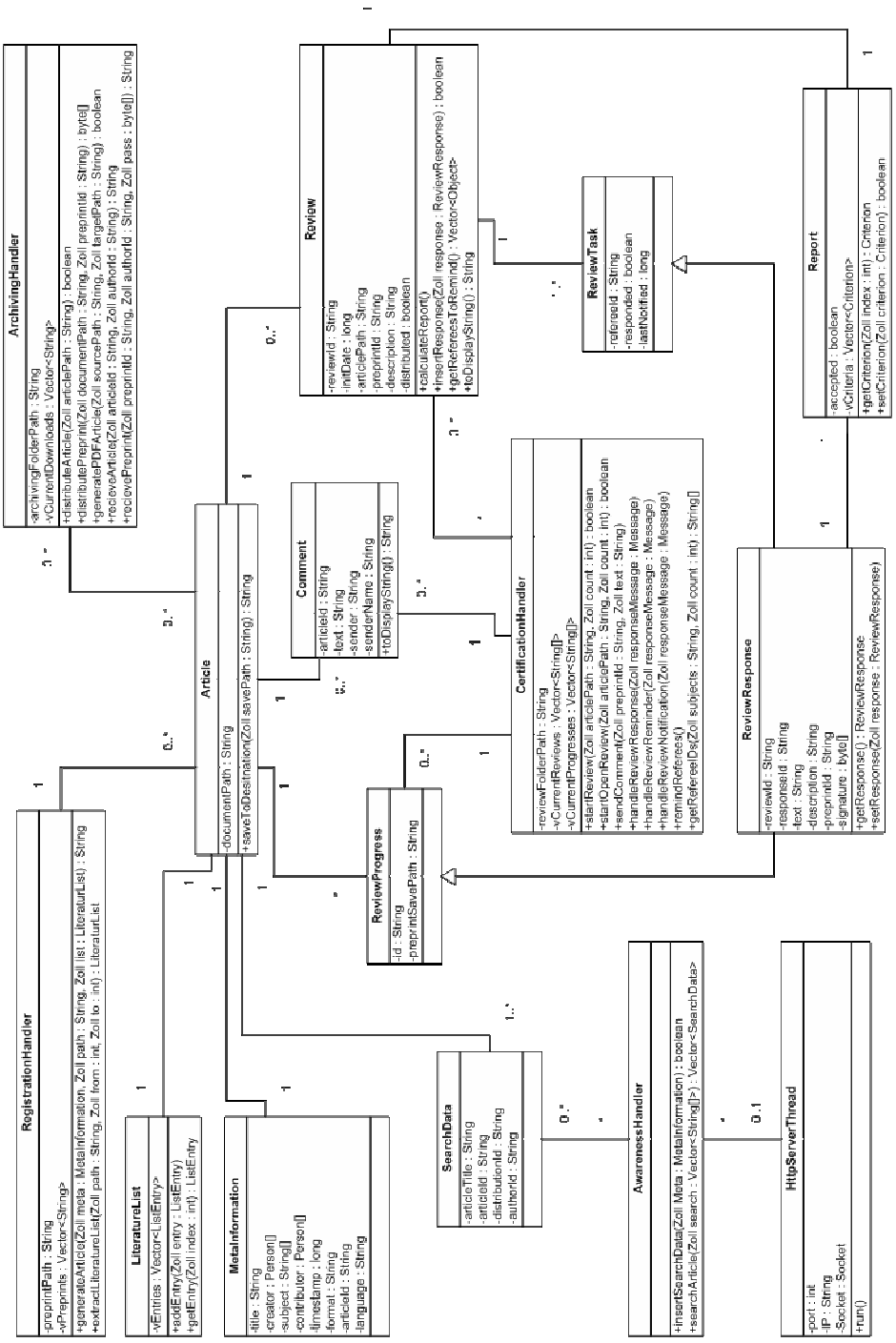


Figure 4: class diagramm of the prototype

6.2 Experiences

The prototype has been tested by the faculty at our institute in a small testbed (20 PCs) with satisfying results regarding stability and scalability. However, we identified that the overlay architecture may lead to performance bottlenecks as a lot of documents were assigned to i.e. stored on only one or two

