The Research and HE Information Market
Hans E. Roosendaal

1 Vision

The research and HE information market will in future be based on a federated network of repositories of information relating to research and education that conform to open standards, and an accommodating architecture that allows users the easiest and fastest possible access to information in all of these repositories.
The information available by such a network will not only comprise of information material for research and HE, but also of management information relating to this information.
The market is the research and HE community; its main focus is open standards.
This federated network will be global.

This generally shared vision describes a real life network of repositories of information relating to research and education, containing both research and education information in the widest sense and management information to support access to and disclosure of this information. The user, be this a student, a teacher or a researcher, will be able to make use of this information from any site and in all possible ways.

2 General requirements

The aim is to make scholarly and educational content more widely available in a coherent way complementing the objectives and aims of the research grid and e–science initiatives.
Many research and HE institutions and other knowledge–intensive organisations and companies world wide are developing novel but often disparate approaches to the management of on–line scholarly and educational resources, a.o. by creating institutional repositories. Others will follow suit. It is necessary to ensure that users will have access to research and learning resources in a coherent manner within a common information environment. Thus researchers, teachers and students should be able to find and to access a wealth of on–line resources from across universities and publishers. Currently, access to quality assured resources is limited to a single institution or, at best, national or disciplinary repositories.

The vision in the research and HE information market is to create a federated digital network for repositories and libraries in the field of scholarly and educational information. These repositories can be institutional repositories containing information products of an institution's research and education, or disciplinary repositories for either research or educational information, or combinations thereof. To be widely acceptable such a network will have to be able to represent a comprehensive choice of different and disparate webs. The network will connect existing and future repositories and libraries.

To achieve wide acceptance of the network it is mandatory that it contains a sufficiently large critical mass of information material. Critical mass is also needed to be able to support a variety of value chains for the Information market. These different value chains can represent different organisational models, legal models and business models as the individual stakeholders see fit for the exchange of specific information products.
The network should be able to support a comprehensive variety of business models that can and will be agreed between the different stakeholders for different classes of information or different ways to access the content in this network of repositories. A main requirement of the network

then is that not any feasible business model now and in future should a priori be ruled out or should be principally impossible.

Next to the research and HE community the general public can benefit as well by being able to access research and education information through public institutions, such as libraries and schools. This will provide a substantial contribution to embedding science better in society. Also, the use of information in the research and HE community is rather intensive and this means that results achieved in this market will yield a wealth of spin−offs for other markets making use of complex information[1].

The creation of a cohesive and coherent network guarantees the best return on investment, in whatever form, for all stakeholders on their own terms, be they public (such as e.g. universities) or private (such as e.g. publishers) organisations. It is in the interest of each individual stakeholder to strive for maximum flexibility in the market place. This can best be achieved by developing a strategy including an organisational configuration that allows maximum compliance with the vision in the market place. Stakeholders should share a basic conception of a high level strategy as starting point for developing their individual strategies.

3   The high level strategy

Creating a network of repositories of information relating to research and education requires a basic conception between the stakeholders of a high level strategy. This strategy can only be successful if it comprises in the best possible way the major interests of all stakeholders. This means that such a strategy can only have one focus: the USER, as an author and a reader, as the generic stakeholder in the value chain and the primary beneficiary of the network. The USER is the learners, teachers, researchers and students in knowledge institutions and organisations. This means that a comprehensive approach to user behaviour and to the consequences of such behaviour for the value chain of information is desirable.
The value chain of information from the viewpoint of the USER can be summarised in the following comprehensive way[2]:

" Authors want to publish more, readers want to read less."

This means that wide exposure is paramount to the author and (pre)selection to the reader of scientific information. Any force like the use of Information and Communication Technology (ICT) by the actors involved (authors, readers, academic libraries, scientific publishers etc.) that allows better fulfilling this law is an engine for change of the value chain. And this will prompt changes in the roles of the stakeholders in scientific communication.

