

A model of scientific communication of a global distributed information system

Bo-Christer Björk

Swedish School of Economics and Business Administration (HANKEN)

Bo-Christer Björk, "A model of scientific communication of a global distributed information system." *Proceedings of the IATUL Conferences*. Paper 3.

<http://docs.lib.purdue.edu/iatul/2007/papers/3>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

A model of scientific communication as a global distributed information system

Draft paper, 20.11.2006
submitted to Information Research (<http://informationr.net/ir/>)

Professor Bo-Christer Björk
Swedish School of Economics and Business Administration
P.O. Box 479, 00101 Helsinki, Finland,
Bo-Christer.Bjork@hanken.fi

Summary

A formal graphical model of the scientific communication process is presented in this paper. The purpose of the model is to act as a roadmap for policy discussions and research concerning the process. In comparison to earlier models found in the literature this model is more detailed, hierarchical and includes more modelling constructs (activities, inputs, outputs, controls, mechanisms). The modelling methodology used is IDEF0, a process modelling method, which previously has mainly been used for business process reengineering in the manufacturing industries.

The scope of the model is the whole communication value chain, from initial research to the assimilation of research results to improve every-day life. The model treats both informal and formal communication, as well as the publishing of data, but the major focus is on modelling the publishing and indexing of traditional peer reviewed journal articles, as well as the activities of readers to find out about them and access them. The new business models and parallel functions enabled by the Internet, such as open access journals and e-print repositories, are also in focus.

The current version of the model consists of 33 diagrams, with 113 different activities and over 200 different inputs, outputs, controls and mechanisms.

1. Introduction

Technology developments as a trigger for changes in scientific communication

The speed of progress in science has always been strongly dependent on how efficiently scientists can communicate their results to their peers and to lay persons willing to implement these results in new technology and practices. For centuries the communication chain was very slow, relying on tedious copying of scientific texts by hand. Communication was to a large extent local, taking place orally in the few universities then existing. The invention of the printing press was a major step forward and enabled the cost-effective reproduction of monographs, as well as the establishment of more systematic forms of communication, in the form of regularly appearing scholarly journals. Around the same time scientists started organising learned societies, the main aim of which was to facilitate the spread of knowledge.

During the 20th century science became recognized as the major driver for economic development and the number of scientists increased dramatically. In addition to journals and monographs conferences became an important form for communication, due to the increased possibilities for travel. During the latter half of the century IT had a profound impact on the scientific publishing process. First it enabled the setting up of data bases of bibliographic data, which greatly facilitated the search for relevant publications. Secondly word processing has meant increased efficiency in both the writing of manuscripts and in the handling of them during the printing process.

But the most dramatic effects on the overall process have occurred during the last fifteen years in information distribution and retrieval. It is perhaps no coincidence that scientist have been among the pioneers in using both email and the web. Science is by its nature both global and collaborative and the sorts of networking capabilities now offered are perfectly aligned with the open knowledge sharing goals of the academic community.

A large part of this communication process takes place as a distributed peer production process. Scientists usually require no monetary rewards for sharing their results, in contrast to producers of music or popular literature. What they are interested in, in addition to advancing science itself, is building up relationship with other scientists, or building a reputation which enables them to advance in their careers, get better grants etc. The more others read their publications and cite them the better. Unfortunately the legal, economic and behavioral infrastructure that underpins part of this communication process was shaped by the developments of the many decades preceding the Internet and has now become something of a straightjacket that hinders progress.

Traditionally scholarly journals are sold or licensed to libraries on a subscription basis by journal title or bundles of titles from the same publisher. Users affiliated with the subscribing libraries get access to the journals either in print, electronic, or both formats. As electronic publishing of journals has become common, new economic models have been proposed, notably open access publishing (a new model for publishing scholarly papers, where the full text is retrievable on the web for free by anybody). Open access can be achieved either in conjunction with the traditional economic model or in place of it¹. In conjunction with publishing in a traditional subscription based journal, publishers may give permission to authors to place an author copy of their final paper in an institutional repository, subject repository, or on the author's own web site. Open access economic

¹ For an excellent introduction to Open Access see Suber [2006].

models that replace traditional licensing models rely on either author payment to cover publication costs or publication costs being met by other means, such as advertising or subsidization.

The potential merits and negative effects of an increased use of the open access models for scientific publishing of peer reviewed journal articles is currently the subject of quite heated debates [Goodman and Foster 2004]. Strong and sometimes emotionally loaded arguments are put forward by proponents of Open Access; equally vociferous are defenders of the subscription/licensing model, the current dominant publishing economic model. The debate has also reached into the general media, in particular in connection with certain government initiatives to intervene in favour of open access parallel publishing of articles published in subscription based journals (The NIH proposals in the US, the UK parliamentary enquiry and the UK research councils plans).

Many rather superficial arguments for and against have been proposed. Some open access proponents state that it is morally wrong for a number of big publishers to use their quasi-monopoly situation to make excessive profits from content they get for free from academics and then selling the same content back to the academic sector. On the other hand, open access opponents are worried about the decreased possibility for scientific societies to use revenue from journals to support other activities and about their loss of membership if journals become free rather than bundled with membership. Some large universities whose faculty are prolific authors are concerned that their overall costs will actually increase if they move from a subscription model to an author-pays open access model.

Although a number of empirical studies of the economic effects of going electronic/and or open access have been made, it is difficult to compare the results of such studies since they are often measuring different aspects of the overall process. Thus there is a clear need for models which structure the overall scientific communication process, and can be used as a basis for comparing and integrating the results of different studies.

Earlier models of the scientific communication process

There are earlier models or studies of the scientific communication process, which have been presented in the scientific literature. Garvey and his colleagues at the John Hopkins University published a model in the early 1970's, based to some part on empirical observation of scientists in the domain of psychology [Tenopir and King 2000 pp. 88-89]. The Garvey-Griffith model was a good description of how the communication process functioned at a time when IT-support was still lacking. The modeling was done using verbal descriptions supplemented by a few graphical diagrams. A central aspect was to depict both formal and informal communication of research results and also the inclusion of the research into the body of scientific knowledge in its domain through citations in other publications, mention in review articles etc.

In the mid 1990's Hurd re-examined the status of the scientific communication process and took explicit account of the emerging effects of the Internet (i.e. e-mail and listservers and electronic publications) [Hurd 1996]. She has recently revisited the subject [Hurd 2004] taking into account recent developments such as self-publishing on the web and institutional repositories. Figure 1 illustrates the central aspects of the Garvey/Griffith and Hurd models. When this article was about to be finalised the author received a hint by a colleague about the article of Søndergaard at AI [2003] which proposes a revised version of a model of scholarly communication developed in the 1970s (the UNISIST model). The revised model takes into account the effects of the Internet and the authors also stress the need to start analyzing the differences between scientific domains.

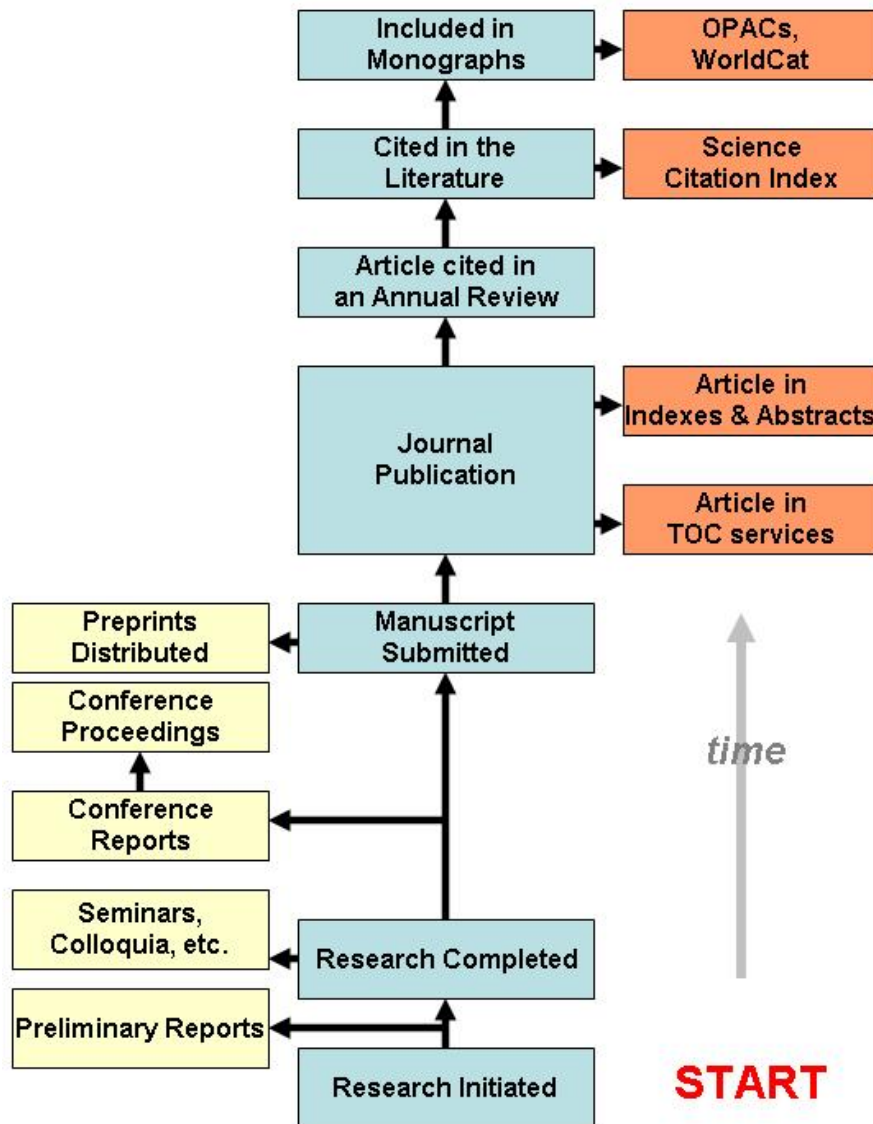


Figure 1. An illustration of the scientific communication process including facets of both the Garvey/Griffith model and Hurd's additions to it [Swisher 2005].

The book by Tenopir and King [2000]; *Towards electronic Journals – Realities for scientists, librarians and publishers*, contains a comprehensive discussion of the scientific publication process from a life-cycle perspective, and in particular synthesizes a large body of empirical evidence concerning the cost of different phases.

An interesting slightly different viewpoint is offered by Lewison [2006], who discusses what happens to a research publication after it has been read, in terms of how it is quoted in the popular press, influences clinical guidelines etc.

Scientific communication viewed as an information system

One interesting aspect of the scientific communication process is that it is a global interconnected information system. The academic discipline which studies the development and use of information systems in companies and organizations is usually called Information Systems Science. Another related but separate discipline is Information Studies or Sciences. Most researchers who have published studies of the scientific communication process and its aspects have come from the latter discipline. This study tries to combine perspectives from both these fields.

Information Systems Science typically studies IT-systems that commercial or non-commercial organizations build to support their activities. The systems can also span different organizations or be interfaced to customers (i.e. e-commerce systems). Typical for these systems is that they usually are purposefully planned and built in a top-down fashion. A good example is provided by so-called enterprise resource planning (ERP) systems which large companies build for themselves. In this respect the scientific communication system, and the IT-support it uses, is different, because it has grown in an organic way over decades, through the integration of tools produced by a large number of different players in a non-hierarchical and uncoordinated fashion. Nobody owns or has control of the scientific communication system, just like the Internet.

An important aspect of large integrated IT-systems in corporations, is that they fulfill multiple functions. Firstly they support transactions, such as registering and controlling sales in an e-commerce setting. Secondly they provide management with a basis for decision support by providing aggregate information based on often huge amounts of low level transactions [Turban and Aronson 1998]. The quarterly accounts of large companies is a good example. Also the scientific communication process fulfills two kinds of functions. The primary is of course to help in communicating interesting research results to interested recipients. The secondary is to provide decision support to research administrations to help in deciding about research grants, professorial appointments etc. In the case of scientific publishing the fulfillment of the second function has as a by-effect led to a situation which strongly favors the existing system, making this area less amenable to innovations and new business models than other areas of e-commerce.

There are several stages in the development of information systems, including requirements analysis, design, programming, implementation. In the early stages formal modeling methods, usually supported by graphical tools are typically used. Methods typically used include data flow models, semantic data models, object models [Sommerville 1995]. In his book on scientific publishing and knowledge sharing Hars [2003] includes example diagrams using both flowcharts and object models. A significant benefit of using some of these techniques is that they are supported by IT-based modeling tools, and that they can be used as basis for more detailed design and programming in an integrated fashion.

Despite the fact that the scientific communication process has not been designed but has evolved it might be useful to model it using some suitable formal process modelling technique. The technique which was chosen in this work is called IDEF0. The traditional uses of IDEF0 models has been in illuminating current and alternative processes in business process reengineering projects, typically focusing on the design and manufacture of industrial products like submarines or buildings. The choice of IDEF0 was partly a matter of convenience, the fact that the author was well familiar with the method from previous business process reengineering research [Karstila et al 2000].

2. Modelling methodology and scope of the model

The IDEF0 Modelling methodology

The main concepts of the IDEF0 method are the activity and the flow [NIST 1993]. Activities are shown as rectangles and their names start with verbs. Flows are represented by arrows and the names are nouns. A flow can be either an input, output, control or mechanism. An input represents something, which in an activity is consumed to produce an output. Typical inputs could be raw materials, energy, human labour, but also information when the purpose of the activity is to transform the information. Outputs can be reused as inputs to further activities, and feedback loops are possible. The carrying out of activities is guided by controls. Outputs which take the form of information can also be used as controls. Mechanisms, which point at activities from below, are persons, organisations, machines, software etc. which carry out the activities. The presentation of the IDEF0 diagrams is hierarchical in a way that individual activities contained in diagrams are broken down into further sub-activities in diagrams lower in the hierarchy.

As an example of using the IDEF0 method consider the preparation of a spaghetti meal (Figure 2). The top level context diagram contains only one activity, describing the overall activity. On the next level this activity is subdivided into three sub-activities: *prepare the sauce*, *cook the pasta* and *serve the dish*. These can inherit some of the flows of the mother diagram, but alternatively new flows can appear at this level. Thus the aggregate input *ingredients* from the context diagram is on the next sublevel disaggregated into five different ingredients (*minced meat* etc.).

IDEF0 and similar modelling techniques are frequently used in process re-engineering efforts to clarify the process and propose changes in it. Using a formalized tool helps in communicating about the process. For this modelling exercise a particular tool called BPwin has been used for making and editing the IDEF0 model. Compared to a simple drafting tool BPwin enhances the speed and consistency of the modelling work, especially for larger models and when changes are needed.

Scope of the SCLC-Model

The overall aim of the modelling work undertaken here was to understand the scientific communication process and how it has been affected by the Internet, in order to provide a basis for a cost and performance analysis of various alternative ways of organizing it. The model can also work as a roadmap for positioning various new initiatives, such as e-print repositories and harvesting tools, within the overall system. Although the current model also includes communication more in general, the emphasis is on publishing as traditional journal articles.

The model explicitly includes the activities of all the stakeholders in the overall process, including the activities of the:

- ***Researchers*** who perform the research, write the publications and act as reviewers
- ***Research funders*** who strongly influence the process
- ***Publishers*** who manage and carry out the actual publication process
- ***Libraries*** who help archiving and in providing access to the publications
- ***Bibliographic services*** which facilitate the identification and retrieval of publications
- ***Readers*** who search for, retrieve and read publications
- ***Practitioners*** who implement the research results directly or indirectly

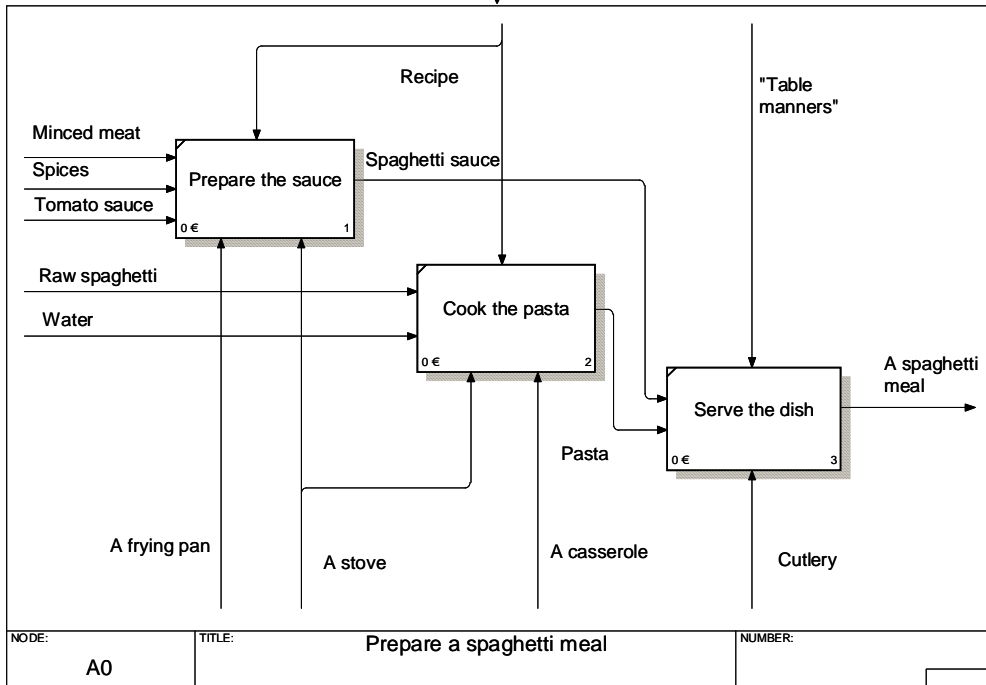
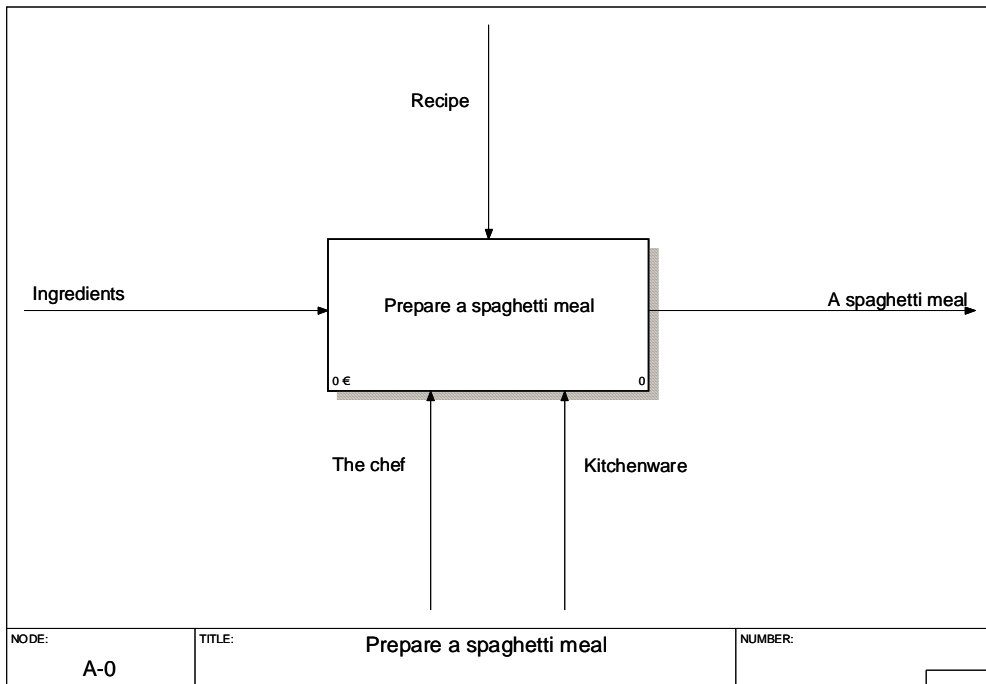


Figure 2. An illustration of the organisation of IDEF0 models. All models start with a context diagram, containing only one activity, which then is split into sub-activities, which in turn can be split into sub-activities etc.