peers. However, we assume that this happened due to the small testbed because only few peers were available for storage. Therefore we plan to modify the architecture and have the superpeers to perform some kind of load balancing when new papers are stored to ensure an equal distribution of papers. Furthermore, the system needs to be tested in a bigger environment. However, for such a test, time is required to convince scientists to publish there papers through the system. This could be done by scientific societies that recommend this way of publishing to its members. Thereby, a discipline wide adoption of the system could be realized, that could be used for further analysis regarding the scientists usage behaviour of the system. This would allow us to test the system behaviour under more realistic circumstances and to improve its functionality according to scientists requirements. In contrast to the possible bottleneck described above, the review process (i.e. assignment of documents to reviewers) and the automatic linkage between documents proved to work very well. Furthermore, we realized that some kind of central institution needs to be implemented in order to ensure the correct set-up of users and communities. This may be a role libraries could play in the future. Apart from proving the technical possibility, the prototype provides a good testbed for studies on reputation mechanisms for scientist (e.g. other than the impact factor of journals) which we work on with our colleagues from the social sciences. So far we have implemented a reputation system that relies not only on citations but also on an agreement rate which is given by the author to every article he has cited. In the actual system reputation is solely determined by citation numbers and therefore the quality of an article is not measured (e.g. articles of poor quality can obtain high citation numbers when people cite them as negative examples). In our system the agreement rate allows us to determine whether an article has been cited because of its good quality (high level of agreement) or e.g. because of its questionable research method (low level of agreement). The development of alternate "impact factors" is also important from an economic point of view since the traditional reputation system is seen as one reason for the serial crisis.

7 CONCLUSION

The paper developed a concept and a prototpye for the distribution of electronic scholarly articles over a highly decentralized network. It is shown that a P2P system is generally able to support electronic scholarly communication. In contrast to the existing system it distributes market power among its participants rather than bundling it on one player. Thereby, it exploits the fact that today scientists already work for free for the system (as authors and referees) and bundles both roles in its client software. As a result, only one client application is necessary to fulfill both roles, in contrast to the existing system which requires different tools for publishing and information retrieval. However, the system of scholarly communication is not going to be changed fundamentally (i.e. making publishers redundant in short or middle term) by this system, but it provides a useful tool that has the potential to advocate open access on the green road by providing easy and standardized self archiving by authors and easy access and convenient search ability by OAI-PMH-compatible search means and thereby changing the structure of scholarly communication in the long run. In this context, the system may provide a means to promote the "green road" to open access.

8 REFERENCES

Adamic, L. A., Lukose, R. M., Puniyani, A. R., Huberman, B. A. 2001: General methods of statistical physics - Search in power-law networks. Physical review / E, 64, 4, pp. 461351-461358.

Association of Research Libraries 2004: Association of Research Libraries: Monograph and Serial Expenditures in Research Libraries. http://www.arl.org/stats/arlstat/graphs/ 2004/monser04.pdf.

