Comparing consortial repositories: a model-driven analysis.

Josh Brown, SHERPA-LEAP Project Officer, University College London.
josh.brown@ucl.ac.uk

Abstract:

This study aims to provide a comparative assessment of different repository consortia as a reference to inform future work in the area. A review of the literature was used to identify repository consortia, and their features were compared. Three models of consortial repositories were derived from this comparison, based on their structure and aims. The consortial models were based around either: creating a shared repository for the members, developing a repository software platform or creating a metadata harvesting service to aggregate content. Using case studies of each type of repository consortium, each model was assessed in terms of its particular strengths and weaknesses. These strengths were then compared across the models to enable those considering a consortial repository project to assess which model, or combination of models, would best address their needs and to aid in project planning.

1. *Introduction*:

Institutional repositories are a relatively new and heterogeneous phenomenon, and have achieved considerable diversity in aims, structures and implementation in a short space of time. Many institutions have chosen a collaborative approach to repository provision and have taken part in repository consortia, defined here as *a collaborative venture in which participating institutions share the development and implementation of bespoke repository services.* Repository consortia have emerged in many nations, and have taken varied forms. This variety has made it difficult to assess the relative merits of the possible consortial approaches as, while individual repository consortia have presences in the literature and there has been some discussion of the benefits of a consortial approach[1], there is to date no significant attempt published to synthesise the experiences and expertise they represent.

The aim of this paper is to provide a comparative assessment of different repository consortia as a reference to inform future work in the area. In order to achieve this, the objectives are to:

- Classify repository consortia according to their distinctive features
- Identify exemplars of each type of consortial model
- Compare different exemplars of each model to identify its essential features

- Compare the contexts, strengths and weaknesses of the different models

2. Method:

Initially a literature review was conducted, and a number of repository consortia were identified. On the basis of this review, the original aims of repository consortia were compared and classified. This classification produced three distinct classes of repository consortium. These are:

- Shared: a single repository serving multpiple institutions

- Platform: customised repository software made available to particpating institutions

- Aggregator: a search service based on central harvesting of data from participating repositories

Within each model, exemplars were selected to serve as case studies for the basis of analysis. These exemplars were compared to one another in order to arrive at a clearer picture of the unique strengths and challenges of each model. The literature was supplemented by contacting managers or developers from each exemplar via email and/or telephone. A total of 7 consortia from 6 countries (Australia, Japan, the Netherlands, Norway, the UK and the USA) were used for case studies.

Each model was assessed as to its perceived strengths and weaknesses, based on the successes and experiences of the example consortia in achieving their original aims. This assessment yielded a typical set of circumstances in which each model tends to arise, and was then used as a basis for comparing the models to one another. The comparison took two forms. A simple ranking (strong to weak) was used to show in which areas a given consortia tended to perform well and which areas could require monitoring by potential consortium managers. For aspects of the consortia which were too broad for meaningful ranking, the strengths and concerns were discussed under a series of appropriate subheadings. The comparison of these sets of advantages also serves to highlight the key benefits that derive from a consortial approach to repository development, whichever model is adopted.

3. The consortial models:

A literature survey of the stated founding aims of each consortium yielded three distinct types of repository consortium. It is worth noting that these categories are not entirely fixed and that, by entering into new collaborations, or developing from one phase of the consortium project to another, the distinctions between models

can become blurred. However, there is some benefit in analysing these consortia in terms of their original aims as this gives a clear guide as to the suitability of each model to those seeking to assess their effectiveness, or considering embarking on a consortial repository project. The division into these models also provides the means of mapping the development of the consortia over time, which provides useful evidence for longer term or strategic planning. These issues are discussed in more detail in chapter 4.

3.1.   The 'Shared' model.

The shared repository model is one in which a group of institutions, generally without individual pre-existing repositories, come together to develop a single instance of a repository to house content from all participating institutions.

The exemplars used for the analysis of this model are the White Rose consortium in the UK, with their repository White Rose Research Online (WRRO), and the Washington Research Library Consortium (WRLC) in the USA, with their repository ALADIN Research Commons (ALADINRC).

For the purposes of straightforward comparison, the key details of each consortium are presented in table 1 (below).

	White Rose	ALADINRC
Dates	2004-present	2004-present
Original Aim	To build on existing collaborations; to exploit economies of scale; to aid regional research collaboration; to leverage the White Rose 'brand'; to pool resources to create dedicated staffing.[2] [3]	"To provide institutional repository services to their academic communities" as well as "a means for providing access to each institution's scholarly output and a means for making each institution's contributions accessible to all member institutions"[4]
Funding source(s)	Start-up funding from JISC and CURL, via the SHERPA project, ongoing funding from the consortium with additional funding from the JISC for specific projects.	The consortium
Original partners	White Rose University Consortium (The Universities of Leeds, Sheffield and York).	The Washington Research Libraries Consortium (American University, Catholic University of America, University of the District of Columbia, Gallaudet university, George Mason University, George Washington

		University, Georgetown University and Marymount University)
Platform	EPrints	DSpace
Developer	The consortium	The consortium

3.1.1. White Rose.

White Rose Research Online is one of the larger institutional repositories in the UK, both in terms of the volume of research it contains[5] and in terms of the size of the research community it serves (see below). The consortium has moved from a goodwill basis to a formal memorandum of understanding which secures the repository for the immediate future and indicates an ongoing commitment from the partners.