The current version of the model has some limitations, which should be kept in mind. Its main emphasis is on the publication and dissemination of research results in the form of publications that in the end can be printed out and studied on paper (irrespective of whether the publications are distributed on paper or electronically). Thus forms of communication such as oral communication, unstructured use of email and multimedia, which all are essential parts of the scientific knowledge management process, as well as publishing of data and models, are only shown on a high level of abstraction in the model. Details could be added at a later stage, but would also add to the complexity of the model.

One important aspect of the process, which has been added to the current version, is the funding of the activities. Although parts of the overall process are carried out by commercially operating parties, almost all stages are predominantly funded by public finance via university budgets, research grant organisations etc. Also the model depicts publishing and value added services using both paper and electronic formats in an integrated way. Pure electronic or pure paper-based publishing could be described by subsets of the model. The same goes for free publishing on the web (“open access”), which resembles traditional publishing, but where certain activities such as negotiating, keeping track of and invoicing subscriptions can be almost entirely left out.

The model includes some activities, which would be typical for a scientific publisher publishing several journals, allowing for economies of scale. The activities of single-journal publishers could be described by a subset. The reason for including activities such as the general activities of a publisher is that these significantly influence the cost of individual journals in the form of the general overhead costs that publishers add to the subscription prices.

The central unit of observation in the model is the single publication (in particular the journal article), how it is written, edited, printed, distributed, archived, retrieved and read, and how eventually it may affect practice. The scope is thus the full life-cycle of the publication and the activities of reading it, which also is reflected in the name chosen for the model. This means in practice that most of the activities take place during 5-10 years after the initial writing of the manuscript, but in some cases there will be a demand to access the results after decades. Note also that compared to some of the earlier models the downstream life of manuscripts and electronic copies of publications is also modelled quite extensively.

Analysing the whole process in this way should help in highlighting how different actors provide added value to the end customers at each stage. It is therefore close in spirit to the concept of value-chain or value system analysis as defined by Porter [1985]. In the long run the customers (authors and readers) will decide on which business models prevail based on how much added value different intermediaries, such as OA journals provide them.

Overall organization of the model

The current version of the SCLC-model includes 33 separate diagrams, arranged in a hierarchy up to seven levels deep. There are typically 3-4 activity boxes on each diagram, although there are a couple of diagrams with more activities and some with only two. Official IDEFO guidelines recommend using up to six activities per diagram, but it was felt that models with fewer activities per diagram are easier to read and understand. There are altogether 113 activity boxes and around 250 labelled arrows. The overall hierarchical breakdown of the model is shown below in table 1 (detailed activity breakdowns have been omitted in some parts).

Table 1. Major hierarchical breakdown of the SCLC-model. Not all diagrams have been broken down here.

A-0 Context diagram

- A0 Do research, communicate and apply the results
 - A1 Fund R&D
 - A11 Evaluate prior research of applicants
 - A12 Evaluate research proposals
 - A13 Make funding decisions
 - A2 Perform the research
 - A21 Study existing scientific knowledge
 - A22 Collect data from existing repositories
 - A23 Do experiments and make observation
 - A24 Analyse and draw conclusions
 - A3 Communicate the results
 - A31 Communicate the results informally
 - A32 Communicate the results through publications
 - A321 Publish the results
 - A3211 Write Manuscript
 - A3212 Choose where to submit or negotiate publishing
 - A 3213 Produce publication
 - A32131 Publish as Monograph
 - A32132 Publish as Conference Paper
 - A32133 Publish as Scholarly Journal Article
 - A321331 Do publisher's general activities
 - A321332 Do journal specific activities
 - A321333 Process article
 - A3213331 Do peer review
 - A3213332 Negotiate copyright
 - A3213333 Pay article charges
 - A3213334 Do technical phases of publishing
 - A322 Facilitate dissemination and retrieval
 - A3221 Facilitate retrieval globally
 - A32211 Bundle publications from different sources into electronic services
 - A32212 Make manuscript or copy of publication openly available on the web
 - A32213 Integrate meta data into search service
 - A3222 Facilitate retrieval locally
 - A32221 Negotiate Subscriptions and Licenses
 - A32222 Make paper publication available inside organisation
 - A32223 Make electronic version available inside organisation
 - A3223 Preserve publication
 - A323 Study the Publication
 - A3231 Find out about the publication
 - A3232 Consider buying access to publication
 - A3233 Retrieve Publication
 - A32331 Retrieve paper publication
 - A32332 Retrieve electronic publication
 - A3234 Read And Process Publication
 - A324 Publish secondary account of the results
 - A33 Share the data
- A4 Apply the Knowledge
 - A41 Educate professionals
 - A42 Regulate industry and society
 - A43 Do industrial development
 - A44 Apply in practice

Only the major diagrams are shown in the table. Some diagrams are further broken down into separate activities. In the following model walk-through each diagram is explained separately. The diagrams are numbered using the standard IDEF0 numbering scheme, which helps keeping track of the hierarchical position of each diagram.

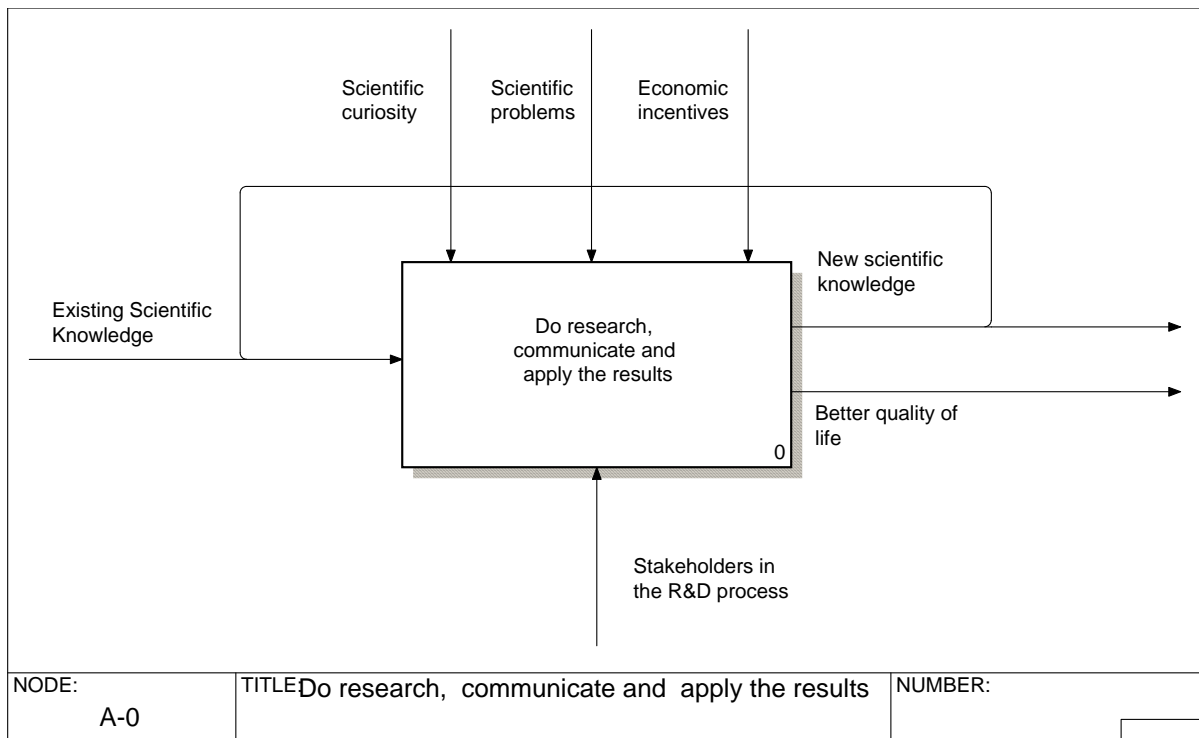
One aspect of IDEF0 modelling which readers might notice is the ambiguity concerning the use of information ICOM's as either controls or inputs. A good case is the review of manuscripts. The earlier version of the manuscript is used as an input being revised to produce a better version as output, and this is controlled by the reviews. There is no general rule for this and often either choice could be justified.

Note that this version of the model is the fourth and that the model has continuously evolved based on the feedback received. Due to the enlarged scope the model has been renamed *Scientific Communication Life Cycle Model*. The earliest published version was called the *Scientific Publishing Life-Cycle model* [Björk and Hedlund 2002]. In addition a conference paper [Björk and Hedlund 2003] and a journal article [Björk and Hedlund 2004] discussing parts of the model have been published. Version 3.0 of the model has been posted on the web pages of the OACS project [Björk 2005].

3. Model walk-through

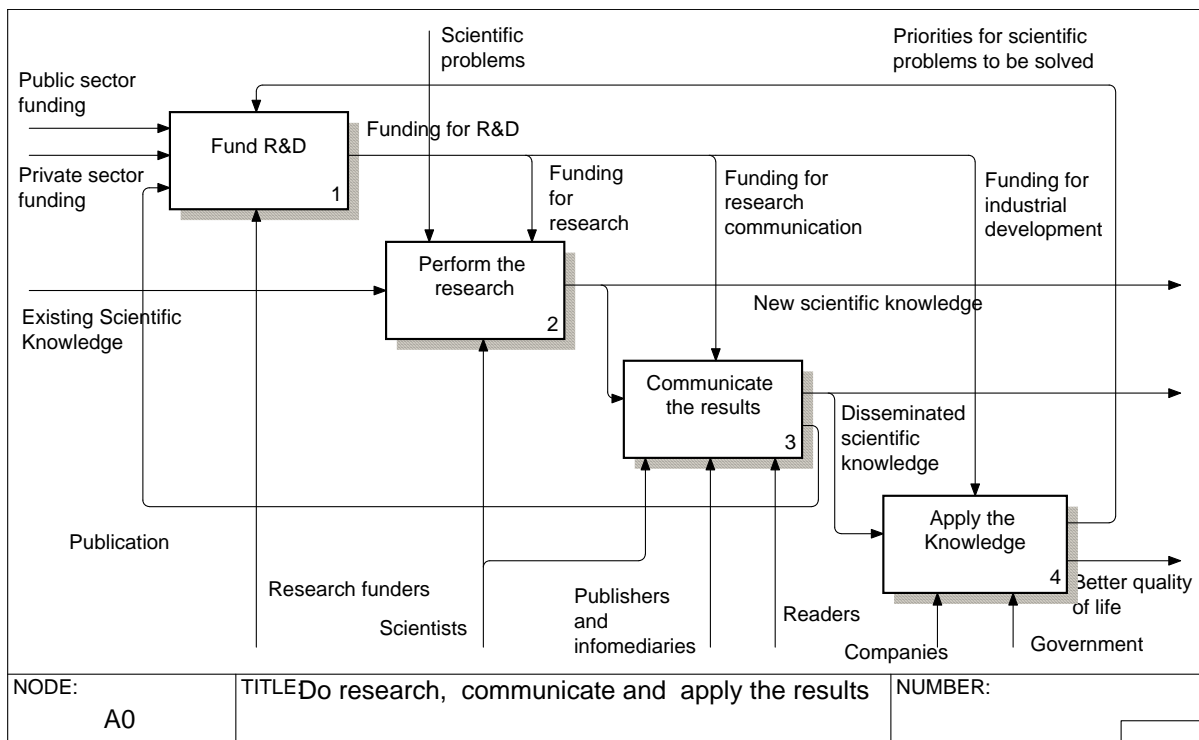
A-0 Do research, communicate and apply the results – context diagram

This is the so-called context diagram for depicting the overall model. The context diagram is traditionally the starting node of all IDEF0 models, and contains only one activity describing the overall process. The philosophy of this diagram is to show how science as a global knowledge creating and sharing system can help improve everyday life as well as create new scientific knowledge, which is fed back into the existing body of scientific knowledge. The main stakeholders in the process are collectively shown as a mechanism arrow coming into the activity box from below, and the main drivers controlling the behavior of the stakeholders are shown coming in from above (*scientific curiosity, economic incentives*). Also *scientific problems* to be addressed by the research and the whole accessible body of *existing scientific knowledge* are seen as controls. From an academic viewpoint the main output is *new scientific knowledge*. From the viewpoint of society that funds research the most important outcome is *better quality of life*.



A0 Do research, communicate and implement the results - breakdown

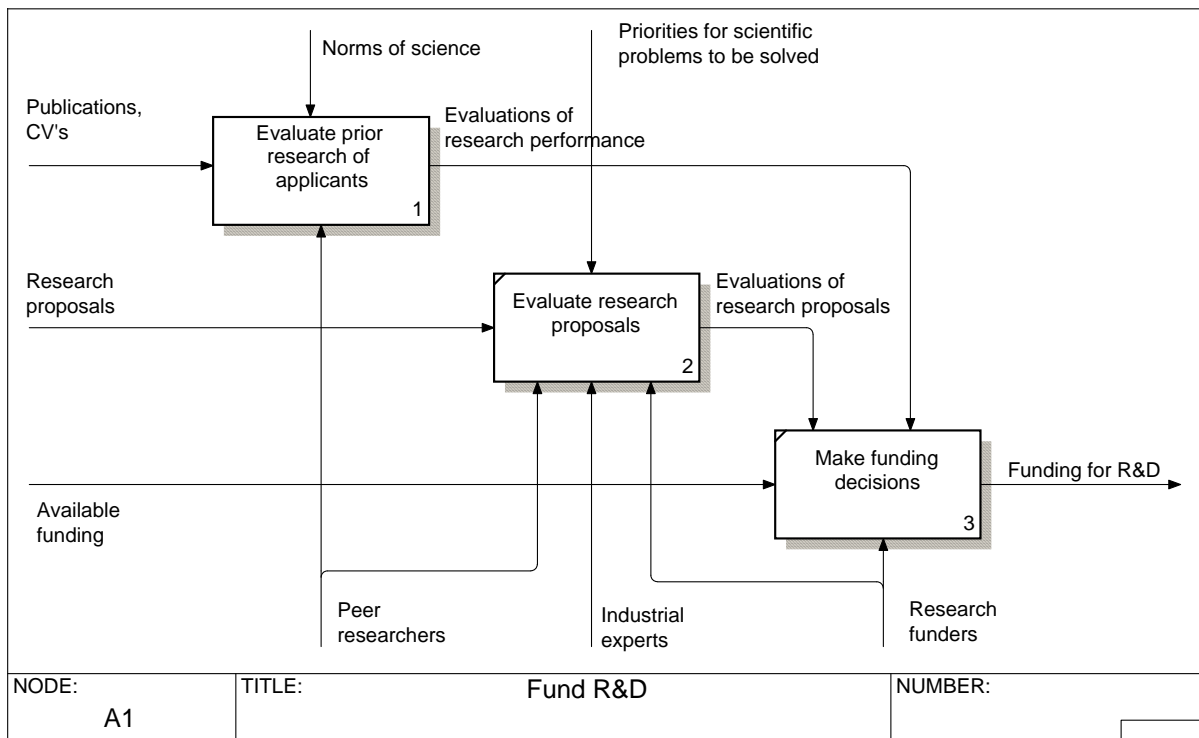
This diagram is crucial for understanding the life-cycle view adopted in this modelling effort. The whole life-cycle is seen as consisting of four separate stages. *Fund R&D* is included in the model as a separate activity. One reason for this is the importance research funders (understood in the widest sense including basic university funding) have in the shaping of the scientific communication chain, since they, through research contracts and university guidelines, potentially have the power to strongly influence a move towards Open Access. This can for instance be seen in the recent mandates and recommendations of the NIH, Wellcome Trust, Research Councils UK or Finland's Department of Education. *Perform the research* is the most resource demanding part of the system. *Communicate the results* is the most extensive part of the model. The end result of this activity is called *disseminated scientific knowledge*, reflecting the viewpoint that scientific results which have been published, but which are not read by the intended readers are rather useless. The downstream activity *apply the Knowledge* is important in order to achieve the improved quality of life which research funders mainly are looking for.



A1 Fund R&D

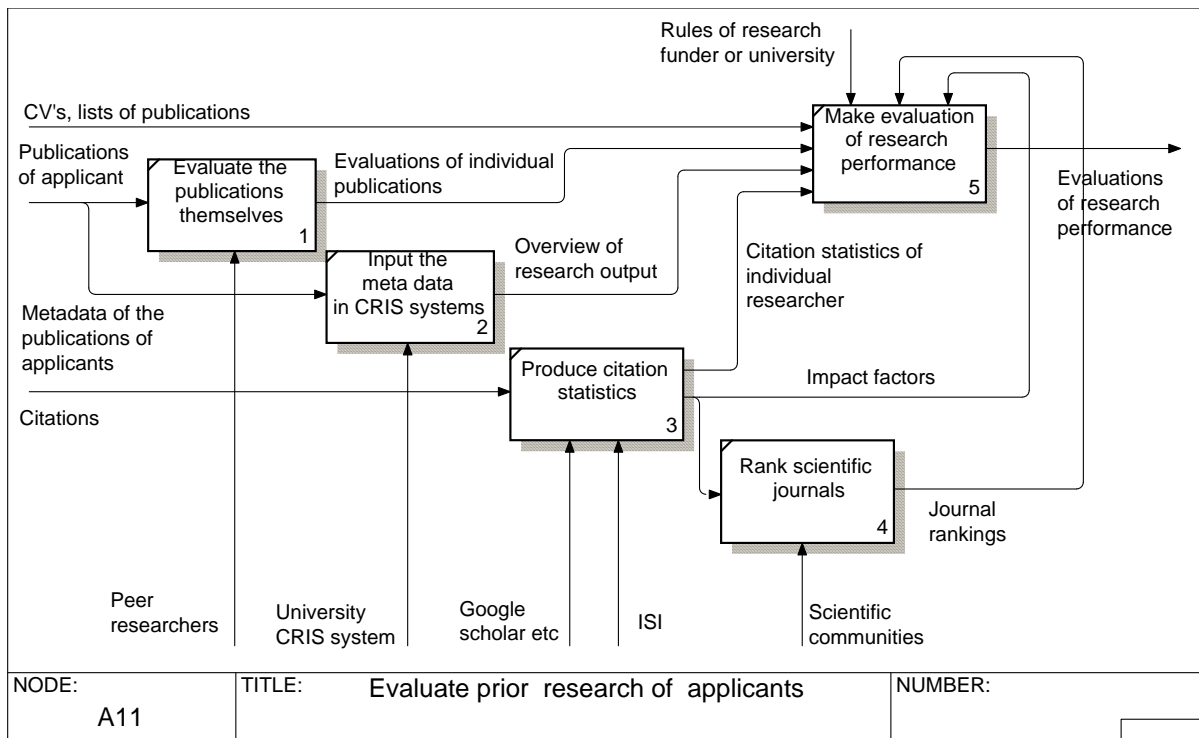
The global scientific communication system fulfils two functions; one is to communicate the knowledge as efficiently as possible. The other is to act as a decision support system for university administrations, granting agencies etc. This part of the model depicts the decision support functions of the overall system. It is important to include in the overall model, since certain aspects of this part, for instance the use of journal impact factors as a proxy for quality, constitute strong barriers for innovations in the communication parts of the overall system.