---

[1] Complex information is information originating from different sources and representing different forms of information each having its own terms, rules, standards and values. Complex information is then information integrated from these different sources. These sources can be the links in a value chain and complex information is then information integrated over the entire value chain. Complex information is therefore by definition strategic information for an organisation, be this a complex organisation like a knowledge−intensive organisation such as a university, an enterprise, an entire business or industry. Research and HE information is a prototype of complex information.

[2] Hans E. Roosendaal, Peter A. Th. M. Geurts and Paul E. van der Vet, "Developments in scientific communication: Considerations on the value chain". Information Services and Use, 21 (2001) 13−32

The institutional stakeholders in the research and HE information market, and beyond, will as enablers be the secondary beneficiaries. As stated before, the foremost goal for every stakeholder is to develop an individually tailored strategy to comply with the high level strategy and to place this stakeholder in this way at the forefront of developments in on–line information management. Only then the stakeholder will be able to make an invaluable contribution to a network for world wide information provision in research and education. A key aim of the strategy is making universities, other knowledge institutions and scientific publishers, non–commercial or commercial, professional by helping to make use of this network and ensuring that the architecture will best serve all stakeholders' needs.

The network should be able to support the USER in the following strategic tasks:
- **creation**, i.e. its technical aspects: how to assist the author to create the work;
- **accessing:** how to assist the reader to be able to access the desired work;
- **archiving:** how to assist the reader to find and retrieve the desired work;
- **quality control:** in technical terms and in terms of certification and accreditation of the content of the work;
- **disclosure:** how to assist the reader in selecting and filtering the information to obtain the desired work;

and
- **use of information:** what do authors and readers need to make most effective and efficient use of the information for their individual research and educational purposes.

The following strategic issues are then relevant:
- **organisational configuration:** at the market level there is a necessity for an organisational configuration supporting the development of this network while representing many stakeholders in the market of scholarly and educational information; the individual stakeholder level needs an organisational strategy supporting the high level strategy compliant with the market organisational configuration.
- **user requirements:** a minimum requirement is to obtain and maintain a comprehensive view of what the current and developing behaviour of academics and students as they search for scientific content in the various repositories, and translating these requirements into technological and organisational conditions for the network.
- **system integration development:** main issue here is the middleware[3] to connect existing and future repositories. The network will most probably require a new generation of middleware.
- **standardisation:** dynamic standards with the goal to support a common, interoperable, developable and federated infrastructure and to support the ability to integrate single modules of a workflow between the stakeholders into the general infrastructure are necessary. The strategy should encourage activities in standardisation agencies of open standards like OAi, W3C to develop a comprehensive and integrated range of standardisation–oriented services and products.

---

[3] See for further details also 5.2. The relationship between all stakeholders is given in the value chain. As stated before the network should be able to support a comprehensive variety of business models that can and will be agreed between the different stakeholders for different classes of information or different ways to access the content in this network of repositories. As a main requirement of the network we have stated that not any feasible business model now and in future should a priori be ruled out or should be principally impossible. It goes without saying that this requirement is also valid in connection with requirements on middleware.

- **reference models:** real scale reference models and best practices for the scholarly and educational communication and information system of sufficient critical mass are required to further progress and to lower risks.
- **policies and legal issues:** policies and legal arrangements between the different stakeholders in the scholarly and educational information domain need be further developed in a coherent way, i.e. on the basis of an adhered high level strategy.

