

Sakai/DSpace Integration White-Paper

Table of Contents

Sakai/DSpace Integration White-Paper.....	1
Overview.....	2
Abstract.....	2
Introduction.....	2
Participating Communities.....	2
Sakai and DSpace as Joint Enterprise Applications.....	2
Technical Interface and Interpersonal Agreement.....	3
Virtual Learning Environments for Research.....	3
Observations.....	3
Motivating Uses.....	4
Conducting Research.....	5
Researcher Webpages.....	6
Resources for Small-Group Teaching.....	6
Broad Dissemination of Material.....	6
Monitoring Automated Submissions.....	6
Librarians Preparing Collections.....	6
Teacher/Researcher Preparing Collections.....	7
Librarian Offered Incomplete Work.....	7
Issues at Project Termination.....	7
Menu of Options.....	8
Menu Overview.....	8
Extending Resource Tool.....	8
Site Archive.....	9
Showcase Tool.....	9
Bibliography Tool.....	9
Learning Object Creation Tools.....	10
Search Tool.....	10
Researcher Webpages.....	10
Teaching/Research Collections.....	11
Realtime Discovery.....	11
Sakai-Delegated Permission.....	12
ERobots.....	12
Item Semantics.....	13
Collections Workflow.....	13
Policy Manifestation.....	14
Migration Services.....	16
Summary.....	16
Discussion.....	17
Primary Tasks.....	17
Classification of Tasks and Roadmap.....	17
Recommendations.....	18
Outcomes.....	18
Acknowledgements.....	19
Glossary and Further Information.....	19
Appendix: Example Interventions.....	20
Resource Tool Extension.....	21
Policy manifestation.....	21
ERobots.....	22

Overview

Abstract

We seek to identify obstacles and opportunities in integrating DSpace with Sakai in practice, in an institutional context. Through identifying existing user tasks which could benefit from such an integration, along with problems which may be encountered in attempting to achieve this, we identify tools and services which could lead to practical integrations. We focus on these tools, rather than on technical integration issues, which may already largely be resolved with existing APIs, protocols and standards, or else are already being actively addressed.

Introduction

This white-paper attempts to present a route to integration between Sakai and DSpace in particular, and between Virtual Learning and Research Environments and Institutional Repositories, in general. The issues which it addresses are not exclusively technical in nature, however we conclude this paper with a set of proposed technical interventions which may aid the overcoming of broader issues.

Formal quantitative research data in this area is sparse, and the data sets for such studies tend to be small. Much qualitative research also tends to have occurred in informal circumstances. Therefore, we rely to a large extent on unpublished reflection of practitioners which has been explicitly solicited for this paper. The testing of the hypotheses in this paper through more formal studies is actively encouraged.

We have attempted to consult widely, incorporating the practice and attitudes of many institutions and projects. However, the available formally published information [Repo, TLRP, IMLS, CREE], and best practice guides and prominent opinion pieces [LEADIRS, InstRep], have proved of great value in informing this work.

Participating Communities

The following opinions were strongly represented in our research [Repo, CREE, IMLS] and informal conversations.

Sakai and DSpace as Joint Enterprise Applications

There is perhaps a tendency for the Sakai community to regard Sakai as *the* enterprise application for education, and therefore to relegate the repository to the role of database or of distributed multimedia file-store *at its command*.

Having an advanced file-store, such as this within Sakai is, indeed valuable for the Sakai project. Work is already under way to add the ability to use Fedora as a backing store for Sakai's Content Hosting Service, for just such a purpose.

However [Boston], the DSpace community perceive DSpace as a full-height application, embodying modelling, design, functionality, and user-interaction paradigms which are quite different to those of Sakai, but nevertheless valuable to users in many circumstances. DSpace continues to make invaluable mature contribution in a number of areas including issues relating to:

- long-term preservation
- searching and cataloguing
- managed redistribution of whole items and their metadata.

It seems likely that DSpace will remain a valuable full-height application (from database to user-

Sakai/DSpace Integration White-Paper

interface) into the indefinite future.