Funding decisions are here understood to include both decisions about basic university funding (i.e. UK research assessment exercise), decisions about individual research grants and academic appointments. The process can be seen as consisting of three separate parts, of which the evaluation of research proposals only applies to the decision-making about individual project applications.



A11 Evaluate prior research of the applicants

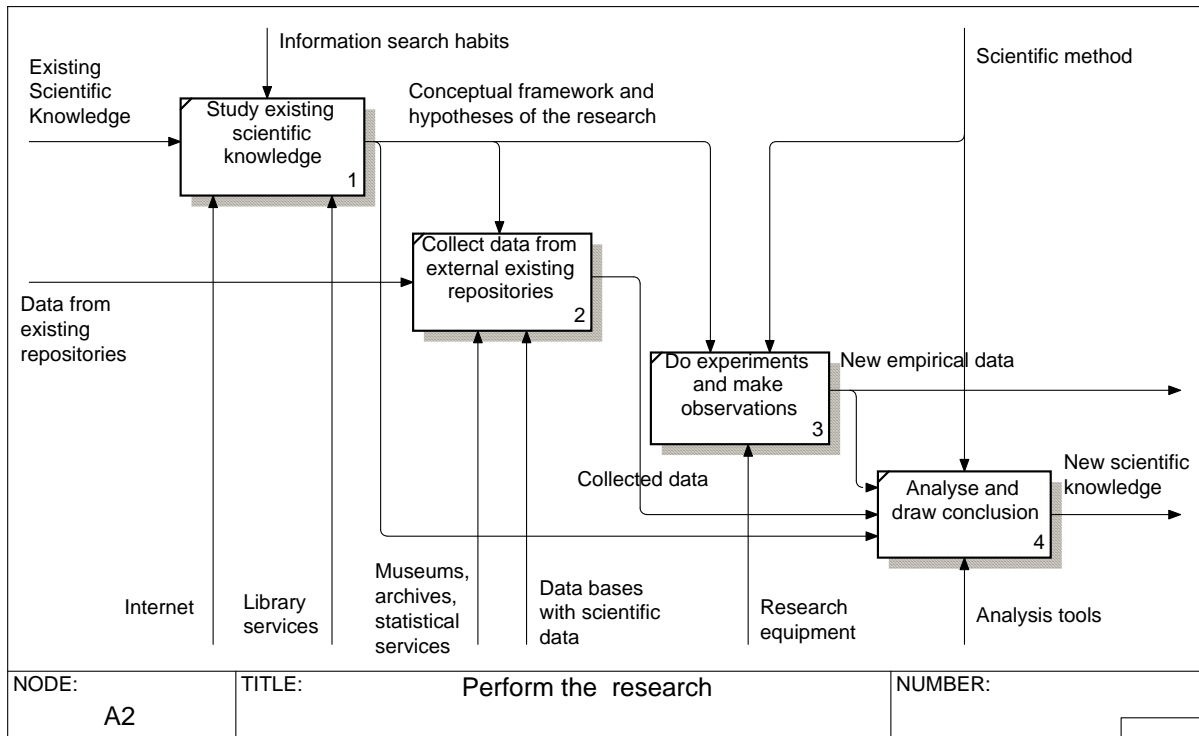
This diagram has been added since the earlier versions of the model and shows more in detail the part of the global information system that acts as a decision support system for university administrations and research funding organisations. At best the publications themselves are assessed by peers, but very often due to time and resource constraints the status of the journal where a researcher has published is used as a proxy for quality. Citation counts, using a system such as the ISI Web of Science, provide a reasonably objective measure of the impact of a particular publication, but only after a considerable time lag. The uploading of the metadata of a publication to a CRIS (Current Research Information Systems) is interesting from the information system development viewpoint, since the author is usually asked to do this, and since it would be very useful to integrate the CRIS system with the institutional repository of the same university [CRIS 2006].



A2 Perform the Research

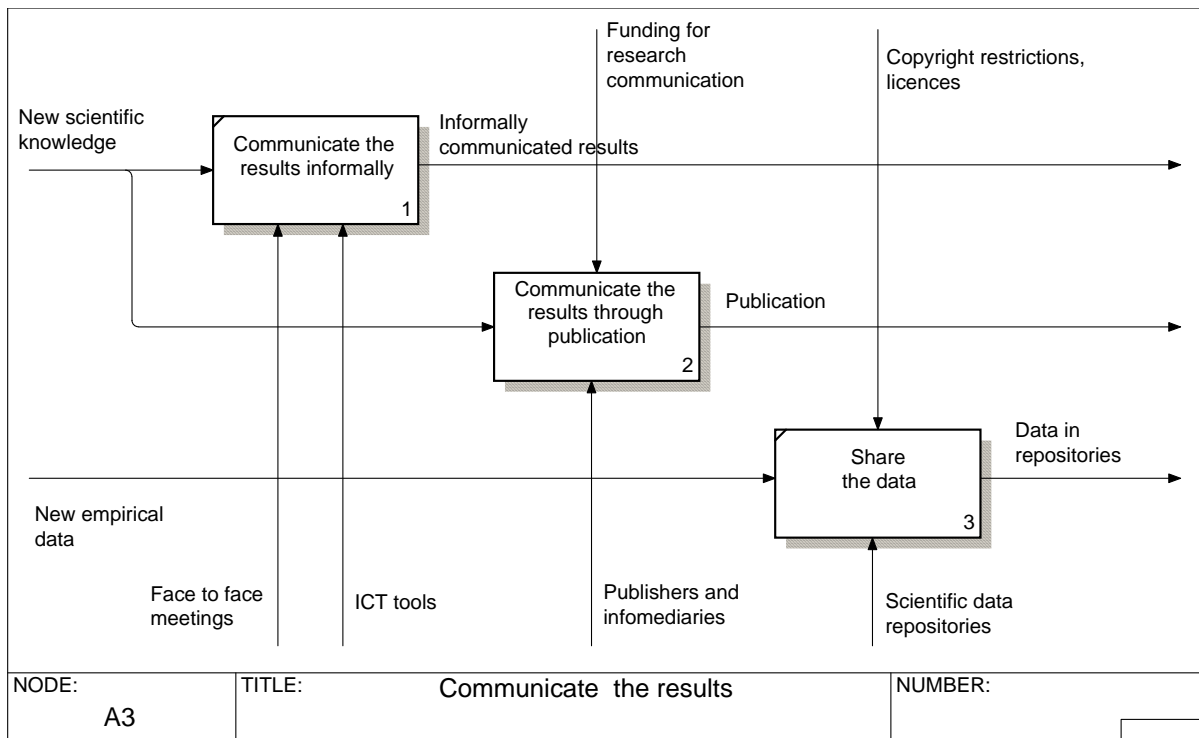
This diagram shows a highly simplified view of a typical research project. Note that one important feature of IDEF0 diagrams is that the consecutive activity boxes do not necessarily imply a strict order in time as in scheduling methods. Thus the activity *study existing scientific knowledge* can go on after the other two activities have started. The important thing is that it provides input to these. Clearly this is only one possible way of looking at the research process. The reason for choosing this view is that it clearly distinguishes the knowledge acquisition activity, which also is the topic of the later stages of the whole model. Here it is seen as providing input to the research that produces *new scientific knowledge*, whereas the later stages of the model show how other researchers utilize the results of this research for their own separate research projects.

Note that according to studies of reading habits of academics they spend around 2-3 months a year retrieving and reading scientific literature, in particular journal articles (King et al 2006). The efficiency of this activity, in terms of minimising the time and efforts spent on search and retrieval and in terms of being able to identify and getting access to the most relevant literature is the central issue of this modelling effort.



A3 Communicate the results

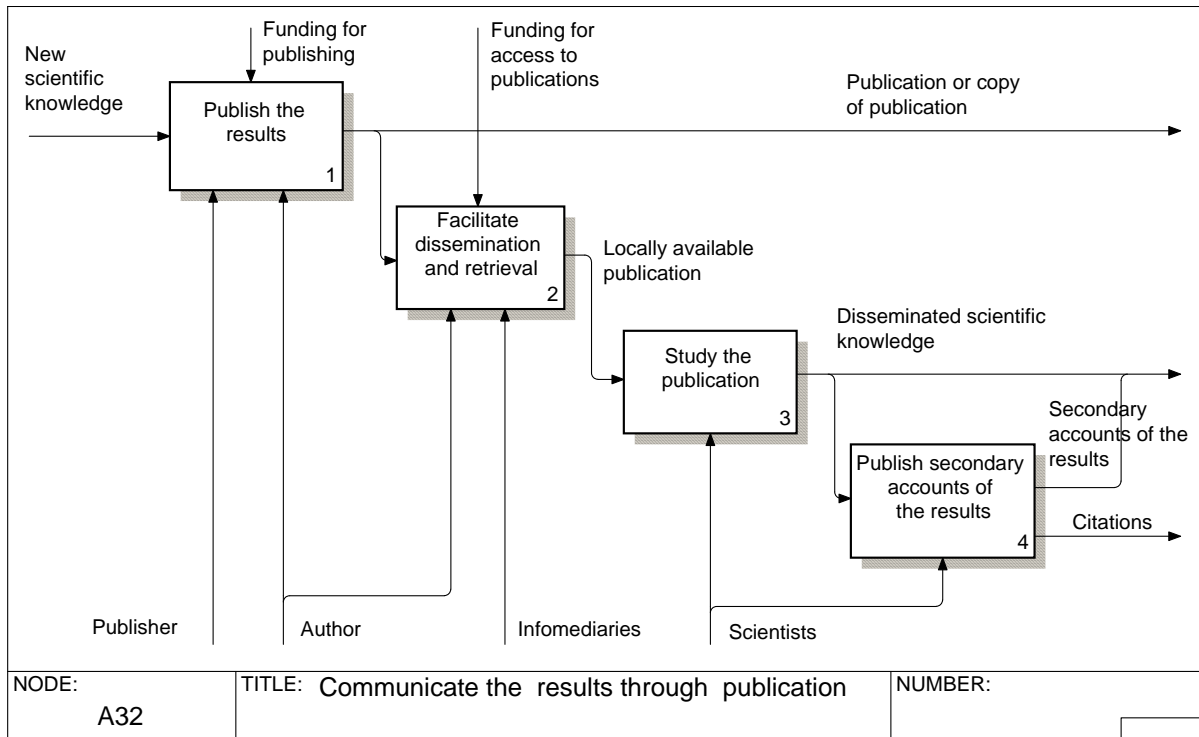
The scientific communication process is divided into an informal and a formal part. Informal communication is carried out in the form of oral presentations of all sorts (person-to-person meetings, conference presentations) as well as email messages, whereas formal communication (publishing) relies on written texts and on quality control by peers. A central difference is that in informal communication the producer of the information usually has full control or awareness of who the recipients are (emails to selected colleagues, presentations at seminars etc.). In science formal publishing usually has a highly specific meaning. It is often carried out in peer-reviewed outlets (working papers of universities are for instance usually not included) and in particular it is assumed to establish priority of new discoveries. In the model a more functional view is taken, where the pre-stages of formal publishing (working papers, posting manuscripts to preprint servers etc) are lumped together with traditional formal publishing. This diagram has been revised since the earlier versions of the models, and now takes into account the fact the scientists not only publish traditional-looking textual accounts (“papers”) but also data and models. Examples of the latter could include astronomical observation data, virtual reality models of historical artifacts, genome charts, computer code. Until now the Open Access movement has mainly concentrated on facilitating access to the textual account of research results, in particular the peer reviewed journal literature, but unrestricted access to research data and models is currently receiving increasing attention (OECD 2004).



A32 Communicate the results through publications

This part of the model is split into four different parts. The first part, *publish the results*, consists of the activities which contribute to the communication and initial publishing of the results, typically involving the researcher himself and a publisher. The *facilitate dissemination* activity describes activities carried out by a large number of organisations such as infomediaries, libraries, as well as IT infrastructure such as web search engines that facilitate for readers to find out about and retrieve publications of interest. This is in contrast to the informal communication where the author usually is directly communicating with the recipients of the information.

The third part of the communication chain is carried out by the recipients of the information in searching for, retrieving and studying publications. In any life-cycle studies this part is extremely important, and it has also been profoundly affected by the Internet. The last part consists of the activities of readers in communicating further particularly valuable research results, through citing them, incorporating them in university textbooks etc.

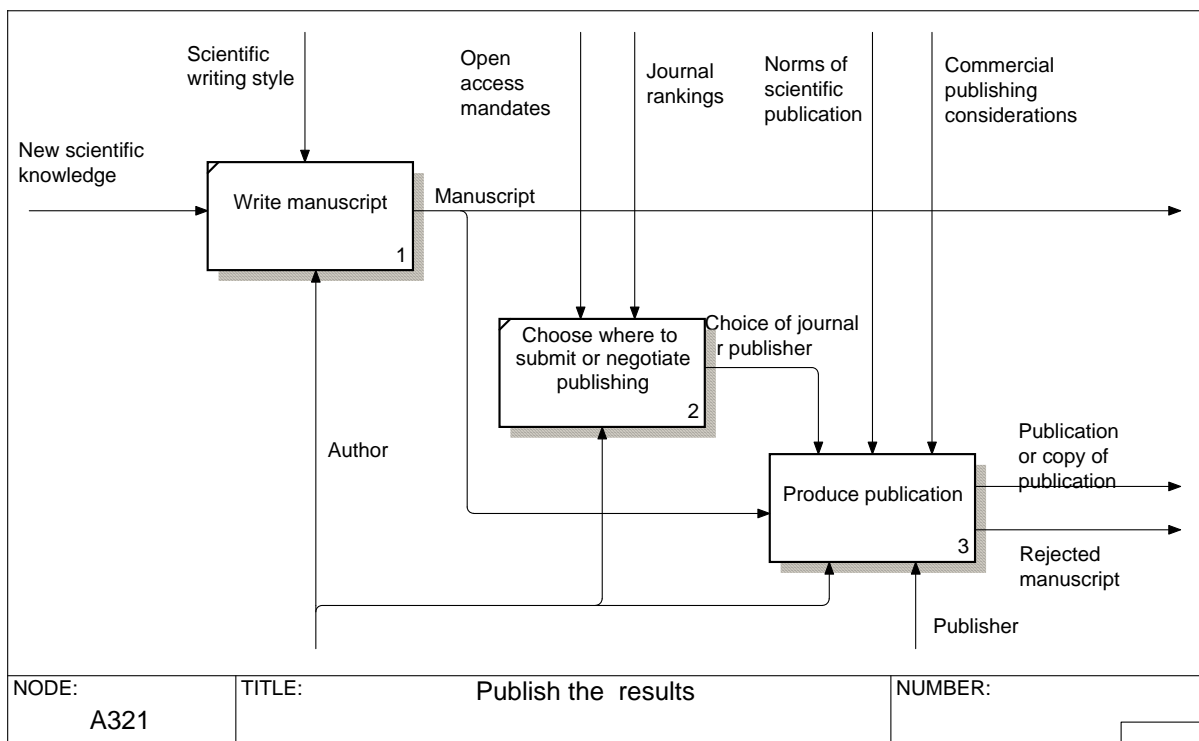


A321 Publish the results

Publishing consists of two separate activities, the writing of the manuscript, which the researcher carries out alone or in a small group, usually taking into account feedback from colleagues, and the more formal publishing process, in which outside persons, such as conference organisers, journal editors and staff etc. participate. Note that a manuscript intended for later publication can have a life of its own, since it can be uploaded to open access repositories on the web.

The writing of the manuscript is guided by a control called *scientific writing style*, which is a label used of a collection of formal guidelines and informal tradition taught to students via supervision of their work by more experienced academics. The production of a proper publication in turn is partly guided by the norms of the scientific community (which journals to publish in etc.), partly by commercial considerations (what journals have been established, decision to publish a particular book).

In order to highlight the importance of the choice of where to submit a journal paper, a separate activity for choosing where to submit or negotiate publishing has been inserted in the middle of the chain.

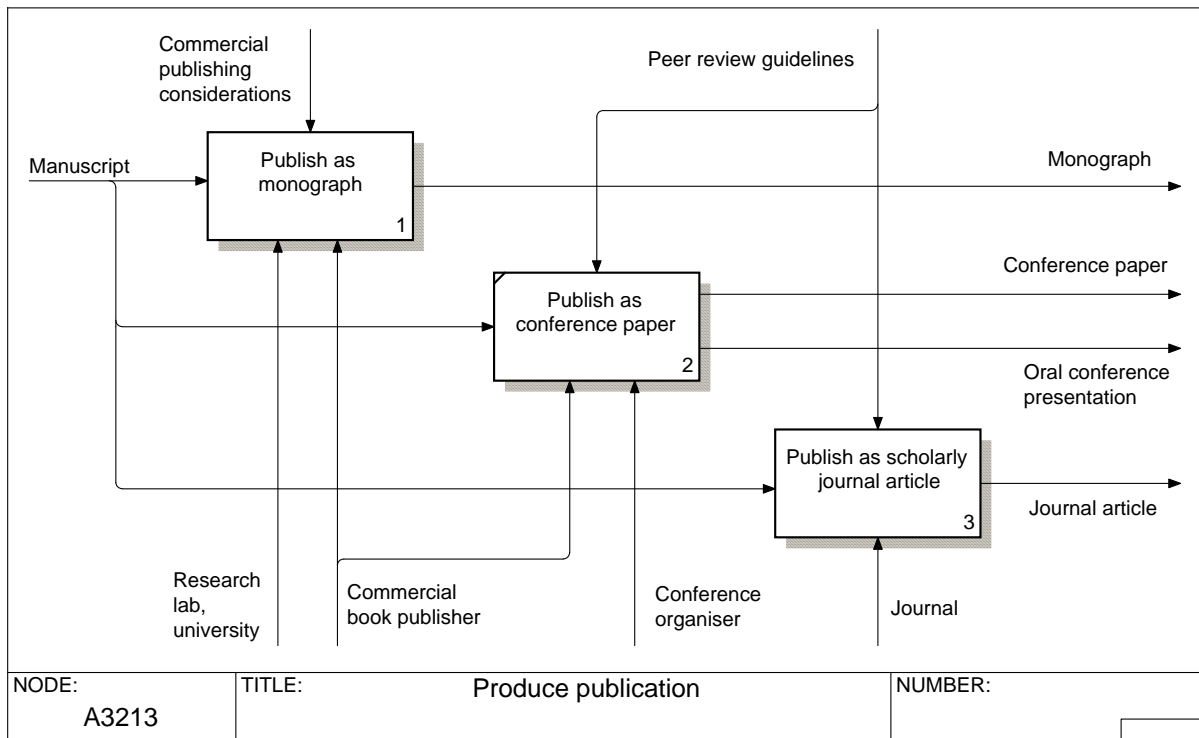


A3213 Produce Publication

At this stage the model is split into three parallel tracks which all take the generic *manuscript* as input. The term *monograph* is used to denote scientific publications which usually are published by the university of the researcher and which are not part of a scientific periodical journal or conference proceedings. Typical examples include working papers, research reports and Ph.D. theses. In some cases monographs might also be published by commercial book publishers, if there is a market demand. Monographs constitute a more important channel for communication in the arts and humanities than in STM (science, technology and medicine).