Beier/Velden 2004: Beier, G./Velden, T.: The eDoc-Server Project: building an institutional repository for the Max Planck Society. High Energy Physics Libraries Webzine, 9, URL: http://library.cern.ch/HEPLW/9/papers/4/.
Bergstrom 2001: Bergstrom, T. C.: Articles - Free Labor for Costly Journals?. The journal of economic perspectives, 15, 4, pp. 183-198.
Bethesda 2006: Bethesda Statement on Open Access Publishing, http://www.earlham.edu/~peters/fos/bethesda.htm.
BOAI 2006: The Budapest Open Access Initiative, http://www.soros.org/openaccess/.
Bolman 2003: Bolman, P.: Open Access: marginal or core phenomenon? Information Science and Use, 23, pp. 93-98.
Cozzarelli 2004: Cozzarelli, N. R.: EDITORIAL - An open access option for PNAS. National Academy of Sciences Washington, DC. In: Proceedings of the National Academy of Sciences of the United States of America, 101, 23, p.8509.
Crow 2002: Crow, R.: The case for institutional repositories: a SPARC position paper, Washington, DC.
Dabek et al. 2001: Dabek F., Brunskill, E., Kaashoek, F., Karger, D., Morris, R., Stoica, I., Balakrishnan, H.: Building Peer-to-Peer Systems With Chord, a Distributed Lookup Service, Proceedings of the 8th Workshop on Hot Topics in Operating Systems (HotOS-VIII).
DFG 2005: DFG: Publikationsstrategien im Wandel?, Weinheim.
Doh-Shin/Menicucci 2003: Doh-Shin, J., Menicucci, D.: Bundling Electronic Journals and Competition Among Publishers, Mimeo, Universitat Pompeu Fabra.
Dryburg 2003 : Dryburgh, A.: Open-Access Journals—Nice Idea, Shame About the Numbers? in: Learned Publishing, 16, 1, pp.15-16.
Fay/Mackie-Mason 1999: Fay, S., Mackie-Mason, J. K.: Competition Between Firms that Bundle Information Goods, University of Michigan working Paper.
Frank et al. 2004: Frank, M., Reich, M., Ra'nan, A.: A Not-for-Profit Publisher's Perspective on Open Access. Serials review, 30, 4, pp.281-287.
Gehrke/Seidenfaden/Baule 2005: Statistical Basics of a Reliable World-Wide Peer-to-Peer Memory, Proceedings of the Eleventh Americas Conference on Information Systems. Omaha, Nebraska.
Goellner 2002: The Impact of Library Budget Crisis on Scholarly Publishing. In: Abel, R. E., Newlin, L. W. (Eds.): Scholarly publishing: books, journals, publishers, and libraries in the twentieth century, New York, pp. 273-276.
Guédon 2004: Guédon, J.: The Green and Gold Roads to Open Access: The Case for Mixing and Matching, in: Serials review, 30 , 4, pp. 315-328.
Haank et al. 2004: Haank, D./Lex, A./Hilf, E./Mittler, E.: Thema: Open Science. in: Medienwirtschaft 1, 3, pp. 143-151.
Harnad et al. 2004: Harnad, S./Brody,. T./Vallières, F./Carr, L./Hitchcock,. S. G. Y./Oppenheim, C./ Stamerjohanns, H./Hilf, E. R.: The Access-Impact Problem and the Green and Gold Roads to Open Access. In: Serials review, 30 , 4, pp. 310-314.
Henderson 2002: Henderson, A.: The Growth of Printed Literature in the Twentieth Century. In Abel, R. E., Newlin, L. W. (Eds.): Scholarly publishing: books, journals, publishers, and libraries in the twentieth century, New York, pp.1-23.
Hong 2001: Hong, T.: Performance. In: Oran, A.: P2P - Harnessing the Power of Disruptive Technologies, Sebastopol, pp. 203-241.
Kahan et al. 2001: Kahan, J., Koivunen, M. R., Prud'Hommeaux E., Swick R. R.: Annotea: An Open RDF Infrastructure for Shared Web Annotations, WWW10, Hong Kong.
King, D. W., Tenopir, C. (1999): Using and reading scholary literature. In: Williams, M. E., (Eds.) Annual Review of Information Science and Technology. Medford, p.423-477.
Kircz/Roosendaal 1996: Kircz, J., Roosendaal, H. E.: Understanding and Shaping Scientific Information Transfer, Joint ICSU Press/UNESCO Expert Conference on ELECTRONIC PUBLISHING IN SCIENCE, Paris, 19.-23.2.1996.
Lagoze 2004: Lagoze, C.: Bridging the Past and Future: Scholarly Communication in the 21st Century, DLKC'04.