A user of a library, a *reader* or *librarian*, say, participates in roles and activities which are traditionally at least partially distinct from those of *teacher*, *student* and *administrator*. It is consistent that such roles and activities will remain distinct to a significant degree as this functionality is reimplemented digitally. It is therefore likely that for some time distinct and separate applications will remain to fulfil these distinct roles.

Technical Interface and Interpersonal Agreement

Clifford Lynch has described an Institutional Repository as not a piece of software, but as an agreement by an organization to steward digital data [Lynch]. Similarly, an institutional learning application is not a piece of software, but a set of individuals, policies, support infrastructure, and implemented decisions, to support learning at a particular institution.

The communities associated with Institutional Repositories on the one hand, and with VLEs/VREs on the other, are largely distinct, both within a particular institution and when considered more broadly as a community of practice. Therefore, similar but distinct management and administration infrastructures tend to emerge to support Institutional Repositories and Virtual Learning Environments. (As recently articulated in [TwnP]).

As this distinction exists, there is a need for explicit managerial agreement between the co-operating institutions to negotiate mutually acceptable and beneficial interoperation according to their specific concerns, above and beyond any technical implementation issues.

We describe approaches to this issue in greater depth throughout this white-paper.

Virtual Learning Environments for Research

Research Universities place great value upon deliberately blurring the distinction between teaching and research, particularly at postgraduate level, but also at the later stages of undergraduate teaching. Using a single tool for collaborative learning in both teaching and research tends to promote that indistinction, and it is perhaps significant that Sakai, which predominantly emerges from research-led Universities, resembles historic and current collaborative environments used within other professional domains (such as corporate intranets).

If, according to this model, Sakai is to be used for collaborative research, repository managers should expect to find preprints, experimental data, grant proposals, and similar academic material, being managed within VLEs such as Sakai. A brief survey of two live projects which use both Sakai and DSpace (separately) [TLRP,AERS] reveal that, amongst other items, their Sakai worksite currently contains:

- academic papers
- proposals and reports
- research instruments (questionnaires, and so on)
- abstracts
- lists of references
- records of the development of a controlled-vocabulary for an ontology.

Observations

A number of studies have been published [IMLS,Repo,TLRP] which examine researchers at work, and examine their use of repositories. Individual project managers, academics, and administrators, have also been willing to share their experiences in the compiling of this paper. It should be borne in mind that the available sources are currently largely unverified, uncontrolled, and may be subject to sampling error and skew. The information listed below summarizes the

Sakai/DSpace Integration White-Paper

recurring motifs in these discussions. Much of this material draws extensively upon [IMLS].

- **Time Pressure and Chaos:** academics feel under intense time pressure and feel that their work is chaotic. Any interventions which result in increased workload, or are unduly complex are likely to be avoided, or worked around.
- **Search within a VLE Is Appreciated:** studies [CREE] have revealed genuine user interest in *search* within a VLE context. This interest extends to library (OPAC) search, as well as to digital repositories. However, interest decreases when users are asked about Internet search, as they feel adequate Internet search tools already exist.
- **Disinterest in Repositories:** researchers claim little interest in the terminology or theories of repository practice. They are concerned with the issues addressed by repositories, but lose interest when discussed in the context of Information Science.
- **Hurdle in Publication Process:** repositories are often seen as a part of the administrative bureaucracy of publication, and repository tasks are frequently delegated to personal assistants and summer students. Repositories are strongly identified with the act of publication.
- **Over-Optimism Concerning Dissemination and Preservation:** researchers see dissemination and preservation as important issues, but tend to overestimate the provision afforded by their own ad-hoc practices.
- **Cooperation and Multi-Homed Work:** researchers are prepared to accommodate many problems, and a quantity of complexity if, as a result, they will be able to collaborate effectively with distant authors, or to work from multiple locations.
- **Assistance Sought for IPR and Organization:** researchers are often uncertain about IPR, both in terms of legal theory and practical tracking of the rights of documents in their possession. Users seem to act positively towards any assistance offered in this area. They also appreciate assistance with more general document organization issues.
- **Continued Importance of Face-to-Face Meetings:** The use of telephones and letters is now minimal in the pursuance of research. However, face-to-face meetings remain valuable, alongside newer electronic media.
- **Value of Visualization:** researchers and students find imaging and visualization tools invaluable. Advances in visualization over paper-based representations is seen as a major benefit of digitised resources.
- **Fear of Repositories:** a repository, however deployed, is seen as necessarily a public-facing permanent record of a researcher's output. Researchers would often rather not contribute at all than to later find they had accidentally submitted faulty data.
- **Fear of Metadata:** some repositories are designed with highly-complex metadata requirements (for example, multi-layer controlled-vocabulary ontologies). This can make deposit a complex process. Also, these vocabularies are often the subject of debate within a community, which makes choosing a classification a complex and value-laden process, and can result in the user preferring not to deposit work.
- **Integration of Teaching and Research:** users are often keen to integrate their teaching with their research. This already takes place, informally, with the use of simpler research resource types (such as academic papers) in small group teaching.