Successes:

The consortium has successfully met its original aims, achieving economies of scale, both in terms of capitalising on the combined research power of the three institutions (which ranks alongside that of Oxford and Cambridge) and enabling pooled resources for staffing and equipment.[6]

Information exchange and the central management of policies has worked well (the divergence of policy and workflow around e-theses has led to some degree of devolution, but this has extended the model, rather than undermined it)[7] and the development of overall strategies for repository development has been generally smooth[8]. A majority of researchers "seem to be quite happy with the shared model"[9]

Having a central team of dedicated repository staff has worked well, and a majority of repository queries and issues go through the team. It has been noted that a "highly centralised service does not scale well without significant additional resource"[10], and in response to this pressure, some responsibility for work has been devolved back to 'local' staff.

Challenges:

The repository is not well embedded in any of the three partners[11] for a number of reasons. WRRO deposit is not a normal part of the workflow for White Rose researchers and the repository has been described as a 'silo'[12]. This results from the fact that it is not well integrated with other IT systems at the partner institutions. This restricts the possibilities of automated 'capture' of research, has made the development of local login credentials very problematic and has been further complicated by the adoption of different research systems at the partners[13]. The sheer number of systems in use at the three institutions can mean that it can be a

challenge to identify solutions for particular IT issues and to influence planning at a local level.[14]

This technical complexity is mirrored culturally, in terms of the scale and diversity of activity in the three partners. Raising and sustaining awareness in such a varied community has proved difficult, and to some extent the repository remains outside of the 'library mainstream'[15]. This could be organisational and be due in part to the lack of local branding and the centralised nature of repository staffing or it could be generic to the field and follow the patterns of awareness of repositories more generally, as some subject librarians are more engaged with repository issues than others and so collaborate in advocacy to a greater extent.

This complexity poses a barrier to the expansion of the consortium, even where that could be perceived as appropriate or beneficial and emphasises the importance of the sensitive work undertaken by the repository team to strengthen local ties to the repository.[16]

3.1.2. ALADINRC.

ALADIN Research Commons has a greater number of participant institutions, seven to WRRO's three, but a lower number of records which may be due to the fact that deposit has mostly been on an ad hoc basis, owing to the absence of dedicated staff positions for marketing the repository across the various campuses.[17] On the plus side, the local nature of content recruitment drives means that they are uncontroversial and targeted.

Successes:

The consortial approach, and the use of a shared repository, has brought a number of benefits to the consortium members. Funding is relatively secure, and the consortial activities, including the repository have generally enjoyed strong support from the member universities.[18] The consortial approach has provided an opportunity for institutions "which might not otherwise have the resources to have an institutional repository"[19]. The fact that the consortium has handled the hosting and development has also saved member libraries having to compete with other units for campus IT resources.

The initial implementation and development process has been described as smooth, possibly owing to the lengthy experience of the consortium in coordinating projects.[20] The central maintenance of the repository also provides a stable basis for metadata standardisation, preservation control and a focus for the resolution of any

problems.[21] It also brings economies of scale, as the central team can handle most technical issues without duplicating time and effort across institutions.[22]

Challenges:

It took a lot of development work to enable branding for individual institutions to be built into the repository.[23] While the ability to set local policies preserves institutional individuality within the consortium, it also leads to a risk of divergence on key issues. Within ALADINRC, areas of divergence included naming conventions for collections to link to institutions and departments[24], definitions of acceptable content and legal concerns and opinions.[25] In order to prevent such concerns becoming a barrier to the effective working of the repository, it was decided to set a central set of policies that were general enough to allow for local variations

Further challenges have arisen around 'ownership' of the repository. The perceived level of separation from each institution, due to the lack of resources at institutional level, has slowed progress in populating the repository, and "at some levels having access to the repository without having had to make a significant investment of money and staff time has led to a perception of the repository as a peripheral service"[26]

3.1.3. Discussion.

For both of the exemplars discussed here, a pre-existing consortium was essential to the smooth planning and execution of the development. The ongoing managerial and financial links required between institutions in order to run the service in a mutually beneficial way seem to fit best within a framework of ongoing collaboration.

One key distinction between the two exemplars is in the structure of staffing. WRRO is operated by a central team, which liaises with the partners. ALADINRC has a central technical team, but other activities (content recruitment, metadata creation, copyright checking and the like) are managed locally. The WRRO approach maximises economies of scale, and ensures rigorous consistency in standards, whereas the ALADINRC offers devolved policy making in key areas and allows a degree of individuality to partner institutions. In and of itself, this must pose some risks as the challenge is to allow for individuality whilst maintaining sufficiently prescriptive standards to preserve consortium-wide (and external) consistency. In terms of staffing, the WRRO approach is less flexible and easy to scale up, which is why some areas of workflow are becoming more devolved.

The key strengths of the shared model are:

- Creates ongoing sustainable service
- Economies of scale
- Pooling of expertise in central team
- One team to handle detailed queries from all institutions
- Central management of policies leads to strong handling of metadata and preservation issues
- Opens up possibilities to institutions that may not have had a repository on their own
- Increased volume of content can lead to greater visibility for the whole.

The key drawbacks of the shared model are:

- Complexity: multiple systems across multiple institutions
- Complexity: additional inter- and intra-institutional cultural and political issues
- Requires sensitive handling of cross-institutional policies
- 'Distant' from individual institutions – can lead to low commitment and a diluted sense of ownership
- With scale comes inertia.

3.2. The 'Platform' model.

The Platform model is one in which a group of institutions, again usually without pre-existing repositories, collaboratively develop a customised, customisable version of a software platform which is then used as the basis for repositories at each participant institution.

The exemplars used for the analysis of this model are SHERPA-LEAP (the London E-prints Access Project, affiliated to SHERPA) in the UK, the Australian Research Repositories Online to the World (ARROW) consortium, and the Project for Electronic Publications and Institutional Archives (PEPIA), based in Norway.