Lionshare 2003: o.V. 2003: Lionshare - Connecting and Extending Peer-to-Peer Networks. A Penn State Proposal to the Andrew W. Mellon Foundation, http://lionshare.its.psu.edu/main/info/docspresentation/lionshare_mellon_pdf.
McCabe 2000: McCabe, M. J.: The Impact of Publisher Mergers on Journal Prices: A Preliminary Report. ARL: A Bimonthly Newsletter of Research Library Issues and Actions, pp. 3-7.
McCabe 2002: McCabe, M. J.: Journal pricing and mergers: a portfolio approach. The American economic review, 92, 1, pp. 259-269.
McCabe 2004: McCabe, M. J.: Law Serials Pricing and Mergers: A Portfolio Approach. Contributions to Economic Analysis & Policy, 3, 1.
McCabe/Snyder 2004: McCabe, M. J., Snyder, C. M.: The Economics of Open-Access Journals, Mimeo, Georgia Institute of Technology.
McCabe/Snyder 2005: McCabe, M. J., Snyder, C. M.: Open Access and Academic Journal Quality. American Economic Review, 95, 2.
Miller 2001: Miller, M.: Discovering P2P, London, 2001.
Nejdl/Wolf/Qu 2002: Nejdl, W., Wolf, B., Qu, C.: EDUTELLA, A P2P Networking Structure based on RDF, WWW, Honolulu.
Orsdel/Born 2003: Orsdel, L. V./Born, K.: Big Chill on the Big Deal?. Library journal 128, 7, pp. 51-57.
Phililips/Phillips 2002: Phillips, O.R., Phillips, L .J.: The Market for Academic Journals. Applied Economics, 34, pp. 39-48.
Price 1986: Price, Derek J. de Solla: Little Science, Big Science and Beyond, Columbia University Press, New York.
Reich/Rosenthal 2001: Reich, V., Rosenthal, D. S. H.: LOCKSS: A Permanent Web Publishing and Access System, DLIB, 7.
Roosendaal 2001: Roosendaal, H. E.: Developments in scientific communication: Considerations on the value chain. Information Services & Use, 21, pp.13-32.
Rowlands et al. 2004: Rowlands, I., Nicholas, D., Huntington, P.: Scholarly Communication in the Digital Environment: What Do Authors Want? Learned Publishing, 17, no. 4, pp.261-273.
Rusch-Feja, D., Siebeky, U. (1999): Evaluation of Usage and Acceptance of Electronic Journals. D-Lib Magazine, 5, 10.
Schauder 1994: Schauder, D.: Electronic Publishing of Professional Articles: Attitudes of Academics and Implications for the Scholarly Communication Industry. American Society for Information Science: Journal of the American Society for Information Science, 45, 2, pp.73-100.
Speier, C., Palmer, J., Wren, D., Hahn, S. (1999): Faculty Perceptions of Electronic Journals as Scholarly Communication: A Question of Prestige and Legitimacy. American Society for Information Science: Journal of the American Society for Information Science, 50, no.6, pp.537-543.
Stoica et al. 2001: Stoica, I., Morris, R., Karger, D., Kaashoek, F., Balakrishnan: Chord: A scalable peer-to-peer lookup service for internet applications. In: Applications, Technologies, Architectures,and Protocols for Computer Communication Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications San Diego, p. 149 – 160.
SuperJournal 2006: Super Journal Baseline Studies, URL: http://www.superjournal.ac.uk/sj/baserept.htm.
Swan/Brown 1999: Swan, A., Brown, S.: What authors want: the ALPSP research study on the motivations and concerns of contributors to learned journals.
Swan/Brown 2003: Swan, A., Brown, S.: Authors and electronic publishing: what authors want from the new technology. Learned Publishing, 2003, 16, 1, pp. 28-33.
Tenopir, C., King, D. W. (2001): Lessons for the future of journals. Nature, London, 413, 6857, pp. 672-673.
Woodward/Pilling 1993: Woodward, H./Pilling, S.: The International Serials Industry: an overview. In: Woodward, H./Pilling, S. (Eds.): The International Serials Industry, Hampshire, pp. 1-22.