Motivating Uses

Following consultation and development of emerging ideas, a number of common patterns of usage emerged to motivate integration between VLEs/VREs and repositories. We present these use-cases below.

Sakai/DSpace Integration White-Paper

In order to reduce the complexity of this report, in each section we briefly cover both the motivation and practice of the end user for each use (use cases), followed by abstract technical overviews of how a solution could be achieved (use case *models*). Within our research, these were developed separately. For further details of published use case studies, and other work, drawn upon in this section, see [CREE, Repo, IMLS, LEADIRS].

Conducting Research

A number of sites already use Sakai for collaborative research. A significant example, local to the authors, is the ESRC funded TLRP project [TLRP]. Considerable JISC VRE research funding has also been allocated to the investigation of the suitability of Sakai as a VRE.

- Sakai can be used to share and produce academic papers and datasets, and for discussion of research with colleagues. Were the Sakai user-interface to include components allowing interaction with the library for the purposes of deposit and dissemination, the researcher could deposit their material electronically with the library. It seems likely that control over the use of the material would necessarily remain with the author for there to be appreciable uptake of this service.
- A researcher could also use Sakai-based tools to manage references and bibliography: a number of tools providing this functionality are currently in development. These tools will integrate with both traditional libraries and digital repositories, so that they can allow a researcher to view papers which are available online, to discover library availability of documents, and to be provided with accurate bibliographic metadata, and so on, from within Sakai.
- Researchers often work with material over which they have restricted rights. Consequences of this necessity are already evident within the information environment. One major example of such a consequence is the facility, in a large proportion of archives and libraries, to deposit items on an expiring non-disclosure basis, and other similar arrangement. In Sakai/DSpace integration a researcher may well expect to be given similar ability to mandate limited access on various non-release bases, and be confident that this will be adhered to if such an agreement is made.
- Migration between software systems has shown that the status of complete projects is often indeterminate in the minds of researchers and research groups. Often the desire emerges to search through old “complete” projects, thought at one time to contain no remaining useful information, to recover archive material: machine operating manuals for laboratory machinery, intermediate experimental data and instruments, and so on. It is, arguably, the role of an institutional repository to steward this information through periods of user disinterest, in order that it may be available if subsequently requested. For institutional repositories which choose to take on such a role, an integration between Sakai and DSpace should be capable of easily archiving a worksite, to achieve such preservation at a low cost, and allowing continuity in research.
- If Sakai is used as a research environment, it can be used as a natural place to collate announcements, calls for papers, and so on. The addition of a “meta-search” into a repository (whether push or pull based, technologically) would allow any newly deposited items to be marked as events in a similar manner, and collated into the announcement stream.
- The arrangement of papers and other academic material into folders, within Sakai, represents implicit metadata, and also presents an opportunity for data-mining. Without user-intervention, a text-mining tool can analyse and cluster folder contents and use this to perform searches into digital repositories, and elsewhere. Relevant material can then be retrieved and placed into the folder, in a “*you may also be interested in*” section, if the user elects to use such a feature.

Sakai/DSpace Integration White-Paper

Researcher Webpages

Researchers typically assemble personal webpages representing their professional output. Given sufficient time, technically accomplished authors can create such pages using available tools. However, many other researchers may appreciate a tool to help them create such pages from within Sakai, using authoring tools which allow them to also manage content, incorporate news feeds, and so on.