Conference papers are subjected to some sort of external review either for the abstract or the full paper, and are usually presented orally in addition to the printed version. Conference proceedings are published as one-off books, CD-ROMs or as annual series. Conference papers are also increasingly posted openly on the web by the organisers.

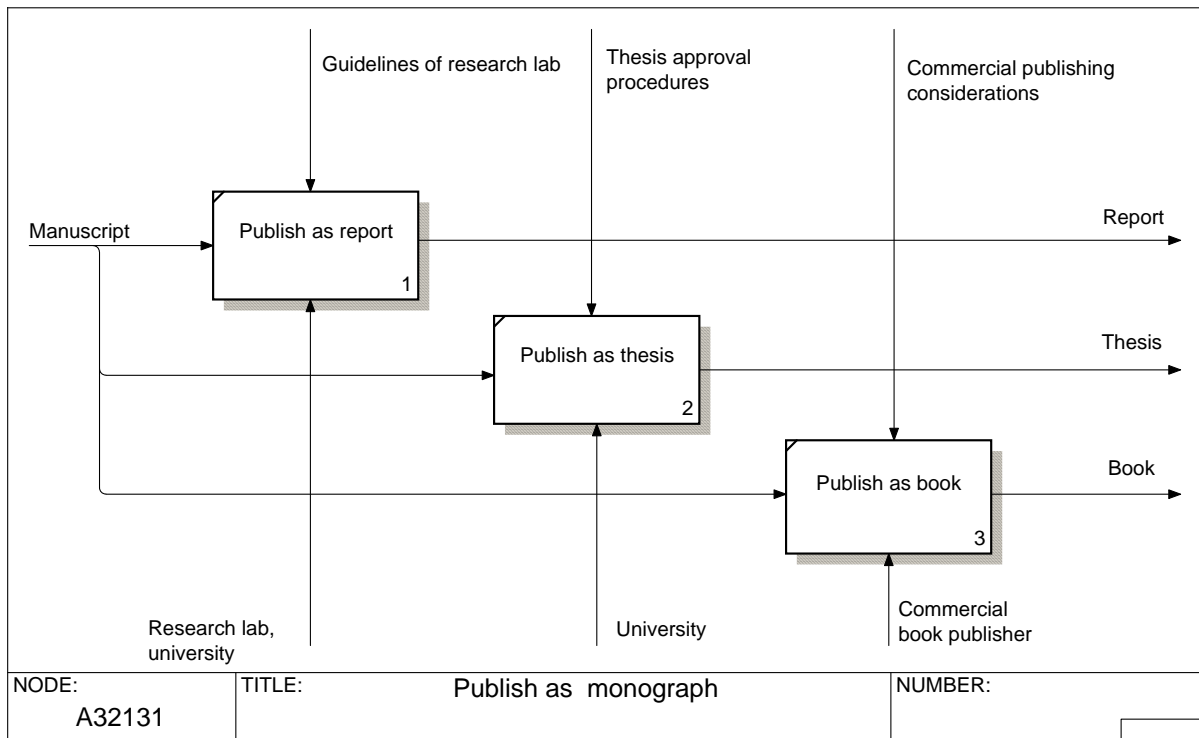
Articles in scientific periodicals are subjected to peer review. It is important to note that periodicals articles have a higher likelihood of being referenced in bibliographical services than the other types of publications. Also journals are usually available by subscription whereas the access to monographs and conference proceedings is predominantly acquired on a case-by-case basis.



A32131 Publish as Monograph

In the model three types of monographs are distinguished. *Reports* are typically produced by government or private research laboratories and some times by university departments. In the past such reports have mostly been sold individually but nowadays they are increasingly put freely available on line, since they offer a good marketing mechanism for the responsible research organisation. *Doctoral theses* are a rather special case since they usually undergo a rigorous quality control and are rather lengthy compared to reports, not to mention articles. Increasingly universities are putting them into their institutional repositories. *Books* printed by commercial or society publishers are a third important category. These can benefit from the marketing and indexing channels of the publisher, but the decision to publish is also strongly based on commercial considerations, in contrast to the two earlier ones.

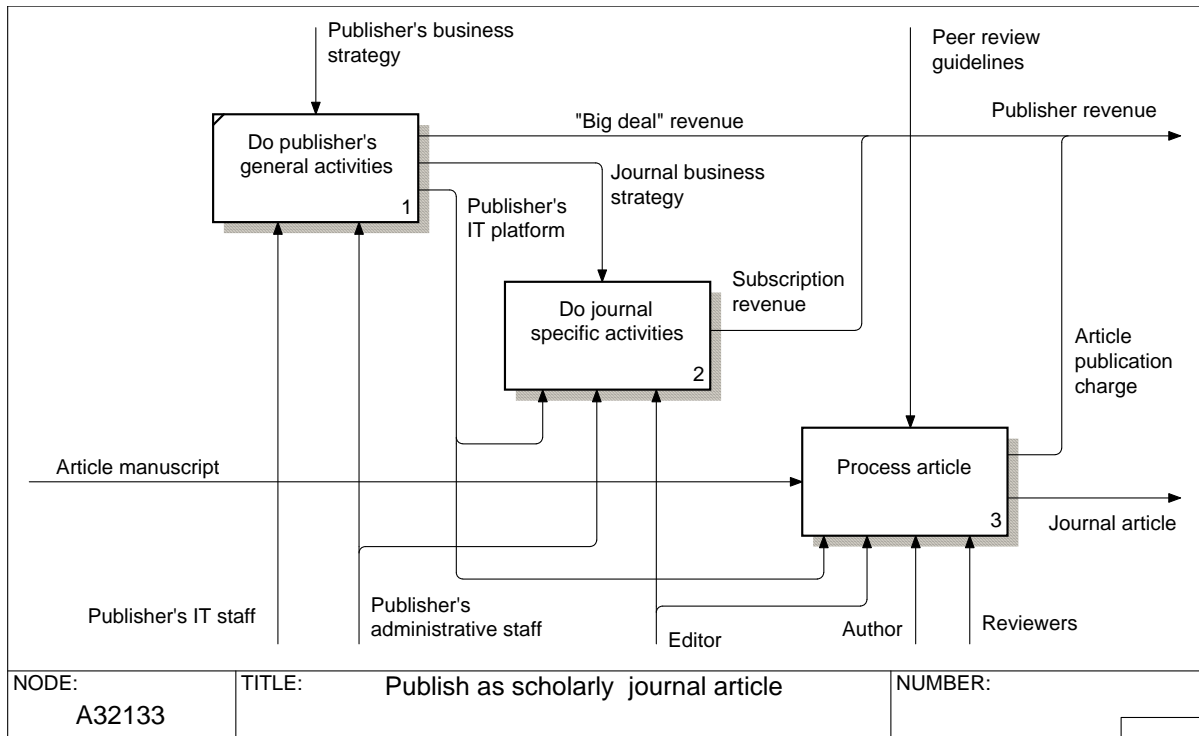
This diagram uses an abstraction mechanism frequently used in the model. The outputs *Report*, *Thesis* and *Book* are separate outputs but merge on the higher level diagram to form the super-type *Monograph* (A3213).



A32133 Publish as Scholarly Journal Article

The purpose of this diagram may at first sight be difficult to understand. The idea is to show all the activities which are carried out by the publishing organization, and thus have a cost associated with them. This is the reason for separating out activities such as *Do general publisher activities*, *do general journal activities*. Both of these demand resources, which cause overhead costs, which then are added on top of the basic variable costs caused by the processing of each individual article (as defined in the activity *process article*). For instance setting up and maintaining the IT-technical infrastructure for a portfolio of journals could be such an overhead causing item. In addition to the technical output the different revenue streams that journal publishing is generating are also shown.

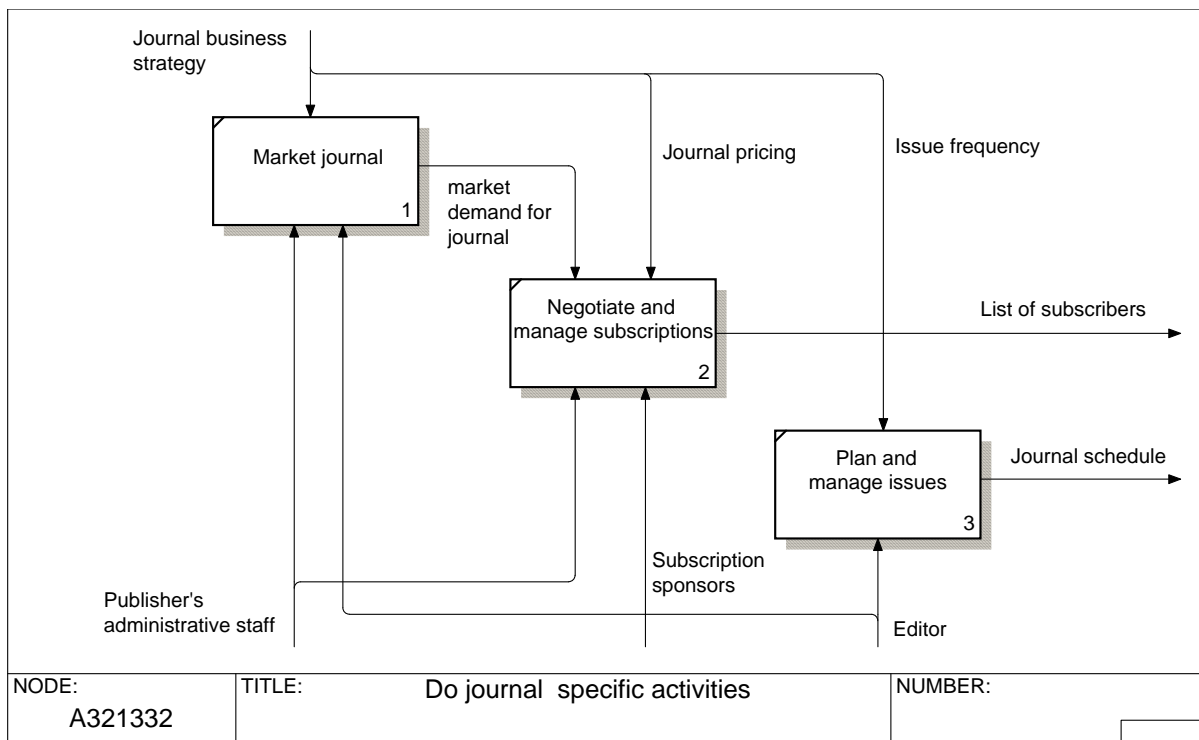
The main pipeline in the model is, however, the input arrow *article manuscript*, which directly enters the activity *process article*. This provides the root for a rather straightforward work-flow like model of what happens to a manuscript on its way to publishing.



A321332 Do Journal specific activities

The activities modelled on the level of the individual journal title, but not the individual submissions or articles, include general marketing both to subscribers and authors, negotiations for subscriptions and the management of them, as well as the planning of the journal (for instance special issues, editorial board). Interesting parameters which are set on this level are the pricing of subscriptions, which have a strong influence on the number of subscribers a journal reaches.

Many of the activities on this level are typically (but not always) handled by the employed staff of the publisher. Part of the activities are also handled by the academics involved in a journal, in particular by the editors.

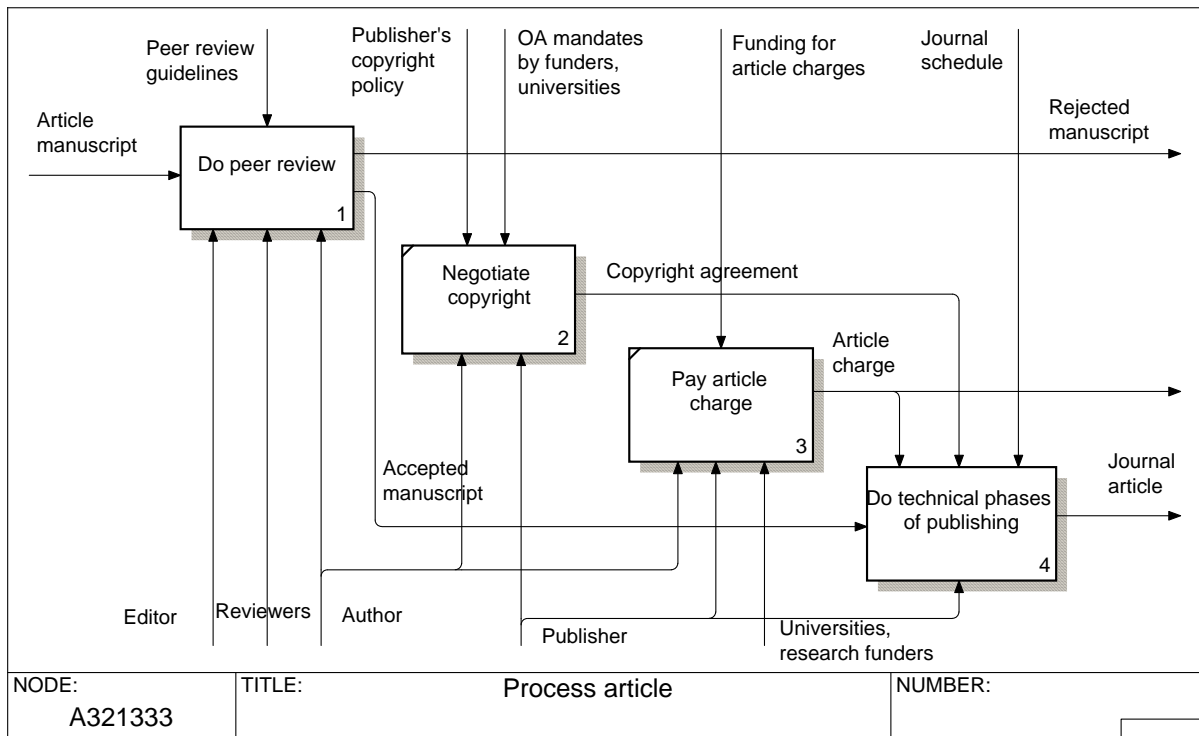


A321333 Process article

This diagram starts by the review activities carried out as a co-operation between the editor, the researcher and anonymous peer academics. This activity demands resources but is usually not a cost item of significance for the publishers since reviewers usually work for free (from society's viewpoint it, however, a significant cost item). The input consists of submitted manuscripts and output of accepted or rejected manuscripts. Note in particular that a significant proportion of manuscripts rejected by one journal are resubmitted as such or with minor revisions to other journals and thus continue adding cost to the overall total.

The last activity in the diagram includes the technical activities needed to publish the article. These are usually well described in the cost accounts of publishers. In between these two minor activities of particular interest from an OA perspective have been inserted, the signing of a copyright agreement and the payment of possible article charges.

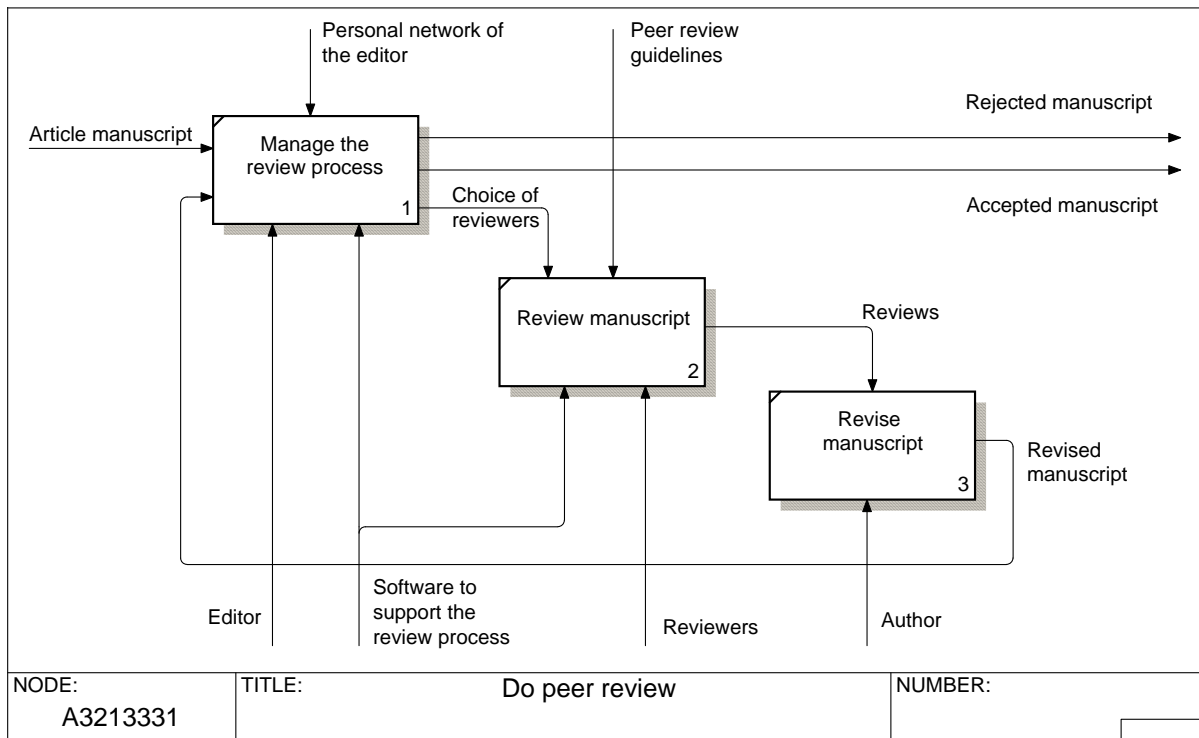
In relating cost to the activities in this figure it should be noted that the costs of the technical phases are highly correlated to the number of accepted papers, whereas the cost of the peer review stage is more related to the number of submissions. Thus journals which have a high rejection rate of say 90 % have a much higher overall cost per published article than say journals with a rejection rate of 50 %.



A3213331 Do peer review

This part of the model depicts the activities carried out during the traditional peer-review process. Note the use of a feedback mechanism, where the reviewers' comments become a control of the subsequent revise manuscript activity, and where the revised manuscript is fed back into the review process. There have been interesting experiments with new forms of quality control using the web, where for instance manuscripts are made openly available on a journal site and readers "vote" on which should be promoted to accepted papers [ETAI 2006]. Nevertheless it seems that in the near future years the current model will prevail.