Table 2 (below) gives the essential features of each consortium.

	SHERPA-LEAP	ARROW	PEPIA
Dates	2004-present	2003-2008	2005-2007
Original Aim	"To create eprints repositories, hosted centrally by UCL (University College London) for each of the partner institutions and to	"Identify and test software and solutions to support best practice institutional digital repositories"[28], defined as "generic storage mechanisms	"To create institutional archives for the participating institutions"[30]

	populate those repositories through collective advocacy"[27]	for a combination of any type of digital content that might arise from the research efforts of Australian universities"[29].	
Funding source(s)	University of London (four separate grants, from 2004-2010, plus occasional external funding for special projects).	Australian Commonwealth Department of Education, Science and Training (two main grants, including funding for additional 'mini-projects')	50% from the Norwegian Archive, Library and Museum Authority, remainder split between the participating institutions[31]
Original partners	Birkbeck, Imperial College London, King's College London, London School of Economics and Political Science, Royal Holloway, School of Oriental and African Studies, University College London (Lead partner).	Monash University (Lead partner), Swinburne University of Technology, University of New South Wales, National Library of Australia.	BIBSYS (developer), Norwegian Archive, Library and Museum Authority and twenty-seven university, college and other libraries.
Platform	EPrints	FEDORA (Adapted)	DSpace (Adapted)
Developer	The consortium	VTLS	BIBSYS

3.2.1. SHERPA-LEAP.

SHERPA-LEAP began within a group of universities that are related not just geographically, all being based in or around London, but organisationally, in that they all formed part of the federal University of London. This remains largely the case with the membership today, although one institution is now an Associate Partner, having decided to leave the University of London. Collaboration is still close, however, and the community of 13 repositories remains cohesive.

Successes:

The original technical model for the consortium was a centrally hosted, single instance of the EPrints platform, with individual configurations for each of the partners. This was designed to allow individual institutions to take their own technical and policy decisions and to create their own branding. Diversity within the consortium was therefore supported, with an emphasis within the community on sharing best practice and advice.[32] At the same time, central hosting allowed the

consortium to roll out multiple repositories in a short space of time, with significantly reduced resource overheads.[33]

The availability of central guidance on the broad range of repository issues faced by partners in the consortium, going beyond the software to metadata, copyright and advocacy, was crucial to those institutions that may not have had readily available in-house expertise at the beginning of the project.[34] This, together with the central funding provided by the Vice Chancellor of the University of London, meant that smaller institutions, for whom a repository was not necessarily high on the agenda or easy to resource, were able to develop repositories, and meant that the larger institutions were offered the means to explore their detailed requirements for repository provision in a real-world, practical environment.[35]

The creation of a closely networked community was also essential to the success of the project. This worked on every level of the project, as each partner institution had both a managerial presence on the steering group and a 'field officer' who participated in the emerging community of practice. It was observed that "the facilitation of opportunities for mutual support is probably the day-to-day aspect of SHERPA-LEAP most valued by its members and a key strength of the consortial approach."[36]

Challenges:

There were significant challenges faced during the initial, setup phases of the project, as there was limited technical support for repository configuration and customisation available from the centre. In this environment, some institutions struggled at times to maintain progress.[37] While this situation was remedied, especially during the second phase of the project in which the membership was expanded and a full-time Project Officer appointed, the commitment to supporting diversity within the consortium initially placed the onus squarely on individual institutions to customise their own repository.[38] The technical architecture of the platform also posed some difficulties at times, as one shared copy of the platform software occasionally led to shared technical problems, an issue that became more significant as the repositories became more individualised. It was realised, with hindsight, that better isolation for constituent repositories would have been an advantage[39], and this was factored in to later stages of the project.

The pursuit of sustainability for the consortium has been a factor from its early days[40]. Once the initial project phase was complete, a second phase, based on expanding the consortium and its community of practice, ensured that the momentum of the consortium was maintained. A third project phase, based largely around developing new services for the members, specifically a search aggregator[41], was a success and the fourth project phase, which aims to build on the community of

repositories and the network of experience sharing, is now under way. The consortium has managed to sustain itself since its inception through discretely-funded expansion and development projects, but the current tranche of funding will end in 2010, and central continuation funding is unlikely to be available. SHERPA-LEAP will be devising an exit strategy in the coming months; the extent and scope of consortial activity will be re-shaped accordingly.

3.2.2. ARROW.

The ARROW consortium was created to develop a software platform for the use of Australian institutions. It took FEDORA and developed it in collaboration with a private company, based in the US (VTLS). This led to the development within the consortium of the repository platform VITAL which has now been adopted at 14 Australian institutions.[42]

Successes:

Partnership with a private software development company brought several distinct benefits to the consortium, in that the consortium did not need to recruit their own developers, the company had already done a great deal of ground work and they brought knowledge and experience to the project that would have had to have been created within the consortium otherwise.[43] The partnership also provided an exit strategy for the consortium upon project completion, in that the company is a persistent entity and provided the institutions continue to subscribe, they receive ongoing software and technical support.[44]

The consortial approach was described as 'a huge help' bringing 'many brains' to the project[45] and creating a space for sharing issues and experiences, development work and providing multiple perspectives on issues of note.[46] This information sharing was managed using regular, frequent technical meetings and bi-monthly steering committee meetings, using a combination of teleconferencing and face-to-face encounters. The community also used a wiki and mailing list to maintain quick and constant communication and there were weekly teleconferences with the developers in the US.[47] This network was particularly useful as the range of experience within the consortium led to a degree of specialisation, with individual partners taking on majority responsibility for one aspect of the work, and so informing the whole group about progress and issues helped keep the workflow under control.[48]

Challenges:

Sometimes the 'many brains' were 'too many', and this could disrupt focus and led to occasional scope creep in the attempt to balance differing demands and priorities. Again, the high level of communication within the consortium was crucial in

addressing this when it happened.[49] The network was also vital in the difficult task of managing the expectations of a large consortium, and this was a challenge that took up some time.[50]

Managing expectations also placed project managers under occasional pressure, as they sought to balance the differing expectations and practices of the private sector developers and the public sector consortium. The priorities of the partners didn't always coincide, and the geographical distance between the US and Australia made effective communication difficult at times.[51] It was also felt that teleconferencing was of limited use, and that the regime of meeting face-to-face with the developers every six months was insufficient to avoid misinterpretations or misunderstandings creeping into the process, which tended to emerge after the fact.[52] Within the Australian partners, it took time-consuming discussion and decision making to evolve standards around issues such as copyright, advocacy, metadata as well as the purely technical features of the repository development process.[53]

3.2.3. *PEPIA*.

PEPIA was based around the development of a customisable, scalable repository platform, Brage, which built on existing, open source software and was developed with the specifics of the Norwegian research environment in mind. A key developmental emphasis was integration with research reporting mechanisms and the Norwegian national harvester, Norwegian Open Research Archives (NORA).[54]

Successes:

One factor that benefitted the integrative aims of the project was the involvement of BIBSYS, the developer. BIBSYS is owned by the Norwegian government and is specifically aimed at serving higher education and research institutions.[55] It occupied a unique position as a collaborator, insofar as it is a non-profit organisation with 30 years experience of working with Norwegian libraries, a factor which aided communication between the development team and the partners.[56] [57]

There were a number of important factors that influenced take up of the repository platform developed. Brage was designed as a stand-alone offering, which ensured that even those institutions that did not have an existing relationship with BIBSYS were able to choose to take part[58]. The cost of the development and implementation process was heavily subsidised by the Norwegian government which meant that the cost of participation in the consortium was kept low[59] and the prices charged for participation were based on the size of the institution, using a formula based on the number of researchers at each. This meant that for smaller institutions, the costs

were correspondingly smaller, a factor which made Brage a decidedly attractive offering to smaller research institutes which may not otherwise have been able to afford a repository of their own[60].

The selection of DSpace as the basis for the Brage platform brought significant benefits for the developers. The fact that it was an established software offering with a large user community aided the development process[61], as did the fact that DSpace was compatible with the existing BIBSYS server platform, which avoided the need to acquire additional competencies at short notice[62].

With regard to sustainability, the consortium benefitted from the presence of an independent company to maintain the product when the project came to an end and the presence of NORA meant that the individual repositories did not start from zero, either in terms of visibility or an active community[63].

Challenges:

The PEPIA development team faced several challenges in the customisation of DSpace. Expanding exisiting software that was essentially designed for a single instance in a single institution was described as 'a technical hill to climb'[64], a situation that was not helped by the fact that DSpace is large and complex. This meant that, although the software was basically compatible with the existing systems, the developers were on a steep learning curve in understanding DSpace even before they came to adjust it to their preferences[65].

The fact that the open source software took some time to tailor, in combination with unpredictable progress along the learning curve and the ad hoc nature of user group support, made it difficult to estimate resource requirements for the project accurately[66]. Developing user management and authentication mechanisms for the repositories running Brage, which was essential for the integration with the national research reporting systems as well as with local, institutional systems proved to be extremely challenging and took up a lot of development time and resources[67]. There were also challenges for the integration with NORA, as problems with OAI-PMH required additional development work outside of the lifespan of the project[68].

3.2.4. Discussion.

The Platform model tends to be adopted by groups of institutions without repositories, but with defined sets of needs and expectations and a clear repository brief. The exemplars represent a mixture of public and private sector collaborations, which afforded a range of technical opportunities and exit strategies to the consortia. The differing nature of these collaborations provides some insight into the breadth of possibilities for such a consortium, and an idea of the challenges that these options can bring.

A common area of concern was the development from a core system designed with a single instance at a single institution in mind to one with multiple iterations and multiple interactions. Both SHERPA-LEAP and PEPIA reported challenges with this process. ARROW seems to have avoided some of these problems by sourcing key aspects of the development process from within its member institutions, a tactic which seems to have required considerable effort to manage, but which has lead to a more detailed understanding of local variations during the planning and development stages of the project.

Some of the key strengths of the platform model are:

- Institutional commitment is clearly defined by project span and tranche of funding.
- Reduced development costs
- Creates a defined community of practice and expertise sharing network
- Specifically goal driven – can avoid some of the challenges and complexity of longer term projects
- Open to expansion with minimal extra cost
- Keeps repository local with staff presence at each institution
- Ample opportunity for customisation
- Platform can be tailored to local conditions and needs
- Can be nimble

Some of its key weaknesses are:

- Tends to be time limited: consortium must grow, develop or die – requires exit strategy as part of planning
- Requires duplication of effort compared to other models
- Requires technical expertise within member institutions
- Relatively greater co-ordination of multiple repositories/partners can be problematic.

3.3. The 'Aggregator' model.

In the Aggregator model institutions, either with or without pre-existing repositories (for those without, the development of a repository is a necessary condition of involvement in the consortium), come together to create a search aggregator to harvest content from their repositories and present it, via a single search interface to maximise the impact of the consortial presence. The exemplars used, the Dutch Digital Academic Repositories (DARE) consortium and the Japanese Institutional Repositories Online (JAIRO) have drawn comment for their similarities before within

the Japanese repository community[69], a fact which reinforces their selection as examples of this consortial model.