This leads to the following conditions for a high level strategy:
- in−depth understanding of the ICT−requirements of users of scholarly and educational information as a basis for further development and integration;
- development of network solutions, including middleware and specific applications, to provide the essential glue for components and functions of the integrated research and education information infrastructure;
- pro−active participation of all stakeholders or of a stakeholder representation in global standardisation efforts in relevant areas including data formats, digital libraries, archival formats and digital preservation practices;
- the promulgation and exchange of reference models and best practices for research and educational information on topics including user ICT−requirements, middleware, data formats, and organisational frameworks;
- the creation of legal arrangements, e.g. in areas such as copyright and licensing, between stakeholders in the research and educational information domain

Even if some universities have already developed institutional repositories for information relating to research and education it is difficult or even next to impossible to make adequate use of the information by other research institutions with the exception of research material accessible via international publishers. This is, however, just a very small fraction of the material available both in research and teaching. A full exchange of this material leading to access to a comprehensive collection of digital content can only enhance the quality of research and teaching as a whole. Examples of possible use are e.g. making use of course material developed at one institution by students and teachers of another institution or accessing in a structured way research information created at one institution by another institution, be this direct from institution to institution or intermediated via a publisher. The essence of a network of repositories is that this access should be possible without relocating the information on intermediate repositories other than strictly necessary for technical reasons.

The above sketches a number of prerequisites for the open exchange of information relating to research and education, for the development of the open information society at the level of research and education as a major component of the knowledge society. A high level strategy as proposed is a necessary condition to bring the knowledge society within reach. This strategy will in turn only be successful if it can be implemented for a sufficient and critical mass of information and this requires the co−operation of all stakeholders.

4   Technology strategy

The strategy aims at creating a federated digital network for repositories and libraries in the field of scholarly and educational information. This can only be realised by connecting as many existing and future repositories, archives and libraries as possible. This requires architecture allowing sufficient integration of repositories. The network should be based on a shared architecture to support research and education that conforms to international open standards and that is responsive to changing contexts for scholarly and educational information use including e−learning and e−science.

Rather than being technology driven the strategy should preferably be driven by specific market needs and should encompass a comprehensive group of stakeholders as to ensure a strategy for the development of sufficient critical mass and required sustainability. This allows each individual stakeholder to contribute to the gradual creation of the network in a way that is fully compatible with his own strategy provided that this strategy is compliant with the high level strategy supporting the vision on the research and education information market. In this way technical activities as needed for clusters of real life projects demanded by the digital information market at large can be brought together. In doing so, stakeholders will be able to build on infrastructures already existing and to make use of state of the art technologies. At present, these infrastructures and technologies are highly fragmented.

This fragmentation in the digital information market place is the real barrier to the widespread use of digital information to the benefit of society as a whole, and in particular to the development of the information society. Only when the issue of fragmentation is adequately solved will society be able to grab all benefits of the open information society. Fragmentation is a real barrier for universities, research and HE institutions and other knowledge−intensive organisations, to become professional in the management of information, and in particular of information relating to research and education. Reducing fragmentation is needed to ensure integrated search possibilities for all content present in the entire network of repositories, be it data or metadata. A major priority for a high level strategy is to combat fragmentation in the market place while at the same time encompassing the individual stakeholders' needs. Adopting a USER focus in relation to the global digital information market place of scholarly and educational information seems the only way to combat this fragmentation originating from disparate approaches by individual stakeholders. The added value lies in providing access to comprehensive collections of digital content within 10 years to come and based on a clear focus on user behaviour and its consequences for the value chain of information[4].

There are many different widely used proprietary systems as well as open standards and platforms in the market place and there is a variety of needs of different stakeholders, either content driven or technology driven. It is in the interest of each individual stakeholder to comply with a high level and novel strategy that aims to harmonise the global digital information market place and in this way to strengthen this market place for the benefit of the USER and the other stakeholders.

---

[4] Hans E. Roosendaal, Theo W. C. Huibers, Peter A. Th. M. Geurts and Paul van der Vet, "Changes in the value chain of scientific information: economic consequences for academic institutions." Online Information Review, 27, 2 (2003) 120 − 128.

A useful technology approach is to make use of a methodology based on identifying the technological gaps in the existing infrastructures and in the existing state of the art technologies. In combination with the USER focus this will result in defining and implementing strategies for research in order to solve these gaps. In this way fragmentation should be reduced and system integration resulting in a fully communicating network of information relating to research and education can best be ensured.