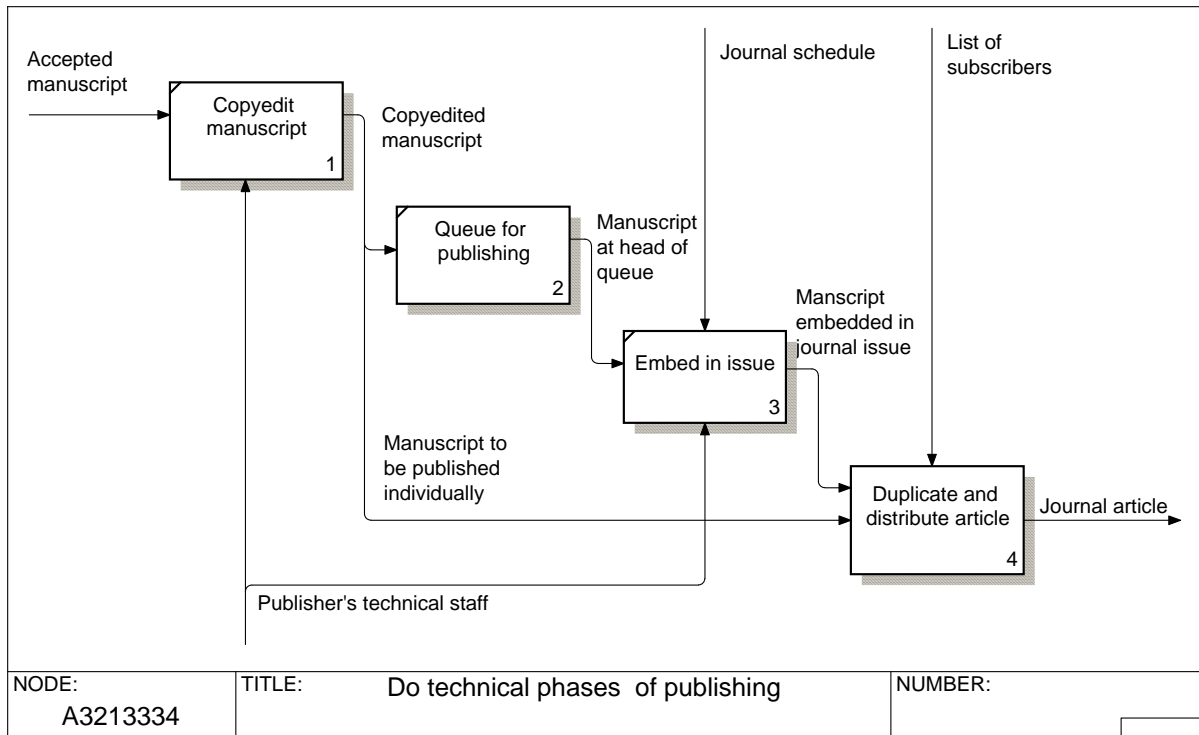
The peer review part of the overall process is interesting because of the way it operates and is financed. It is a good example of the generic type of peer production [Benkler 2006] that due to the web now is starting to have more and more impact is certain niche domains of information and cultural production (i.e. Open Source software and Wikipedia).



A3213334 Do Technical phases of publishing

After a manuscript has been accepted a number of activities take place. One is the final proof-reading and copy-editing of the manuscript to improve the language and detect minor technical errors. Note in particular the value-decreasing activity of *queue for publishing*, where fully processed articles have to wait for several months due to the issue scheduling of the journal. Waiting does not imply a direct cost, but there may be an important opportunity cost involved from the viewpoint of the researcher and society, since the results are poorly spread before the actual publishing. This opportunity cost is different for different domains of science. It might be low for the humanities but is usually higher in the STM (science, technology and medicine) domains. This has been a strong motivator for the founding of both e-print archives and electronic open access journals.

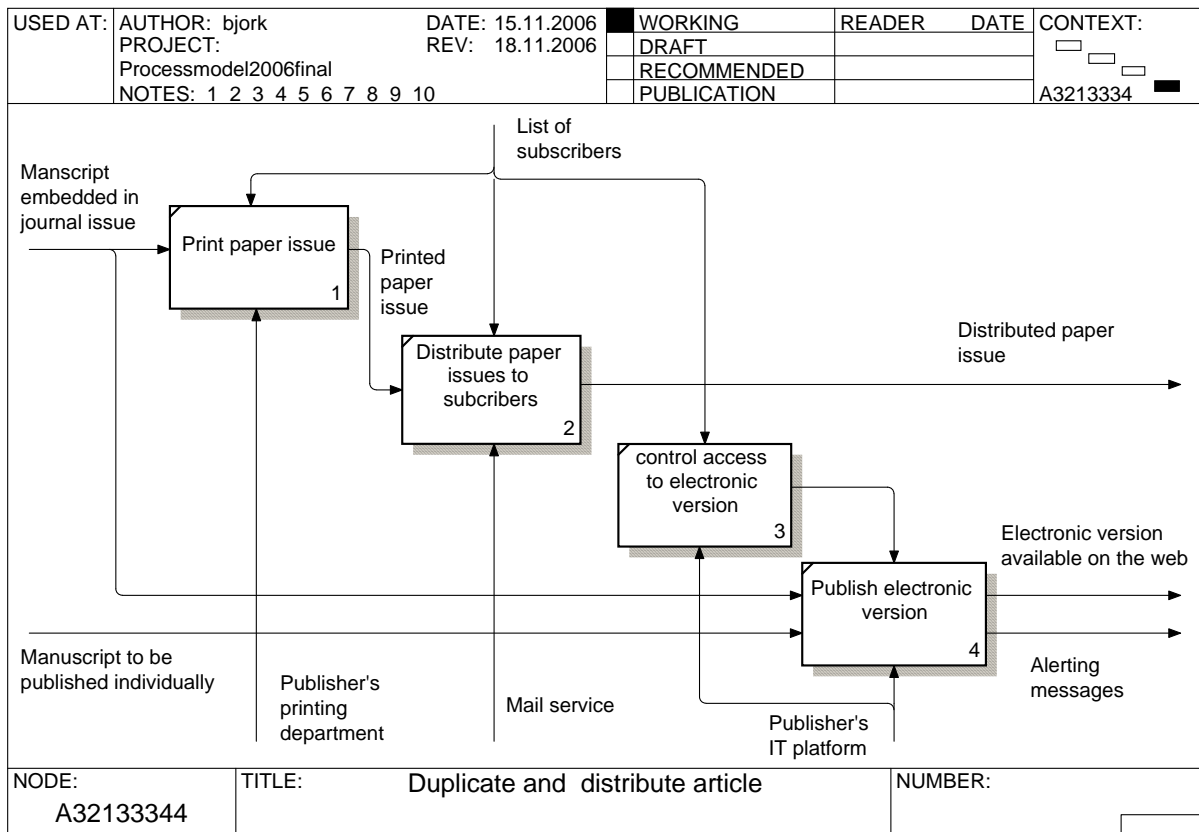
Open Access journals which are not issue based cut down this waiting time to a minimum by publishing an article when it is ready. Subscription based journals which appear both in print and paper, nowadays also sometimes post accepted papers even before formal publication on their publishing platform, thus offering a partial remedy to the queuing problem.



A32133344 Duplicate and distribute article

This diagram models the two parallel activities of publishing a print and an electronic version of an article. Paper publishing involves the traditional printing and distribution activities, which have a much higher marginal cost per copy produced than electronic publishing. Pure paper or pure electronic publishing can be obtained as subsets of this part of the model. Many of the bigger publishers nowadays publish both paper and electronic versions in parallel.

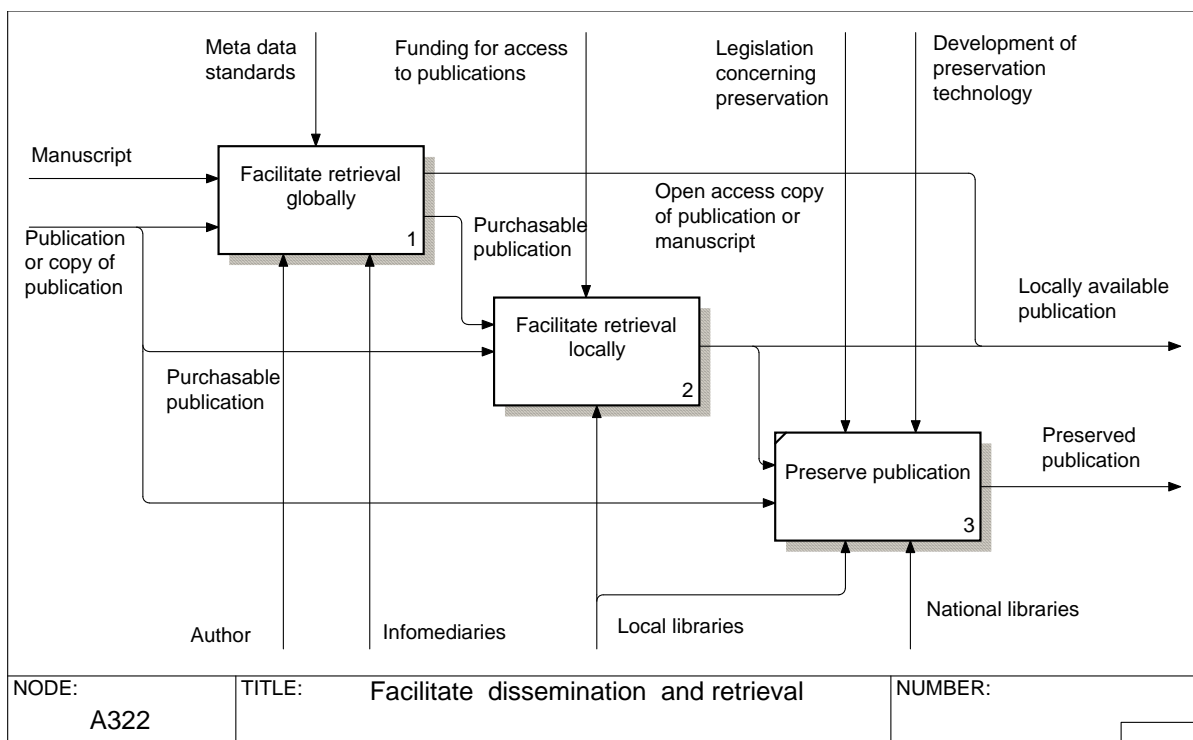
The central difference between paper and electronic publication is that paper (in this context) necessitates a revenue model of subscriptions or pay-per-view, whereas electronic publication due to the almost zero marginal cost of additional readers can be used in conjunction with a larger variety of revenue models.



A322 Facilitate dissemination and retrieval

This is the part of the overall process which traditionally to a large part has been handled by intermediaries and research libraries. In this diagram the process has been split into two sub-activities in which the first models activities carried out by different infomediaries, typically only once for the whole world market, and the second the activities carried out in the local organisations of the readers, thus typically hundreds or even thousands of times for each article or journal issue.

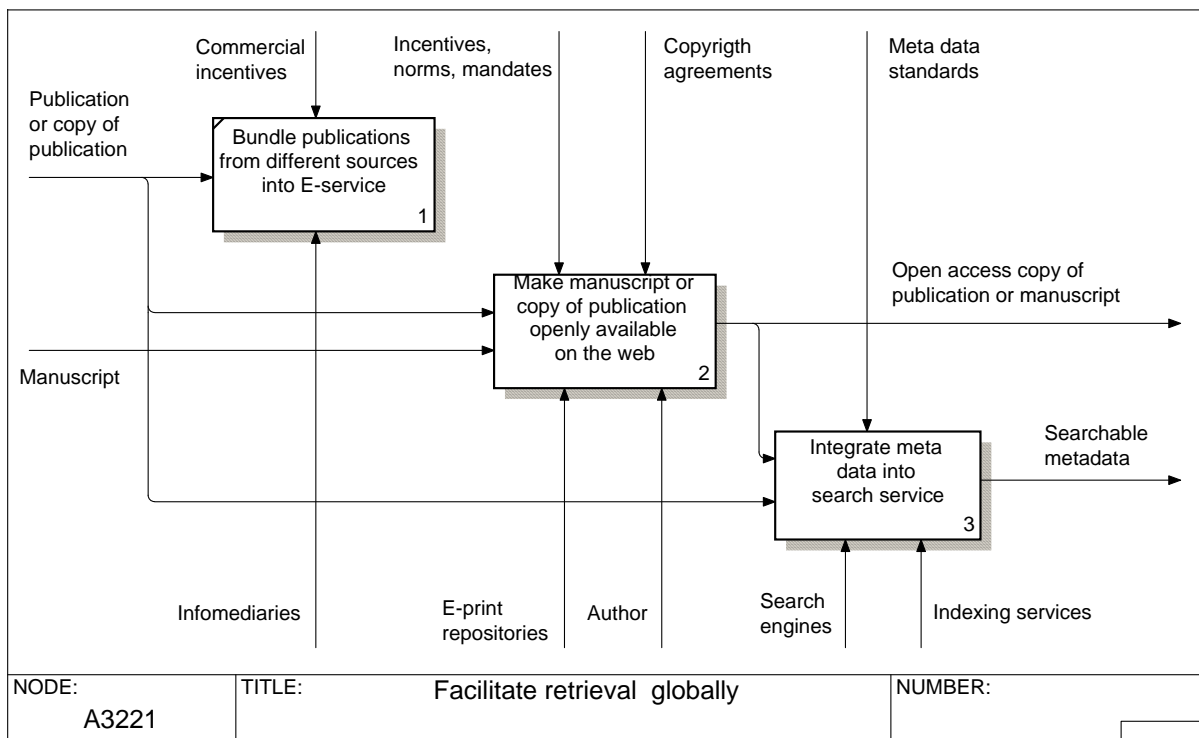
The *preserve publication* activity is currently receiving increasing attention, since the archiving of electronic versions of journals for decades implies a number of problems. National libraries in many countries are getting involved in this. The long-term preservation issue is both technical and organisational, since the subscribing libraries no longer get to possess physical copies of the material they subscribe to. Often it is unclear in the licensing agreement what happens if they cancel an electronic subscription to access to older material. Also the responsibility of publisher's in case of mergers or cancellation of titles is accentuated for electronic material.



A3221 Facilitate retrieval globally

This diagram includes a split between open access material, which can be either in the form of manuscripts or copies of formally published papers posted in e-print archives, and in “toll-gated” material. For purchasable material a further split is made into secondary publishers who bundle full-text material from several different sources (an example is EBSCO) and sell it to libraries, or indexing services, which help in the retrieval function. Note the importance of the *copyright agreement* for an individual publication, which acts as a control of the posting of a copy to an e-print archive.

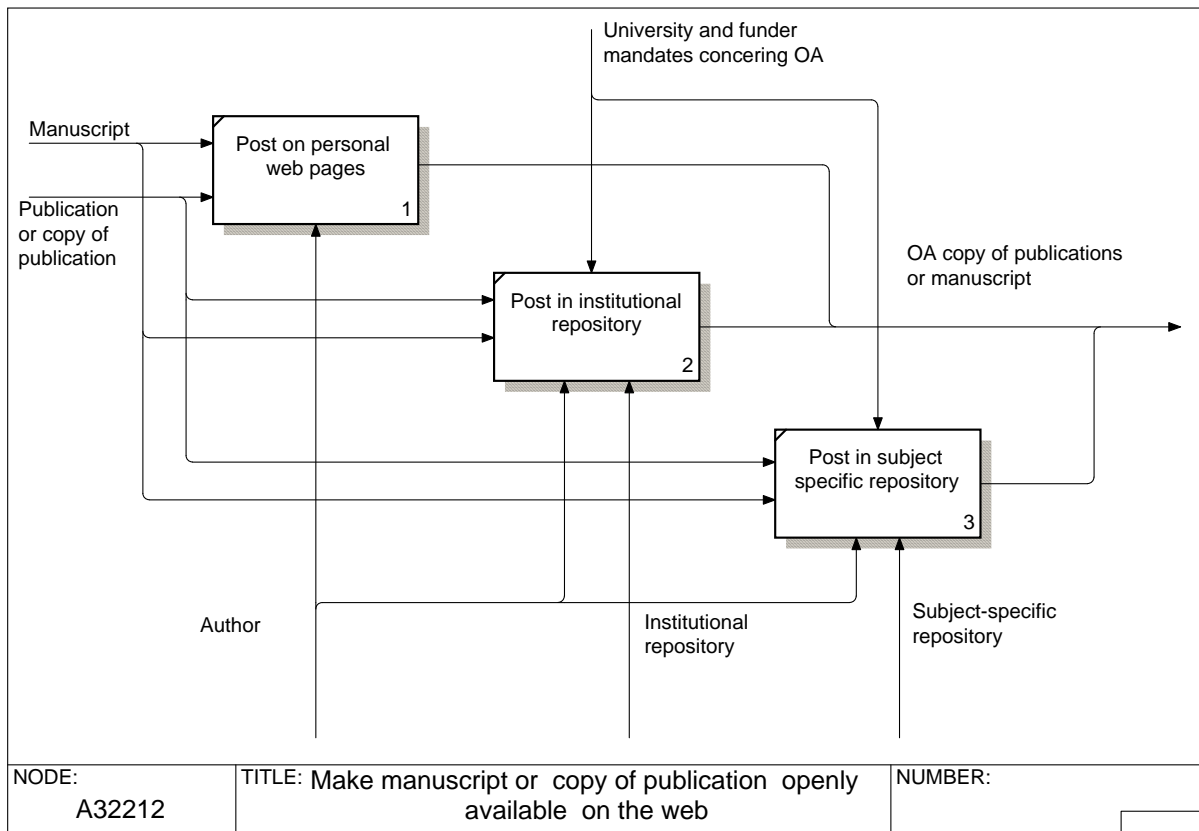
The third sub-activity of this diagram shows the integration of the meta data about a publication (including a possible web address) into different sorts of search services, whether free or subject to subscriptions. This is where tremendous developments have taken place in the last few years. The control activity of such services are different sorts of standards for the semantics and syntax of metadata (for instance the Open Archives Initiative standard for tagging open access repositories).



A32212 Make manuscript or copy of publication available openly on the web

This diagram depicts the different options available to an author for posting a manuscript or a copy of the actual publication in an open access archive. Three main options have been modelled, posting on one’s own home pages, in the institutional repository of one’s organisation or in a subject-specific repository (i.e. arXiv). In practice the flow of manuscripts or ready publications differ for these three alternatives. These are regularly posted in institutional repositories and subject specific repositories have in the early days in particular been populated by copies of manuscripts submitted to journals or conferences (hence the name preprint repository).

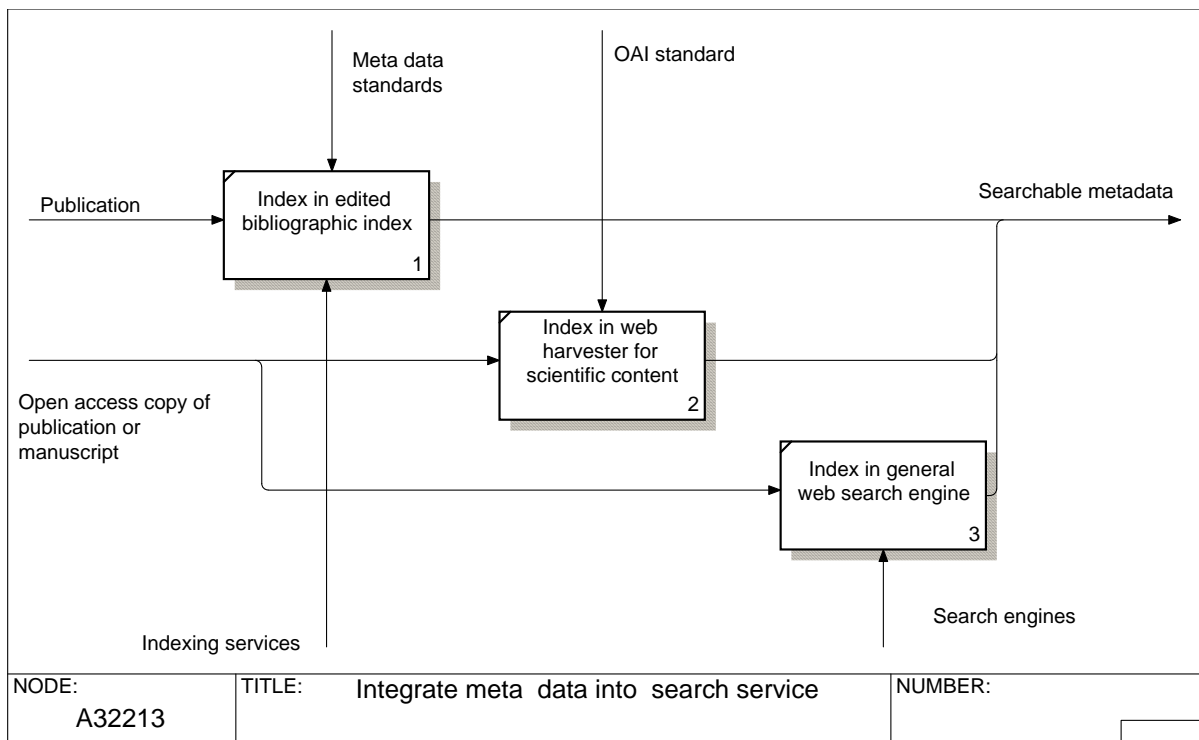
The controls of these activities are quite interesting. They can consist of both reward mechanisms, behavioural norms and in the case of repositories also direct mandates by research funders and universities that require posting copies of publications in such repositories. Finding the proper mix of “controls” will be a key success factor for repositories.