	DARE	JAIRO
Dates	2004-2006	2008-2009
Original Aim	"Implementing the basic infrastructure by setting up and linking the repositories; stimulating the development of services based on the research information made available through the infrastructure" especially focussing on the DAREnet search portal.[70]	To present an international unified search interface for Japanese Institutional Repository content.[71]
Funding source(s)	SURF Foundation.	National Institute for Informatics (NII) Cyber Science Infrastructure (CSI) program.
Original partners	SURF Foundation, Koninklijke Bibliotheek (KB), Royal Netherlands Academy of Arts and Sciences (KNAW), Netherlands Organisation for Scientific Research (NWO) and 13 University Libraries.	NII
Platform	Various	Various
Developer	The consortium.	NII

3.3.1. *DARE*.

The DARE project led directly to the adoption or harvesting of repositories at 13 Dutch universities, and fed into the NARCIS aggregator[72]. The consortium did not specify any repository software for its members, a decision that enabled participants to maintain their independence and avoided the need for consultation and negotiation on which platform to adopt[73].

Successes:

The creation of the DARENet harvester constituted a clear success for the consortium, although participants counted the level of mutual agreement and shared problem-solving as a major contributor to that success[74]. This was managed

in the DARE community by a focus on communal goals, and the setting of milestones within the project, with celebrations once they were achieved. The achievement of these goals was supported by regular meetings and workshops and a coordinated program of knowledge sharing run by a community manager[75].

The visibility of the program and the focus on the community has also helped to embed the role of repository managers in their organisations[76]. The importance of local staff has been helped by the fact that each institution controlled their own data collection, the repository interface and their service offering to their community and the web[77]. At the same time, institutions were strongly linked to the DARE consortium via their repository staff and the DARE community manager, and the sense of 'ownership' was further strengthened by the use of the DARE brand on numerous local-level activities[78].

Challenges:

Technically, the variety of repository platforms meant that harvesting was not always as straightforward as was hoped. Institutions with platforms in common tended to have to work together to solve problems as they arose, which meant that the national program, rather than having a single set of national problems, had four or more local variants[79].

The fact that this was a consortial program meant that it was inevitably distant from individual researchers, who could have benefitted from more contact with the project[80]. On the other hand, the active members of the community experienced some degree of information overload, and particular working groups experienced 'meeting overload'. The milestones, which were intended to provide measurable successes and to help promote the progress of the project, created deadlines, which were often experienced as a nuisance[81].

For the community, it was found that the knowledge sharing that was central to the project relied on the community manager being the most active member of the channels of communication. The community manager also had to shoulder some of the burden of reporting local information as well. This created a sense of dependence despite the willingness of members to share their knowledge[82]. The limits of the 'facilitator' role adopted by the SURFfoundation also needed to be redefined occasionally when the community looked the foundation for answers in order to keep the activity local, and to encourage the creation of solutions within the community.[83]

3.3.2.   *JAIRO*.

JAIRO is one strand of the Japanese Cyber Science Infrastructure (CSI)[84] program, which has achieved remarkable success in expanding the number of Japanese institutional repositories from 13 in 2006 to 110 as of October 2009[85]. It operates as a part of a nexus of programs and projects aimed at developing Japan's institutional repository network and at creating new services and interactions between them. JAIRO is "a service in which academic information accumulated in Japanese institutional repositories can be searched... cross-sectionally"[86] The use of JAIRO for the purposes of this comparison is of particular interest, given the general underrepresentation of Japanese repository developments in the English-language literature.[87] [88]

Successes:

The scale of material covered reflects the rapid expansion of repositories in Japan, and the national scope of the project has enhanced the sense of ownership of the JAIRO service amongst the fledgling repository community, where the sense of progress and 'problems shared' has greatly fuelled the development of institutional repositories.[89] The fact that the JAIRO service ties in with a group of initiatives has greatly helped with its development. For instance, the concerted and shared development process has led to the junii2 metadata standard being adopted in repositories across the country, which has simplified harvesting across the ten repository platforms currently deployed in Japan.[90]

The national level of coordination has won government level support for the aggregator service from the Ministry of Education, Culture, Sports, Science and Technology which has further reinforced compliance with JAIRO's standards and energised the collaboration.[91] Similarly to DARE, one of the main strengths of the project has been the development of the repository community, represented by the Digital Repository Federation (DRF).[92] As well as supporting the work being done by and for the repository community, the DRF develops international links and publicises JAIRO abroad.

Challenges:

There are a number of challenges facing the JAIRO consortium. Although the metadata harvesting has been a predominantly smooth process, there have been local costs incurred in achieving compliance with the junii2 standards, which may in turn need further central development to address the emergence of what have been described as 'local dialects'.[93] A study of the JAIRO aggregator highlighted a "concern whether similar policies were being used at each of the institutions"[94], a finding which substantiates this concern. Other potential threats include the absence of funding for the DRF. In order to continue its work and enhance the offerings of JAIRO

and the CSI by drawing together a community, this must be made independently sustainable.[95]

The JAIRO initiative suffers to some extent from its distance from researchers. While it has impressive coverage and scope, the DRF identified key areas where additional work is needed, some of which are due to the extremely rapid expansion of the Japanese repository community, for which JAIRO has been one of the major drivers. There is a need for content recruitment, improved quality control of contents and a dearth of trained repository staff.[96] The suggestion for a national training course to ensure that repository tasks become part of the library skill set is designed to address these issues, and will also help to address the problem of 'local dialects' mentioned above.

3.3.3. Discussion.

The aggregator model offers several advantages in that it stimulates and provides a focus for repository growth, and it tends to arise in relatively heterogeneous repository ecologies, with a mixture of established, new and planned repositories and platforms. It can present fewer immediate technical challenges to its members, although the co-ordination of standards requires an active and committed program of community action.

Some of the key strengths of the aggregator model are:

- Creates ongoing sustainable service
- Offers vastly increased visibility for researchers
- Boosts site traffic
- Creates a 'go-to' site for researchers seeking content
- Offers single solution for other services looking to harvest metadata etc
- Minimal involvement (and strain) for individual repositories beyond data level if they so wish
- Can be nimble

Some of its key weaknesses are:

- Relies on a community of goodwill
- Multiplies opportunities for standard clash/mismatch
- Remote from institutions and researchers
- Requires relatively open ended funding

4. Comparing consortia.

Having compared exemplars within each model and identified the key features of the models used for this analysis, two tasks of analysis remain. These are to assess the models used in light of the findings of the case studies, and then to compare the models to one another. The comparison takes two forms, a matrix presenting the key strengths of the consortial models alongside one another, and an extended discussion broken down by topic.

4.1. The models.

The models discussed here display certain key characteristics which mark them out as distinct from each other in terms of the repository ecology in which they arise and that they create, as well as in terms of their essential aims. Given that they operate in the same basic field of practice, they naturally share many common characteristics with non-consortial repositories, in that they serve a defined research community, they share common concerns with copyright, advocacy and preservation and they are a relatively new phenomenon.

The fact that they are all consortial endeavours gives them differing advantages and disadvantages to non-consortial repository projects which can be of value to those seeking to assess or create a consortium. The fact that they display differing aims, practical structures and technical challenges is also informative. The shared repositories described here are an 'end in themselves', and while they present certain unique challenges to partner institutions, are essentially designed as a single, ongoing, collaborative service. Platform consortia, on the other hand, are a shorter lived phenomenon, designed to rapidly increase repository provision within a defined community whilst maintaining individuality and independence at the partners. Aggregator consortia share the end with shared consortia of creating a single and enduring service, but are not involved in the architecture of individual repositories beyond the essentials of harvesting data.

The relationship between the platform and aggregator models can seem theoretically fraught, as all the exemplars used have a relationship to an aggregator. SHERPA-LEAP developed a search aggregator for the University of London community in its third phase, ARROW similarly went on to develop an aggregation service, and PEPIA repositories were set up with an eye to harvesting by the Norwegian Open Research Archives (NORA) service. However, there is a clear difference in terms of their original aims and practical focus which justifies the distinction.

The aggregator projects all dealt with multiple repository platforms (4 initially in the case of DARE and 10 in the case of JAIRO as discussed above) and were concerned

solely with harvesting from the outset. The stimulus to repository growth and population afforded by such a project is part of the rationale for its adoption, but establishing and configuring individual repositories is not part of the remit of this type of consortium. The platform consortia, on the other hand, were concerned with exactly this issue from the outset. In the case of PEPIA, NORA harvesting was part of the repository ecology in which the project took place. For SHERPA-LEAP and ARROW, aggregator services came later, as a consequence of the original goals having largely been achieved. In this context, the move to developing an aggregator is a means of adding value to the repositories that have been established, and, especially in the case of SHERPA-LEAP, maintaining consortial momentum and focus.

While the shared model is clearly distinguished by its organisational and technical structure, what distinguishes the platform and aggregator models is their place in the repository ecology and the locus of their main development work. The analysis of their strengths and weaknesses in specific facets of their work shows clear differences in the kind of activity they represent.

4.2. Comparing strengths.

Table 4 (below) gives an at-a-glance overview of the relative strengths of each consortium, as identified from the case study analyses in chapter 3.

Consortial strength		Consortial models		
		Shared	Platform	Aggregator
General	Reduces development costs	●	●	
	Can be nimble		●	●
	Easy to expand membership/services		●	●
	Reduces duplication of effort	●		
	Boosts repository uptake	●	●	○
	Increases visibility of content	○		●
Cultural	Creates community of practice	○	●	●
	Maintains local presence	○	●	●
	Creates pooled expertise	●	○	
Policy	Creates clear policy guidance	●		

Technical	Supports standards development	●		○
	Ease of customisation		●	N/A
	Low cross-partner coordination		○	●
	Reduced software complexity		○	●
	Integration with local systems	○	●	N/A
	Creates technical support service	●	○	
LEGEND: ● = STRONGER ○ =GOOD =WEAKER				

4.3.   Organisation and scale.

As noted above, the shared and aggregator models are structured around providing a centralised, ongoing service whilst the platform models tend towards a temporary 'federation'. The shared model tends to be constrained in terms of both scale and growth, however, as expansion tends to be practically and politically challenging, and new entrants would have to bring a substantial contribution to the consortium to justify the challenges.

Platform consortia can expand as long as institutions are keen to use the platform, and in the case of the ARROW and PEPIA products (VITAL and BRAGE respectively) work with the companies maintaining them. SHERPA-LEAP is constrained by its initial remit to serve the federal University of London, but could potentially grow to the size of the possible user community (in this case some 20 institutions).

For aggregators, the potential scale is global, although most aggregators so far stop at the level of their defined community (be it organisational or geographical). Adding institutions to the harvesting system is relatively straightforward once the infrastructure is set up, and the expansion adds utility to the search service.

4.4.   Sustainability.

A degree of concern around issues of sustainability is inherent in any repository project, owing to the relative newness of the field, but the consortial models described here each present different types of concern. Platform consortia must grow, develop new services or consider their aims achieved and develop an exit strategy. Of the three cases presented here, two have done so, ceding software to the development companies and other services to national bodies, while SHERPA-LEAP continues to operate, with a focus on services for the consortium community

rather than on the repository platform. If no further funding is forthcoming for the consortium, then an exit strategy will be created by the partners.