A32213 Integrate Meta Data into Search Service

Traditionally subscription-based indexing services have dominated this function. Over the past years researchers have increasingly started to use general web search engines for identifying and retrieving interesting publications. An effort to overcome the quality problems related to using general search engines, is the definition of the Open Archives Initiative [OAI 2006] standard for tagging scientific content material on the web, which will enable dedicated harvesting search engines to maintain a much more focused data base of links to relevant publications.

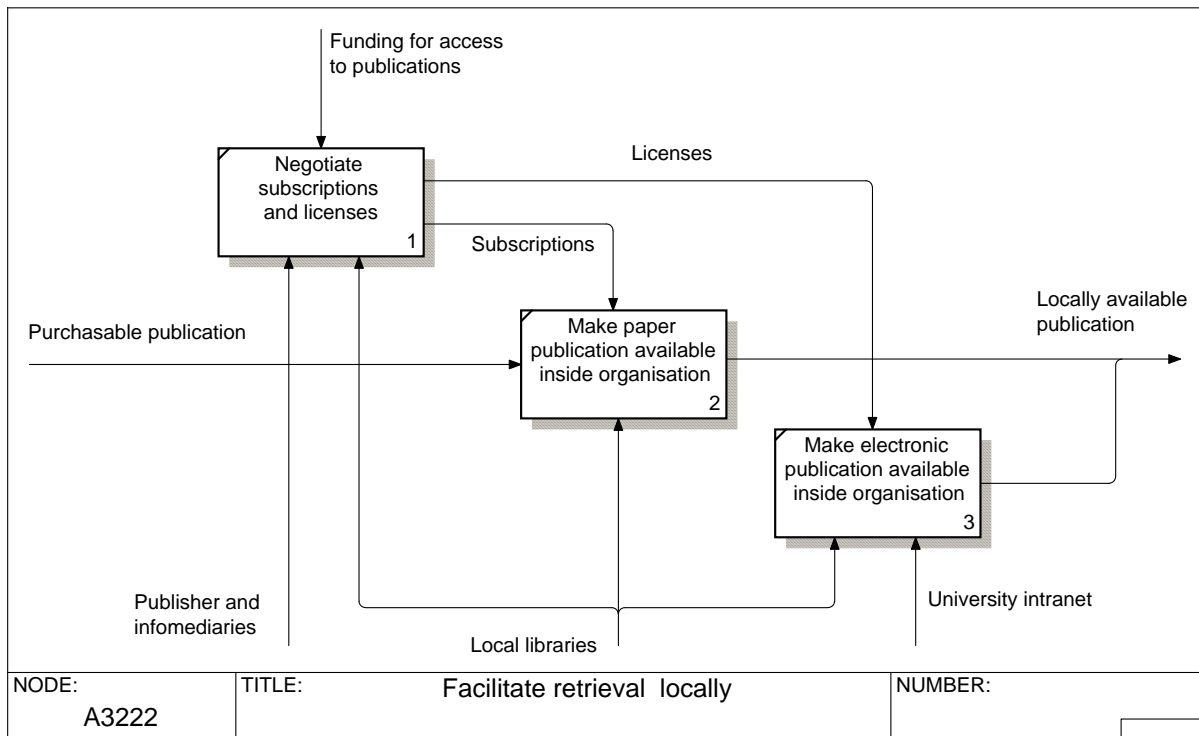
A by-product of the heavy use of IT for these purposes is the possibility of readers to subscribe to services, which based on the interest profiles they define, can send them alerting email messages when something they might be interested in is published.



A3222 Facilitate retrieval locally

This diagram shows the activities carried out by the organisation in which the reader works to facilitate access, for instance the university library of the reader. Note the inclusion of a separate activity for the negotiations that the library carries out in order to obtain the necessary licenses (the activities by library consortia could be included here as well as a sort of overhead cost). The negotiations sub-activity also includes collection planning, making decisions about which journals to subscribe to etc.

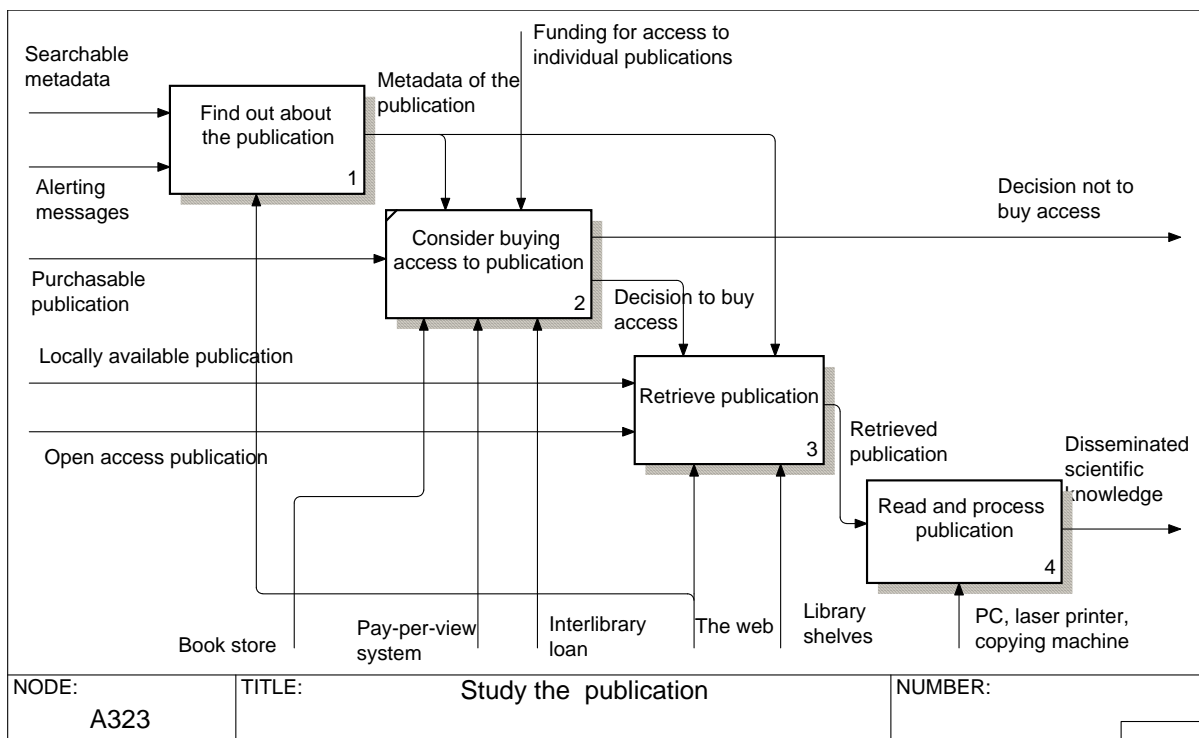
One of the biggest changes that electronic publishing has brought is the dramatic reduction in the activities to make paper publications available inside the organisations. On the other hand libraries now have to use substantial resources to build intranets, which seamlessly organise all the heterogeneous electronic material they have bought licenses to (also to solve the problems of distant access for faculty member or students working from home).



A323 Study the publication

This diagram structures the activities of the readers of scientific publications. Note that from a cost per publication viewpoint the activities of individual readers all over the world and over the whole life-cycle of a publication's readings should be summed up. The *find out about the publication* activity results in the output *metadata of the publication* (including the location from which the paper or electronic version can be retrieved). This output is used as the control of the retrieve publication activity. If the publication is not already available through a subscription or as open access a decision has to be made whether it is worth the monetary cost and effort to purchase a copy. Often this can be done at a cost of say 20-30 USD for articles. One important factor, which nowadays might negatively affect a decision to separately buy an article, is the delay in getting it. Also the researcher may have to pay such acquisitions directly from his own budget which usually means that the barrier for obtaining it are higher than if it was already a part of central library licenses.

Finally the publication is read and the scientific information in question has been disseminated. Researchers often self-archive interesting publications they have read either as paper copies or today increasingly as bookmarks or in a data base. They may also use citation tracking applications such as Endnote to keep track of literature they are likely to reference later on.

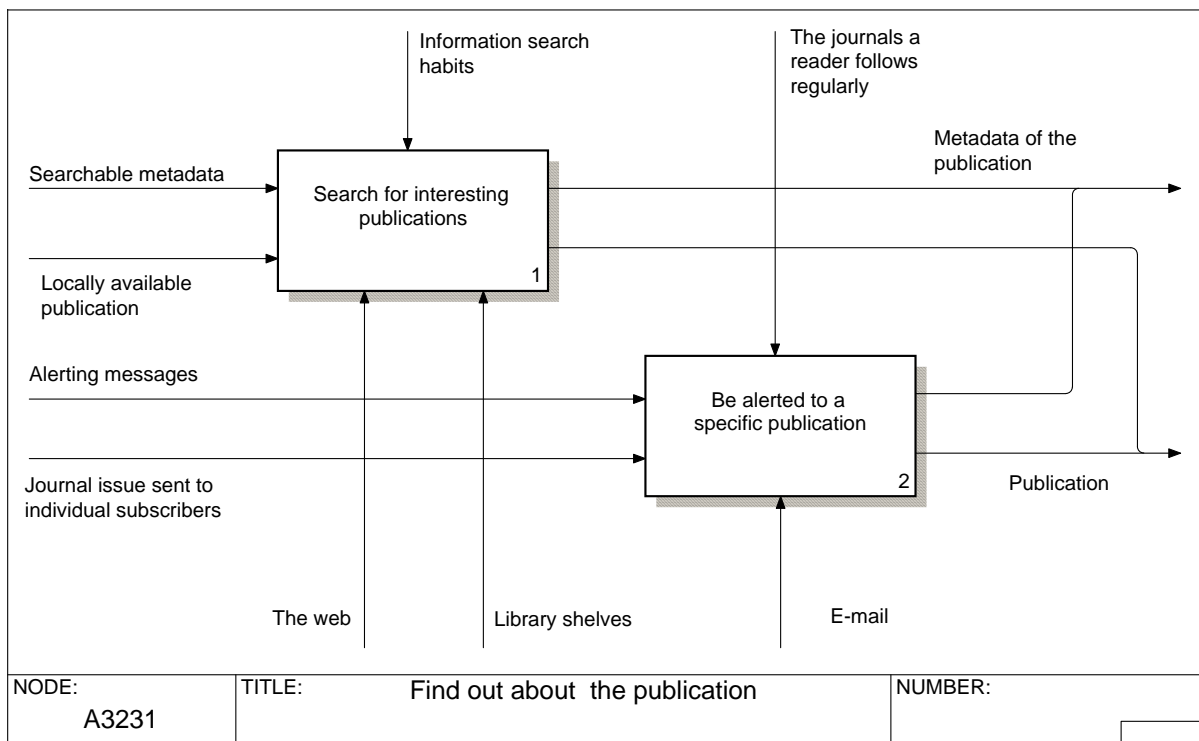


A3231 Find out about the publication

This activity is rather difficult to split up into alternative parallel options. On this first level there is, nevertheless, a split into two generic categories; active search (pull) where the reader is pulling for information and passive push, where the reader receives a notification through some mechanism that something interesting has been published [Björk and Turk 200].

The pull type activities are guided by the information search habits of the researcher in question and the main vehicle today is the world wide web, and the various search engines and dedicated search facilities on offer, including the intranet of the local university library, which offers a search facility of the locally subscribed publications.

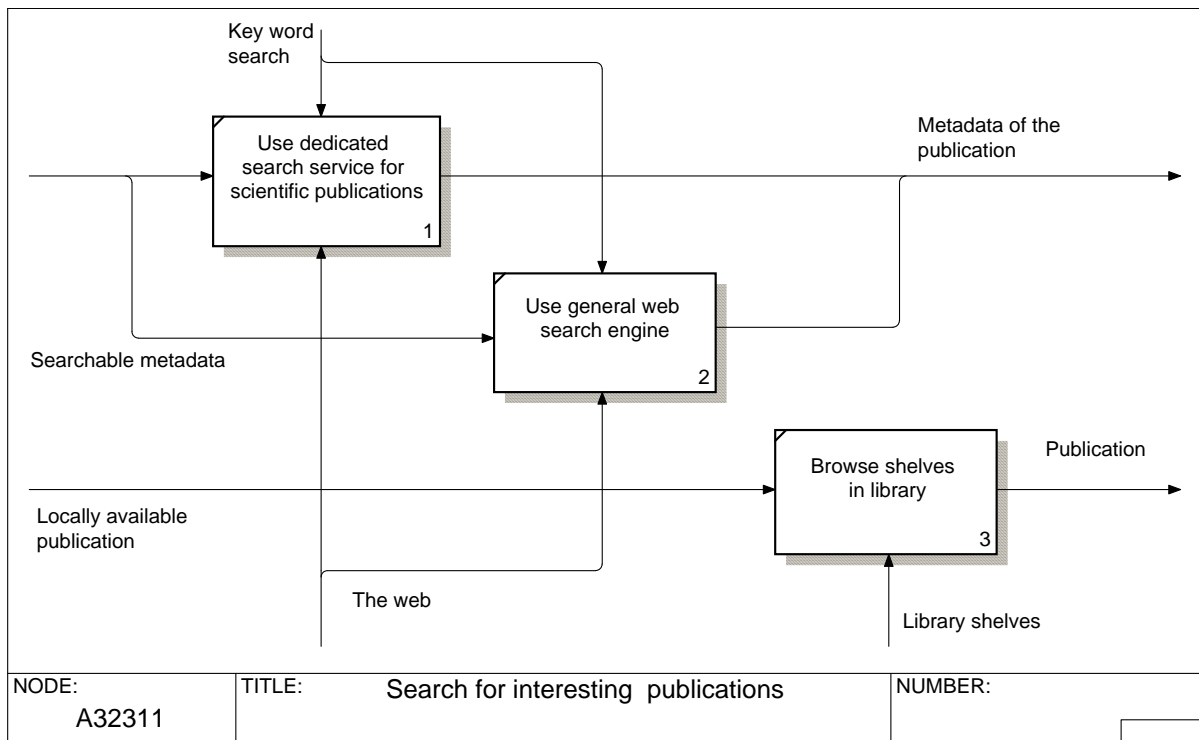
The push variety has been greatly enhanced due to email. Most researchers regularly follow a handful of journals and screen everything that's published in them, at least on the title or abstract level. In the print world this was achieved by getting physical copies of the journals, either as personal subscriptions or as a node in the internal circulation in a university department, but today the same functionality can be achieved faster via emailed tables of content.



A32311 Search for interesting publication

The first modelled option of the pull variety is a traditional bibliographic data base search, for instance using keywords. Other possibilities include more unstructured web searches using a general search engine or just browsing from one hyperlink to the other. Less and less the traditional method of physically browsing the bookshelves in a library is used.

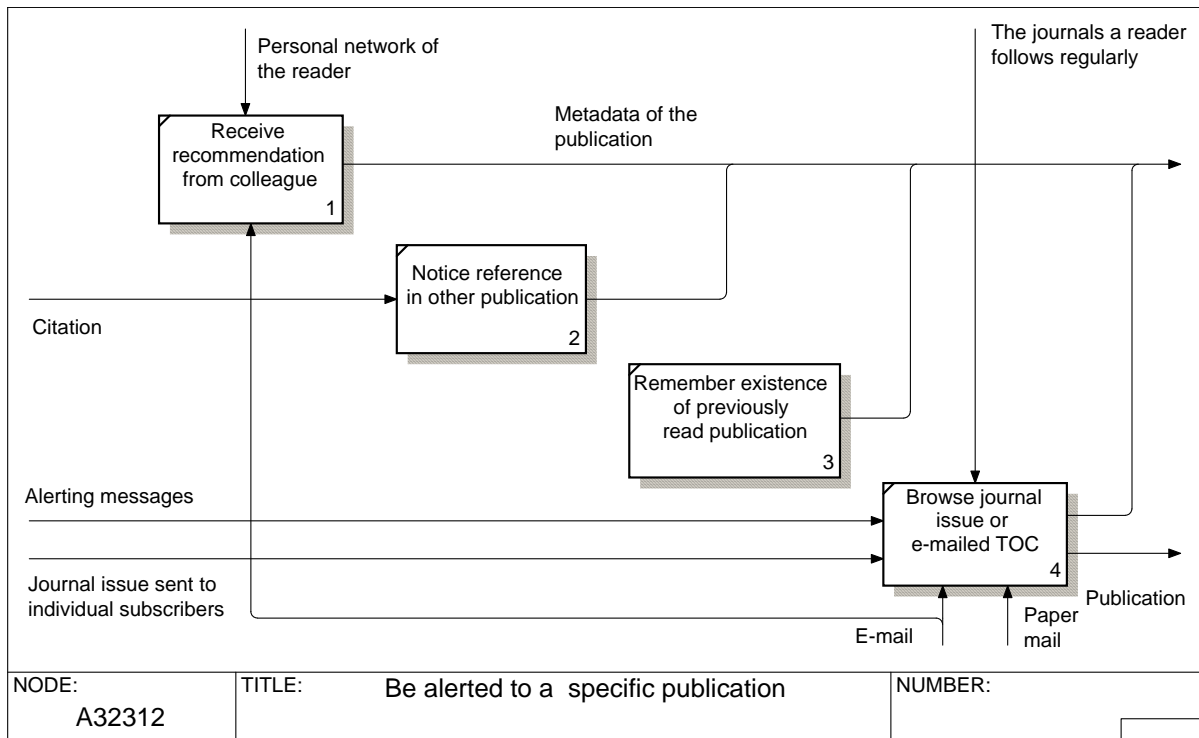
The pull variety is much used by students when they are working on a thesis and in the pre-study stages of a research project, to find out what's been written about a subject. It is less used by senior academics and practicing physicians, who more tend to read for current awareness.



A32312 Be alerted to specific publication

The important distinction between pull and push is, that the bibliographic search is triggered by the researcher himself (pull), whereas the hint is coming from the outside (push). Tracking a reference in another publication and retrieving the full text is very common. In practice a lot of references are not followed up, if the potential reader finds that access is blocked due to a lacking subscription. From society's viewpoint this means an opportunity cost, since the reading not done could potentially have been an important input to further research or other activities.

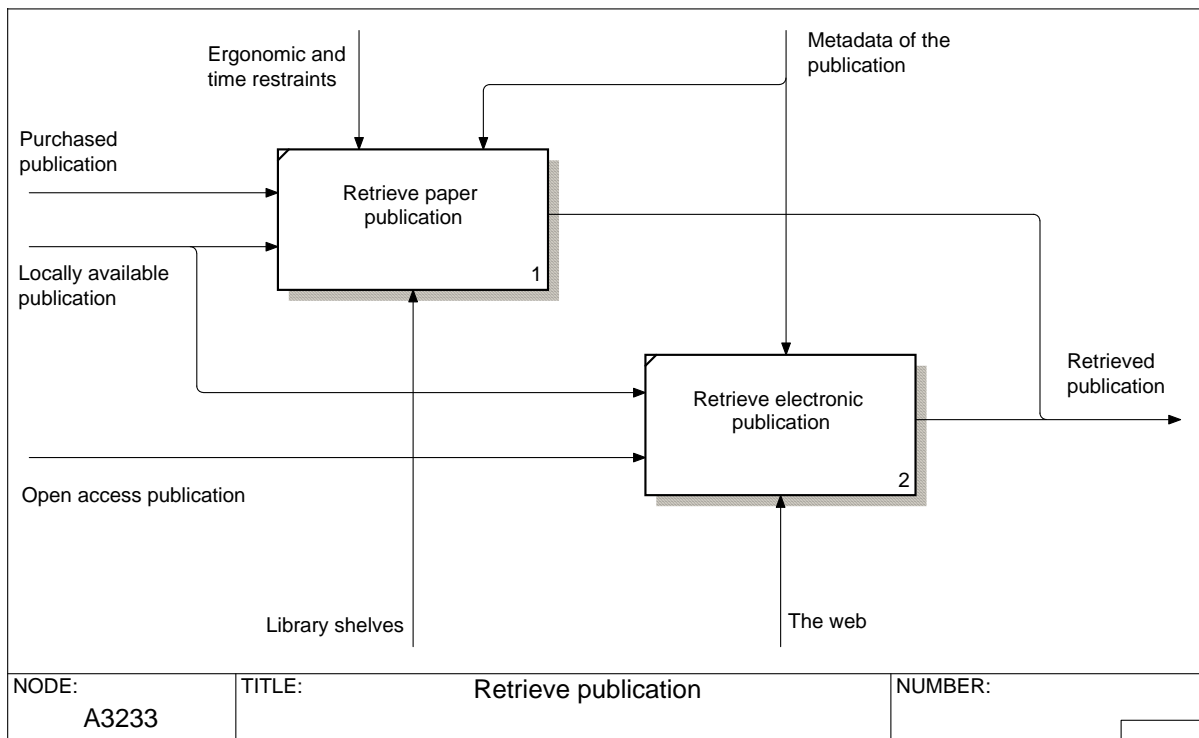
Browsing a journal issue which is physically circulated in an organisation is less and less common as libraries convert to electronic site licenses, but almost the same functionality is achieved by emailed tables of content (TOC), which offer the additional advantage of no delays. The number of personal paper subscriptions have decreased in recent years, expect for certain fields such as medicine. The option *remember existence of previously read publication* is modelled because this is sometimes used as the basis for retrieval.



A3233 Retrieve publication

The basic split has here been made between the retrieval of a copy of a paper publication and one in digital form. Although the two activities modelled here may look straightforward, they might become rather complex in reality. Underneath retrieving a paper copy is a complex infrastructure of library organisation and storage of material, and electronic retrieval can on the software side become quite complex, especially for material not openly available on the web.

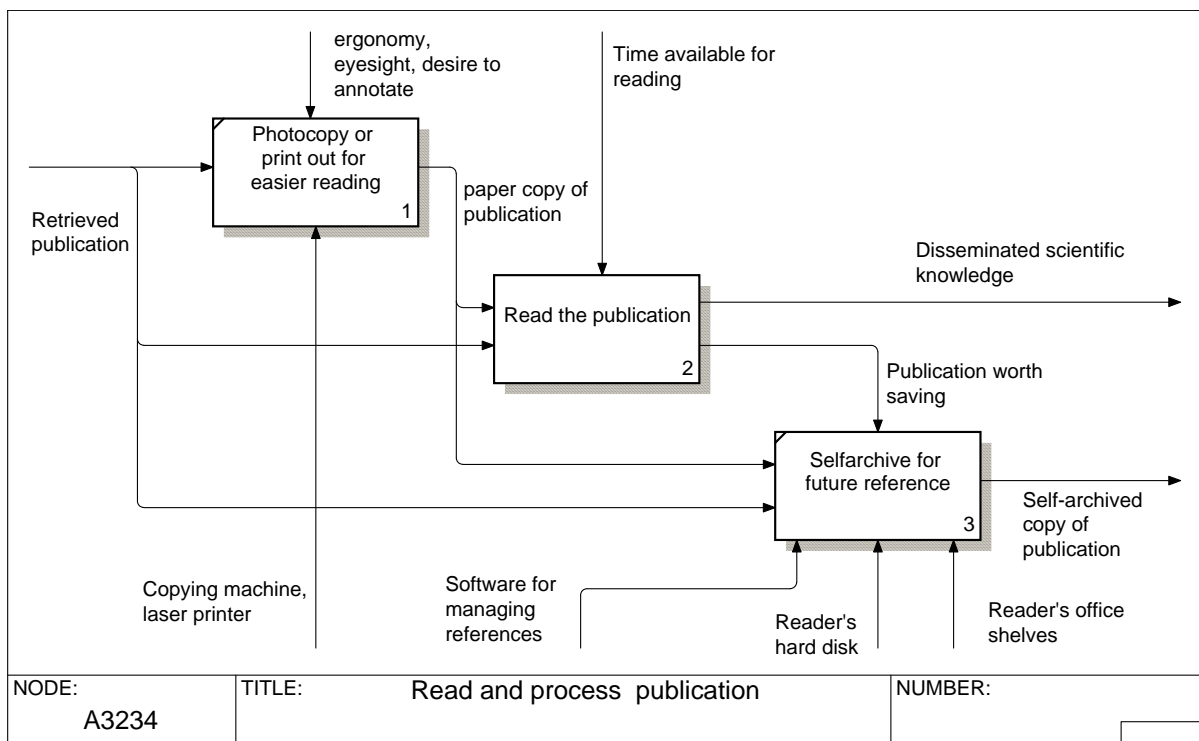
From the viewpoint of the individual researcher this is an activity where the developments with the Internet have had huge impact. A large part of the research literature is now virtually within easy reach, in fact just a few clicks and keystrokes away.



A3234 Read and process publication

This diagram, which also could be called assimilate the knowledge, shows the activities a reader typically does. A typical pre-processing activity is making a printout of an electronic publication (usually an article) for easier reading and annotation. A typical post-processing activity is storing a copy of an article found to be particularly interesting, as the recipient integrates the knowledge into his personal “store of knowledge”. The storage could be physical, by putting photocopies in folders in one’s file cabinet, or electronic, by storing full text files or just web bookmarks on one’s computer.

The reading itself can be on many different levels, from superficial browsing to determine whether an article is worth attention to very detailed reading and annotation. This aspect has not been further detailed in the model.

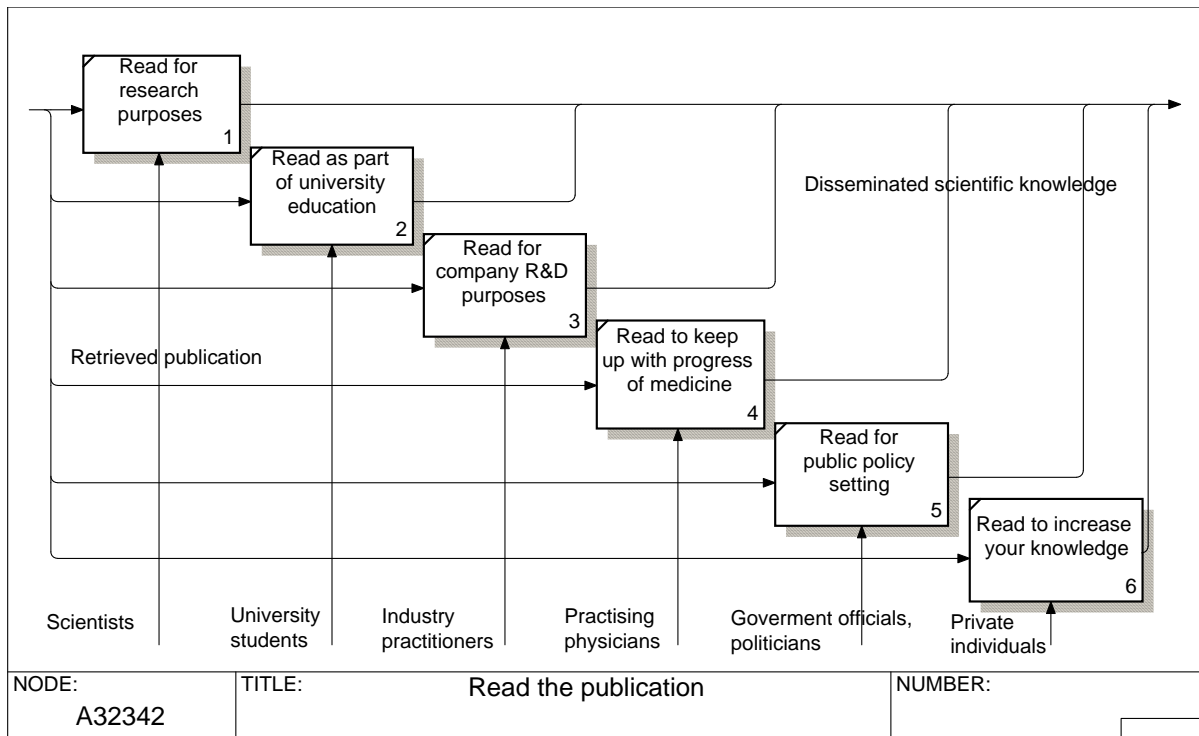


A32342 Read the publication

Since readership and dissemination patterns differ a lot for different categories of readers of scientific publications, the model has here been split up into different activities for the perceived main categories.

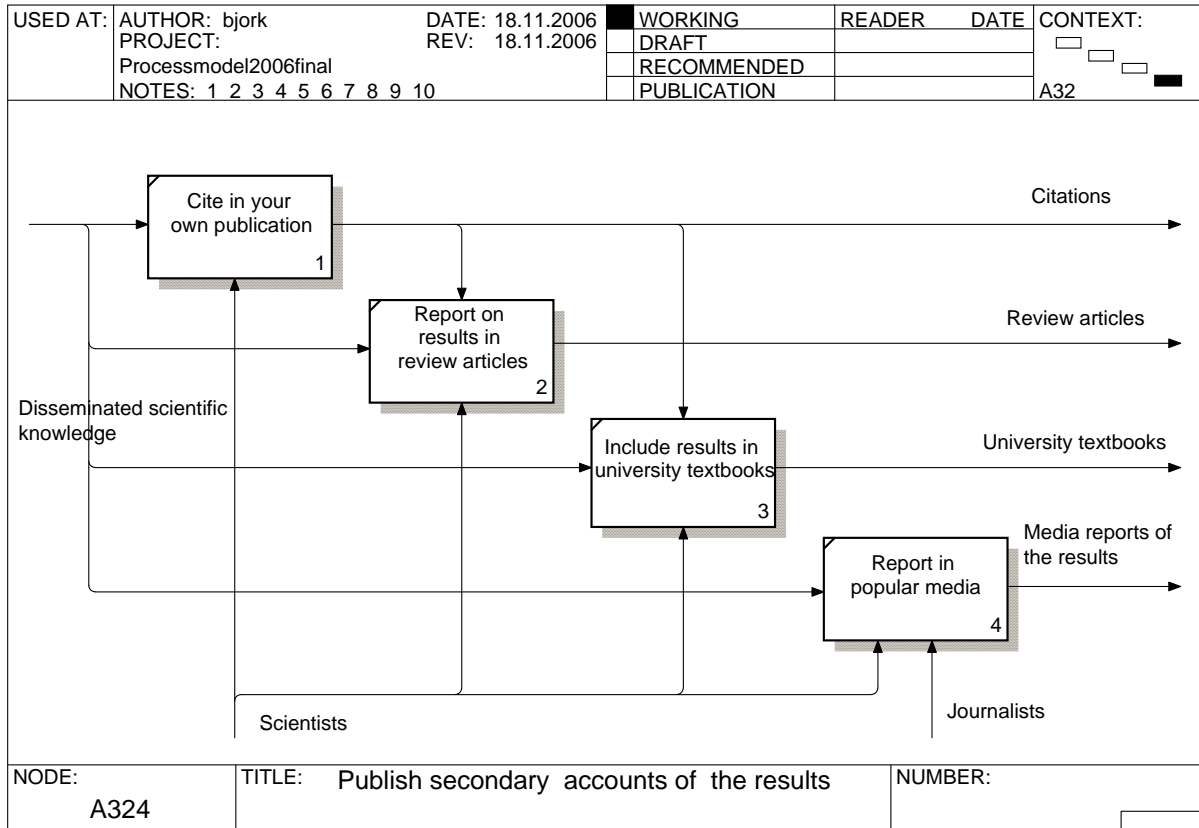
Academic readers tend to read for research purposes. They are usually well supported in terms of large collections of journals that their university provide paid access to. An important subcategory are students, who don't do research themselves, but who in certain fields read a lot, as part of assignments etc. Company experts read in order to carry out industrial R&D and develop better products or enhance the operations of their companies. The usage levels vary tremendously between industries.

Practicing physicians represent a very large volume of readership of scientific journal articles, because they need to constantly keep in touch with the latest developments. Also the journals they read tend to have very large subscription bases and they often get personal paper copies at very reasonable prices. The general public also is some cases follows the progress in science. Here is one area in which open access could have a significant effect, since general libraries as of yet have been able to provide very limited access to the research literature. One important subcategory of readers are also government officials and politicians.



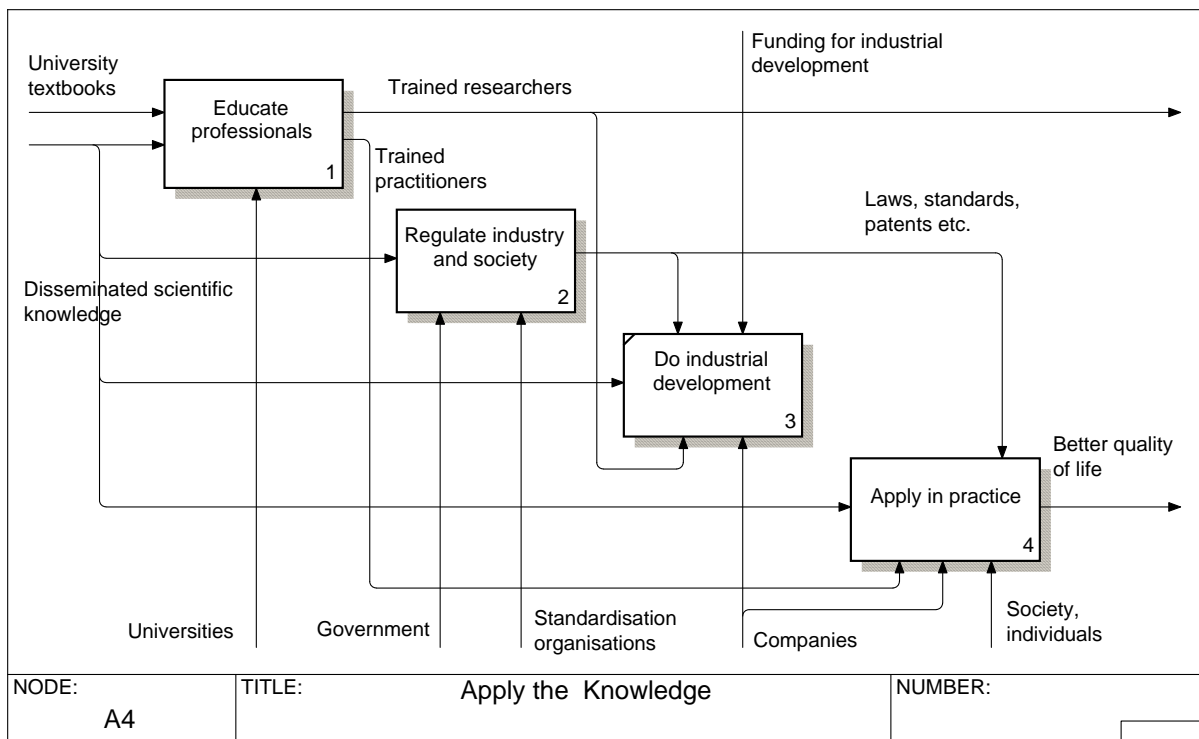
A324 Publish secondary account of the results

A small part of all new scientific publications have a lasting influence on the development of science. From a communication perspective this is achieved via other scientist referring to the results in their own publications via citations. These enable other interested scientists to find out about the results and retrieving the quoted publications. Also the results may become incorporated in the standard textbooks of a field and they may be reported on in articles and news items in the mass media.



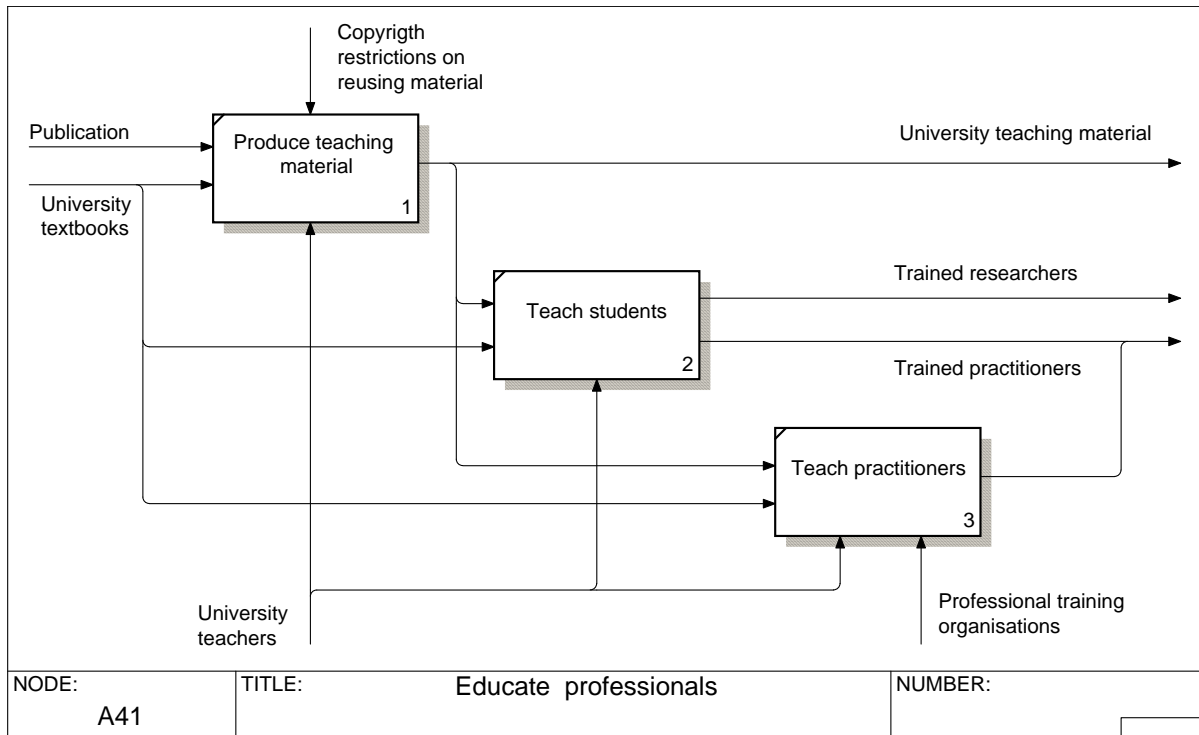
A4 Apply the Knowledge

In the same way as the breakdown of the research itself this diagram is more of a contextual nature. It tries to show how disseminated scientific knowledge can be transferred by several parallel mechanisms into better industrial performance, new products and services and eventually a better quality of life. This is an aspect of scholarly communication, which for instance is discussed in a recent Australian report (Houghton et al 2006), where the estimation of the impact of the increased readership caused by open access on use of the results in industry and society in general is attempted. One of these mechanisms is education and training, which results in better trained professionals who go out into working life (i.e. medical doctors and engineers). There is a rather straightforward link between research and especially university education. A second mechanism is through commercial development work, which translates research results into new products, services and working methods. An important sideline of this is the patenting of inventions, which exerts a very direct control on the application of the results of science in practice, as well as standardisation. Another mechanism is where individual citizens read research publications and are directly affected by them in their behaviour or quality of life. Yet another is where research results as reported directly or indirectly, influence laws and government actions (for instance taxation, public spending).