Shared and Aggregator models, as ongoing services have naturally greater stability, but are as vulnerable to changes in technology or culture as any other members of the new generation of digital services. However, in addition to this, the shared model is contingent on stability within the consortium, whereas the scale of aggregator models tends to render them vulnerable to shifts in national policy rather than to loss or change of membership.

4.5. Relationships with partner institutions.

As noted above, the shared model can face challenges navigating the complexity of local politics and technical systems. It also creates a more 'remote' presence, operating 'between' institutions, rather than being seen as part of any one. The aggregator consortia share this distance from individual institutions, which poses a challenge for the advocacy of the service, but have certain features which assist in addressing this issue, namely the scale of the consortia gives them a visibility which helps advocacy and the sense of ownership, and the systems depend on local repositories which means that there are staff at each institution already working on advocacy and relaying information between the consortium and institutions.

The platform model, which again relies on a local presence in each institution presents the least distance from researchers, and offers the nimblest option for reacting to the needs of differing research communities. The opportunity for collective advocacy work is also a useful feature for publicising the consortial agenda.

In all cases, the consortial models offer a stronger voice for repositories within their institutions and for external marketing.

4.6. Community

The communities created by consortia are cited in each case as one of the main strengths of the approach. In shared consortia, the repository project tends to add another service to the consortial offering, cementing its place within the partner institutions and creating new opportunities for co-operation. The platform and aggregator models generate new communities of practice, which are in every case crucial to the success of the project. In the case of the platform model, the level of community involvement is extremely high owing to the development, configuration and rollout processes, and this can lead to other synergies within the consortium (see, for example, the ongoing community projects within SHERPA-LEAP). The ongoing nature of aggregator projects means that the focus for community development remains longer term than can be the case with platform consortia, but

even in those consortia which have finished (see DARE for instance) the community has remained active, and has been exploited for other projects in the field.

4.7.   Economies of scale.

The presence of significant economies of scale stands alongside that of community-building as one of the greatest benefits of a consortial approach. Shared repositories vastly reduce the resource burden for partner institutions whilst pooled resources maximise the benefit of investment in staff and resources. The reduction in development costs across platform consortia is substantial, and studies have described the aggregator model as the cheapest means for creating a national repository infrastructure.[97]

5.   Conclusion.

It is clear from this comparison that repository consortia offer significant benefits to those institutions that have the opportunity to participate in them. Divergent consortial models have emerged from differing circumstances, and each model poses its own risks and challenges, but overall the consortial approach enables rapid expansion of repository provision at a reduced cost, builds communities of practice, improves the sustainability of individual services and maximises the exposure of repository content.

The differences identified here between the consortial models show that one model does not suit all, and that careful assessment of a range of factors, including the local and national 'repository ecology', the number of partners, the potential for collaborations with private or mixed sector companies, the scale and duration of commitment to the consortium and the experiences of previous consortial repositories should inform the choice of consortial model. The aim of this comparison is to facilitate these considerations, and offer guidance for those assessing the means for each institution to gain the most from a consortial repository project.

References:

[1] Moyle, Martin and Proudfoot, Rachel. (2009) Consortial routes to effective repositories. RSP Briefing Papers: Senior Managers . Repositories Support Project, UK. (In press) Available at: http://eprints.ucl.ac.uk/18775/ [last accessed 1/12/09]

[2] Proudfoot, Rachel. (2005) The White Rose Consortium ePrints Repository: creating a shared institutional repository for the Universities of Leeds, Sheffield and York. ALISS Quarterly, 1(1), pp19-23. P20.

[3] Proudfoot, Rachel. (2008) White Rose Research Online: A Case History. Repositories Support Project case studies [online] available at: http://www.rsp.ac.uk/repos/casestudies/pdfs/leedsyorksheffield.pdf [last accessed 05/10/09]

[4] Hulse, Bruce; Cheverie, Joan F. And Dygert, Claire T. (2007) ALADIN Research Commons: a consortial institutional repository. OCLC Systems and Services, 23(2), pp158-169. P159.

[5] Registry of Open Access Repositories (2009) Available at: http://roar.eprints.org/index.php [last accessed 12/10/09]

[6] Proudfoot, Rachel. (2005) Op. Cit. P20.

[7] Proudfoot, Rachel, "RE: Meeting and Research" email to the author, 20/07/2009.

[8] Proudfoot, Rachel. (2005). Op. Cit. P21.

[9] Ibid, p22.

[10] Proudfoot, Rachel. (2009) Op. Cit.

[11] Ibid.

[12] Ibid.

[13] Ibid.

[14] Ibid.

[15] Ibid.

[16] Proudfoot, Rachel. (2005) and (2008) Op Cit.

[17] Hulse, Bruce. "Re: Aladin Research Commons", email to the author. 05/08/2009.

[18] Ibid.

[19] Ibid.

[20] Ibid.

[21] Hulse, Bruce. (2007) P163.

[22] Ibid, p164.

[23] Ibid.

[24] Ibid, p165.

[25] Ibid, p166.

[26] Hulse, (2009), Op. Cit.

[27] Moyle, Martin; Stockley, Rebecca and Tonkin, Suzanne. (2007) SHERPA-LEAP: a consortial model for the creation and support of academic institutional repositories. OCLC Systems and Services, 23(2), pp125-132. P125.

[28] Groenewegen, David and Treloar, Andrew. (2008) The ARROW project: a consortial institutional repository solution, combining open source and proprietary software. OCLC Systems and Services, 24(1), pp 30-9. P31.

[29] Treloar, A., Payne, G. (2006), "The ARROW project after 2 years: are we hitting our targets?", VALA 2006: Connecting With Users, 8-10 February, Melbourne .