A41 Educate professionals

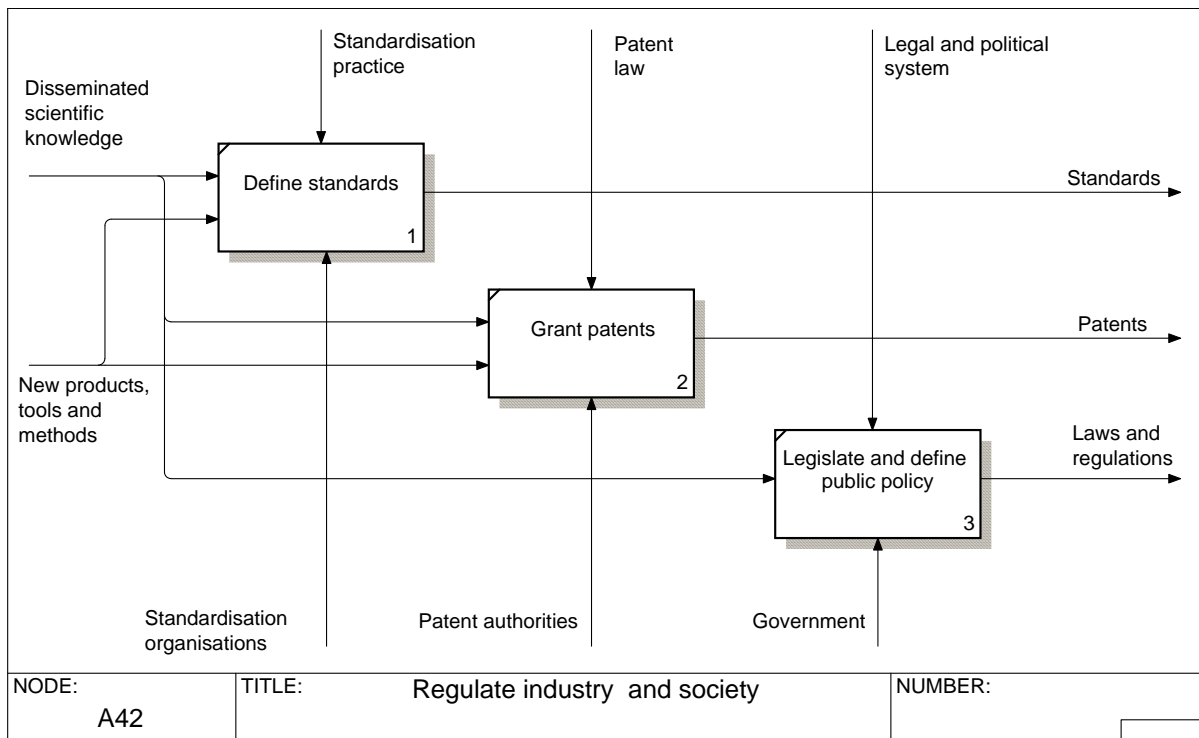
One aspect of the use of scientific research literature is its use as part of the teaching material used in university level education. Traditionally there have been certain rules as to the use of photocopying of research articles, which the university has subscriptions to (“fair use”). The availability of articles as open access clearly facilitates their use in courses and parallels the development called open course-ware (a term originally coined by MIT), which is concerned with the teaching material (course texts, presentation slide, etc) developed by teachers.



A42 Regulate industry and society

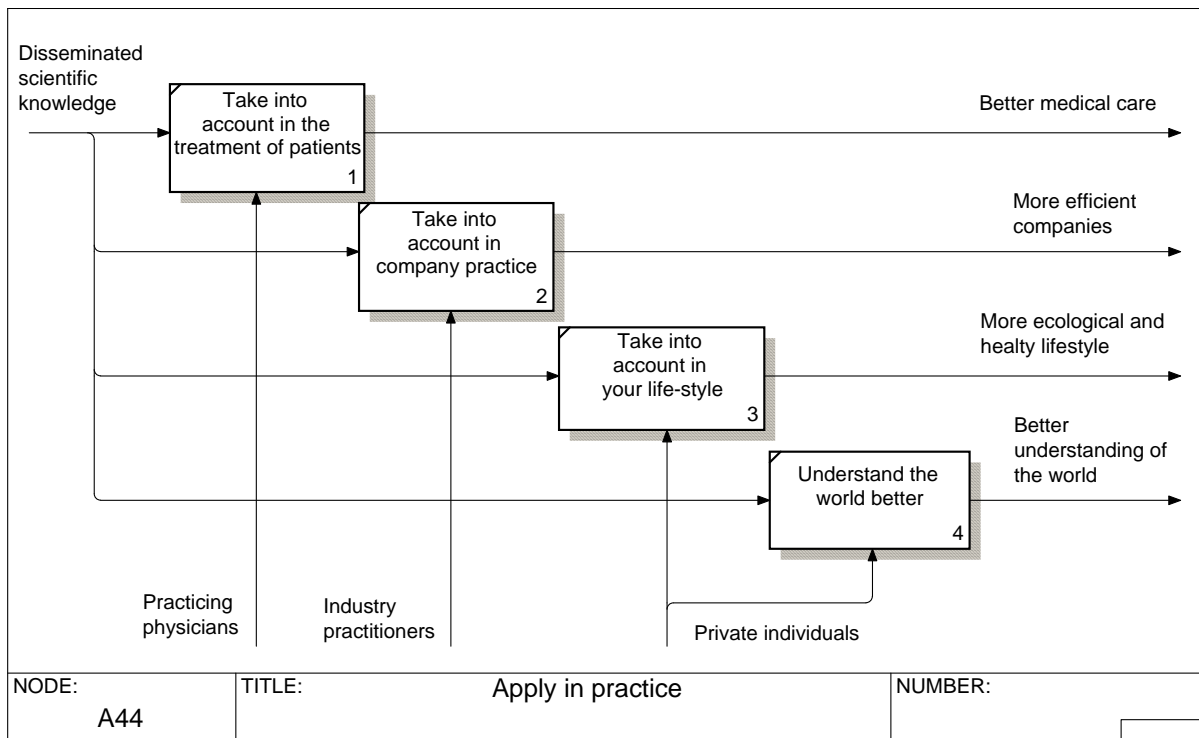
Disseminated scientific knowledge has a big direct and indirect influence on a number of measures, which influence industrial development and daily life. To quote some examples:

- Research on the stability of steel and concrete structures in civil engineering has influenced building regulations and standards.
- Research on global warming and its possible causes has influenced public policy (Kyoto agreement, emission trade between power utility companies).
- Economics research, for instance the work on Keynes, has had a huge influence on fiscal and monetary policy
- Patenting is a very important and highly debated mechanism for protecting the rights of companies making discoveries. Thus the access to both patent information and the possible underlying research publications is very important. In particular in areas such as genome research and the development of medicines this is a central issue.



A44 Apply in Practice

This diagram should be read in parallel with diagram A32342 (Read the publication) which split up the readership of scientific publications into a number of generic categories. The diagram shows the beneficial after affects of the disseminated research knowledge (excluding the feedback effects of science itself). Thus better awareness of the latest research helps doctors in the diagnosis and choice of treatments and medicines for patients. Better awareness of say the latest management theories may help the senior management in companies to the development of company practice. And better knowledge of scientific results may either help the general public in doing better informed decisions in for instance energy saving matter, or may be a value in itself (self-improvement as an end in itself).



4. Conclusions

Discussion

Based on the literature review it is the conclusion of the author that this is the first time a formal process modelling methodology is used to model the system of scholarly communication in this comprehensive way. Publishers employ methods of a similar nature to study the workflows within their organizations, but the whole point here is to study the whole system, including the activities of libraries and readers.

The model in its current shape has not been validated in its details, but has been discussed with several colleagues with encouraging feedback. It would in fact be very difficult to design a method for the validation of the model. Flaws in details of the model could be pointed out but it would be difficult to “test” the model as a whole. Every stakeholder in the overall process has a different perspective on the process. The only realistic test of the model is to show it to people and ask if they find it useful in creating a better understanding of the overall process.

In the modelling itself some rules of IDEF0 modelling have not been strictly enforced. Tunnelling of arrows (marking arrows that will not be inherited to more detailed diagrams) has not been done. Also in several possible ICOM:s which exist in reality have not been indicated. The overall design consideration has been simplicity and showing the essentials, in particular concerning the break into activities. This author has seen many IDEF0 models where the task of communicating the ideas to others is defeated by too complicated models.

Compared to the earlier models discussed earlier in this paper the main differences are:

- Hierarchical structure of the model
- More modelling constructs, i.e. controls and mechanisms
- Much more detailed modelling of many of the functions
- Disaggregation of inputs and outputs on more detailed levels
- Modelling of many of the new system functions that have emerged as a result of the Internet

It is hoped that the model could prove useful in providing a roadmap showing the place of a number of different initiatives for increasing access to scientific publications, within the overall system of scholarly communication.

Further research

When discussing the economics of scientific publishing we need to look at the economic effects of changes in some parts of the system on the whole system. It should be clear from reading the model that changes in one part can have profound repercussions in other parts. Currently much of the emphasis in the debate seems to be only on the publishing phase costs and pricing², but the effects on indexing, library access, readers’ retrieval costs and even on the quality of downstream research and product development should be considered. It is like considering the effects of pricing of underground tickets in a metropolitan area. A sharp decrease in prices (or making it free altogether)

² For instance in the report commissioned by the EC [2006]

will have big impacts on the use of cars for commuting, traffic congestion (and non-productive time spent waiting in queues) and even air pollution.

In an envisioned follow-up study, which would concentrate on journal publishing, there are four major scenarios to compare between. The scenarios are fictitious extreme cases and at any time reality will be a mixture of these four models.

Paper. All journals are published and circulated in paper only. Libraries around the world need to circulate and store them internally. Researchers can search for articles using electronic indexing services. This was the situation until around 2000. This scenario is included to show what changes have been induced by the changeover from paper to electronic publishing

Electronic - subscription based. All scientific journal publishing goes over to an electronic only mode. The subscription model continues to be the overwhelmingly dominating model and pay-per-view is a marginal phenomenon. For the sake of argument it is assumed that there is no OA primary and secondary publishing.

Electronic – Open Access. All journals convert to Open Access primary publishing, using the article charge method. This is by OA advocates often called Gold route to OA [AMSCI 2006]

Electronic – subscription based supplemented by parallel posting of copies in OA repositories. The electronic-subscription model continues to dominate but due to liberal copyright policies all authors are able to upload personal version copies of their articles (perhaps after delays) to institutional OA repositories. It is assumed that this does not affect subscriptions negatively and that only those readers who don't have subscribed access use the green road. This is by OA advocates often called Green route to OA [AMSCI 2006].

The SCLC-model can then be used to construct a spreadsheet containing all the activities involved in the process as rows and the above four scenarios as columns. Differences between the scenarios occur in the form of some activities not being needed in all, and also in the unit costs for different sub-activities. The unit costs need to be normalised to some standard measure and that measure should be the costs per single published article. By multiplying over the number of articles published annually, number of average readings over the life-time etc, global costs over the life-time can be calculated.

The type of costs can be both directly monetary (subscription fees), time based (readers time retrieving and reading an article, a reviewers time for reading and commenting a manuscript) and opportunity costs. This latter type is very important and represents the value lost because at the margin a number of potential readers did not read a paper because of lacking subscriptions, laziness in retrieving a paper not instantly available etc.

A hypothesis to be tested in such a study could be as follows;

A massive move from subscription based electronic journal publishing to OA publishing, financed by article charges would:

Very significantly reduce the global costs of the system, given the same spread and volume of readership as in the subscription based model.

At the same time increase readership very significantly thus adding value in the form of spin-off effects on future research as well as in the form of better application of research results in practice.

Doing this type of research is extremely challenging and would among other difficulties entail the gathering and synthesizing of empirical evidence from many different studies. Hopefully the model presented here could be of use in integrating such evidence.

Acknowledgements

This model has been developed in several stages, including the SciX project (<http://www.scix.net/>), funded by the European Commission, and the OACS project (<http://www.oacs.shh.fi/>), funded by the Academy of Finland. Several colleagues have provided input and feedback to the model as it has evolved. Carol Tenopir, Turid Hedlund and Jonas Holmström have in particular also commented the draft of this article.

References

- AMSCI (2006). **American Scientist Open Access Forum**, Email Listserv. Retrieved 20 November 2006 from <http://american-scientist-open-access-forum.amsci.org/archives/American-Scientist-Open-Access-Forum.html>.
- Benkler, Y. (2006). **The Wealth of Networks – How social production transforms markets and freedom**, Yale University Press, New Haven.
- Björk, B.-C. (2005). **Scientific communication life-cycle model, version 3.0**, Report, 2005-02-10, Swedish School of Economics and Business Administration, Helsinki, Finland, Retrieved 20 November 2006 from <http://www.oacs.shh.fi/publications/Model35explanation2.pdf>.
- Björk, B.-C., Turk, Ž. (2000). **How do researchers find and retrieve scientific publications – a case study of the impact of the Internet on the construction IT and construction management research communities**, *Electronic Journal of Information Technology in Construction, Vol. 5 (2000) / 5*. Retrieved 20 November 2006 from <http://www.itcon.org/2000/5/>.
- Björk, B.-C. and Hedlund, T. (2003). **Scientific Publication Life Cycle Model (SPLC)** *In: Proceedings of the Elpub2003 Conference, From Information to Knowledge, 25-28.6.2003, Guimaraes, Portugal*, Retrieved 20 November 2006 from <http://elpub.scix.net/cgi-bin/works/Show?0317>.
- Björk, B.-C.; Hedlund, T., (2004). **A Formalised Model of the Scientific Publication Process** Online Information Review, Vol 28/1, 2004, pp. 8-21.
- Björk, B.-C., Hedlund, T., Holmström, J., & Cox, J. (2003). **D2: Scientific Publishing: To-Be Business and Information Model**. A deliverable of the project Open, self organising repository for scientific information exchange (SciX), EC fifth framework programme, Information Society Technologies (IST). Retrieved 20 November 2006 from <http://www.scix.net/d/02/d2.pdf>.
- CRIS (2006). **Enabling Interaction and Quality: Beyond the Hanseatic League**, Proceedings of the 8th International Conference on Current Research Information Systems: (CRIS2006), Bergen, Norway, 11-13 May 2006, Leuven University Press, Belgium.
- ETAI (2006). **Electronic Transactions on Artificial Intelligence**. , Retrieved 20 November 2006 from <http://www.etaij.org/>.
- EC (2006). **Study on the economic and technical evolution of the scientific publication markets in Europe**, European Commission. Retrieved 20 November 2006 from http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf.

Goodman D. and Foster C. (2004). **Special Focus on Open Access: Issues, Ideas, and Impact**, Editorial, *Serials Review*, Volume 30, Issue 4, Pages 257-381.

Hars, A. 2003: **From Publishing to Knowledge Networks – reinventing On-Line Knowledge Networks**, Springer, Berlin.

Houghton, J., Steel, C. and Sheehan, P. (2006). **Research Communication Costs in Australia: Emerging Opportunities and Benefits**, A report to the Department of Education, Science and Training, Centre for Strategic Economic Studies, Victoria University, Melbourne. Retrieved 20 November 2006 from http://www.dest.gov.au/NR/rdonlyres/0ACB271F-EA7D-4FAF-B3F7-0381F441B175/13935/DEST_Research_Communications_Cost_Report_Sept2006.pdf.

Hurd, J. M. (1996). **Models of scientific communications systems**. In: Crawford, S et al (eds) *From Print to Electronic: The Transformation of Scientific Communication*, Information Today Inc., Medford, NJ, pp. 9-33.

Hurd, J. (2004). **Scientific Communication: New Roles and New Players, Science & Technology Libraries**, Vol 25, No 1-2, pp. 5-22.

Karstila, K.; Karhu, V; Björk, B.-C. (2000). **Development of process modelling tools for the construction industry** In: Goncalves R. et al (eds) : *Product and Process Modelling in Building and Construction ECPPM 2000*, Lissabon, 25-27.9.2000, Balkema, Rotterdam.

King, D. W., Tenopir, C. & Clarke Michael (2006). **Measuring Total Readings of Journal Articles**. *D-Lib Magazine* Vol 12/10, October 2006. Retrieved 20 November 2006 from <http://www.dlib.org/dlib/october06/king/10king.html>.

Lewison, G. (2006). **Beyond SCI citations – New ways to evaluate research**, *Current Science*, Vol. 89, No. 9, 10 November, 2005, pp. 1524-1530, Retrieved 20 November 2006 from <http://www.ias.ac.in/currsci/nov102005/1524.pdf>.

NIST (1993), **Integration Definition For Function Modelling (IDEF0)** Draft Federal Information Processing Standards Publication 183, 1993 December 21, National Institute for Standards and Technology, Gaithersburg, Md, USA, Retrieved 20 November 2006 from <http://www.itl.nist.gov/fipspubs/by-num.htm>.

OAI (2006). **Open Archives Initiative**, webpages, Retrieved 20 November 2006 from <http://www.openarchives.org/>.

OECD (2004). **Declaration on Access to research Data from Public Funding**. Retrieved 20 November 2006 from <http://www.codataweb.org/UNESCOmtg/dryden-declaration.pdf#search=%22DECLARATION%20ON%20ACCESS%20TO%20RESEARCH%20DATA%20FROM%20PUBLIC%20FUNDING%22>.

Porter, Michael (1985). **Competitive Advantage - Creating and Sustaining Superior Performance**. The Free Press, New York.

Søndergaard, T. F.; Andersen, J.; Hjørland, B. (2003). **Documents and the communication of scientific and scholarly information – Revising and updating the UNISIST model**, *Journal of Documentation*, Vol. 59, No.3, 2003, pp. 278-320.

Sommerville, I. (1995). **Software Engineering**. Fifth Edition, Addison-Wesley, Harlow, UK.

Swisher, B. (2005). **Electronic Access to Social Science Research**, Course LIS 5703, Lecture notes, School of Library and information studies, University of Oklahoma, Retrieved 20 November 2006 from <http://www.ou.edu/ap/lis5703/sessions/s06.htm#Notes>.

Suber, P. (2006). **Open Access Overview**, Retrieved 20 November 2006 from <http://www.earlham.edu/~peters/fos/overview.htm>.

Tenopir, C. & King, D. (2000). **Towards electronic Journals – Realities for scientists, librarians and publishers**, Special Libraries Association, Washington D. C..

Turban, E. & Aronson, J. (1998). **Decision support systems and Intelligent Systems**. Fifth Edition, Prentice-Hall, Upper Saddle River, New Jersey.