[30] Røyneberg, Ellen. (2007) Electronic publishing and institutional archives: utilising open-source software. BiD: textos universitaris de biblioteconomia i documentació. 19. P1

[31] Joki, Sverre Magnus Elvenes. (2007) PEPIA: a Norwegian collaborative effort for institutional repositories. OCLC Systems and Services: international digital library perspectives, 23(2), pp 204-209. P205.

[32] Moyle, Martin; Stockley, Rebecca and Tonkin, Suzanne. Op. Cit. P128.

[33] Ibid. P129.

[34] Ibid. P126.

[35] Ibid. P129.

[36] Ibid.

[37] Ibid. P130.

[38] Ibid. P128.

[39] Ibid. P130.

[40] Ibid. P131.

[41] http://www.sherpa-leap.ac.uk/lasso.html

[42] http://www.arrow.edu.au/repositories

[43] Groenewegen, David. "Re: ARROW", email to the author, 23/07/2009.

[44] Treloar, Andrew. "Re: ARROW, email to the author, 23/07/2009.

[45] Groenewegen, David. Op. Cit.

[46] Groenewegen, David and Treloar, Andrew. (2008) Op. Cit. P205.

[47] Groenewegen, David. Op. Cit.

[48] Ibid.

[49] Groenewegen, David and Treloar, Andrew. (2008) ibid.

[50] Groenewegen, David. Op. Cit.

[51] Ibid.

[52] Ibid.

[53] Groenewegen, David and Treloar, Andrew. (2008) ibid.

[54] Frantsvåg, Jan Erik. (2008) Open access in Norway: where are we and where are we going? Sciecominfo , 1. P1-2.

[55] Frantsvåg, Jan Erik. Telephone interview with the author, 21/08/2009

[56] Joki, Sverre Magnus Elvenes. (2007) Op. Cit. P205.

[57] Cit. P1.

[58] Frantsvåg, Jan Erik. (2009) Op. Cit.

[59] Joki, Sverre Magnus Elvenes. (2007) Op. Cit. P205.

[60] Frantsvåg, Jan Erik. (2009) Op. Cit.

[61] Joki, Sverre Magnus Elvenes. (2007) Op. Cit. P207.

[62] Røyneberg, Ellen. (2007) Op. Cit. P2.

[63] Frantsvåg, Jan Erik. (2009) Op. Cit.

[64] Joki, Sverre Magnus Elvenes. (2007) Op. Cit. P207.

[65] Ibid. P208.

[66] Røyneberg, Ellen. (2007) Op. Cit. P6.

[67] Joki, Sverre Magnus Elvenes. (2007) Op. Cit. P208-9.

[68] Frantsvåg, Jan Erik. (2009) Op. Cit.

[69] Digital Repository Federation. (2008) The DRF (Digital Repository Federation) report during 2006-2008 [online] available at: http://drf.lib.hokudai.ac.jp/drf/index.php?plugin=attach&refer=Digital%20Repository%20Federation%20%28in%20English%29&openfile=DRF%20report%20during%202006-08.pdf [last accessed 27/09/2009]

[70] Kuil, van der Annemiek and Feijen, Martin. (2004) The dawning of the Dutch network of Digital Academic Repositories (DARE): a shared experience.nnAriadne, 41, [online] Available at: http://www.ariadne.ac.uk/issue41/vanderkuil/#3 [last accessed 25/09/09]

[71] National Institute for Informatics (Japan). (2008) New Horizons in Academic and Scholarly Communication: NII National Institutional Repositories Program (Phase 1 report). [online] Available at: http://www.nii.ac.jp/irp/en/archive/pdf/CSI_Phase1_Report.pdf [Last accessed 05/10/09].

[72] http://www.narcis.info/index

[73] Kuil, van der Annemiek. "Re: DARE questions", email to the author, 12/08/2009.

[74] Ibid.

[75] Ibid.

[76] Ibid.

[77] Kuil, van der Annemiek and Feijen, Martin. (2004) ibid.

[78] Kuil, Van der Annemiek. (2009) Op. Cit.

[79] Ibid.

[80] Ibid.

[81] Ibid.

[82] Ibid.

[83] Ibid.

[84] National Institute of Informatics (2008) New Horizons in academic and scholarly communication: NII Institutional Repositories Program Phase 1 report" [online] available at: http://www.nii.ac.jp/irp/en/archive/pdf/CSI_Phase1_Report.pdf [Last accessed 25/09/2009]

[85] National Institute of Informatics Institutional repositories Program, current IRs available at: http://www.nii.ac.jp/irp/en/list/ [Last accessed 13/10/2009]

[86] Sugita, Izumi. "Re: JAIRO enquiry", email to the author, 25/09/2009.

[87] National Institute of Informatics, Op. Cit. P13-4

[88] Matsuura, Katherine. (2008) Japanese Institutional repositories: where did they come from and where are they headed? MSc dissertation, University of North Carolina at Chapel Hill. P2.

[89] Sugita, Izumi. Op. Cit.

[90] Ibid.

[91] Ibid.

[92] Digital Repository Federation, available at:
http://drf.lib.hokudai.ac.jp/drf/index.php?Digital%20Repository%20Federation%20%28in%20English%29 [Last accessed 12/10/2009]

[93] Sugita. Op. Cit.

[94] Matsuura, Op. Cit. p39.

[95] Sugita. Op. Cit.

[96] Digital Repository Federation (2008) Op. Cit. P1-2.

[97] Swan, Alma *et al*. (2005) Developing a model for e-prints and open access journal content in UK further and higher education. Learned Publishing, 18, pp25-40.