A major feature of such pages, typically, is a list of publications, along with links to online copies of those publications, where available. If Sakai is integrated with DSpace, then this tool can be enhanced by a digital repository. The repository can make deposited papers available to the user preparing the bibliography. Also, however, the tool can offer a facility to simply deposit data into the repository upon addition to the publication list for the personal page, which will be accompanied by rich and carefully checked metadata (as such checking is strongly in the user's interest!). Further details of this use case can be found in [IMLS].

Resources for Small-Group Teaching

Small-group teaching, particularly at research Universities, occasionally involves the examination and discussion of research material, such as academic papers, and scientific data-sets. This material has supporting teaching material added (perhaps in the form of introductions, or emerging questions for teaching), whereupon it becomes a *de facto* learning object. A tool could assist this process, by presenting an authoring interface to a teacher.

This provides integration opportunities for interoperating Sakai and DSpace instances, firstly using the institutional repository to select research material around which to build a learning object, and secondly to provide a location for easy deposit into a DSpace acting as a learning object repository when the object is complete.

Broad Dissemination of Material

Format Migration is likely to emerge as an important tool for many repositories wishing to maintain material on a multi-decade time-scale. Tools to handle such migration are already emerging in repository software.

It is also, arguably, a role of a repository or library to make material available in a manner requested by the user, where this is practical. This can be seen to encompass dynamic delivery of digital information in those formats which the user can easily view.

A researcher wishing to communicate contemporaneously with colleagues may face similar challenges with formats. The services available in digital repositories, primarily motivated for preservation reasons, as described above, could be made available through a Sakai tool, to allow format migration for a user's individual needs, where this is economic. This provision is analogous to a library providing book-binding services to the general institutional community, through a need to maintain such skills to manage their own collections.

Monitoring Automated Submissions

The automated nature of a web-services type integration between Sakai and DSpace seems likely to create concerns both for DSpace system administrators, wishing to maintain some administrative control over the repository, and also for archivists concerned with preserving record-status of documents within a repository.

A mechanism within DSpace to allow the fine-grained monitoring, representation and intervention of such automated accesses would allow a wider range of integrations to proceed.

Librarians Preparing Collections

Librarians occasionally create collections and exhibitions from material available within their

Sakai/DSpace Integration White-Paper

library. This can be derived from material available on public shelves, but may also derive from other material available to the library. Using DSpace as its basis, a tool could be developed to allow a librarian to prepare such a collection, and then “publish” it as a learning object, making the object available to Sakai users as an “exhibition” or “special collection”.

If DSpace/Sakai integration includes an indication of which collections or communities are to be regarded as the basis of learning object repositories, Sakai tools can be developed to integrate learning objects developed in this manner with the learning object infrastructure being developed within Sakai.

Teacher/Researcher Preparing Collections

Teachers and Researchers also produce collections and exhibitions in a manner similar to those described in the preceding section. However, in this case it is likely that their tool of choice will be their VLE/VRE, and that learning objects may incorporate primary material not present within the digital repository. A similar tool to that used in DSpace for librarians can be developed within Sakai to assemble these collections. However, this time Sakai/DSpace integration beneficially operates in the “other direction”: the completed object is transferred into a DSpace learning object repository from inside Sakai.

Librarian Offered Incomplete Work

It is likely that, once Sakai and DSpace have been integrated, material will be offered for deposit which is not of sufficient quality to be made publicly available within the library. The library may wish to store this information within DSpace, away from public “display” in order that it may potentially be recovered by special arrangement, by analogy with *dark archives*. However, after some work upon such objects it may be that the object can then become publicly available as a part of a library collection.

The provision of workflow, state management, and various degrees of *visibility* within DSpace would allow repository managers to accommodate those collections whose standard falls short of their repository policies, within *dark collections*.

Whilst the provision of such facilities within DSpace does not directly impinge on the technical aspects of Sakai/DSpace integration, it seems likely that many more successful integrations would take place were DSpace to adopt this, or other, strategies in managing data accompanied by metadata of variable quality.

Issues at Project Termination

As discussed in the earlier Conduct of Research section, experience suggests that there is a great deal of “churn” and instability, in both policy and practice, at the start and end of research projects. This can lead to significant periods where information is unmanaged. In well-funded, technically adept, departments central services provide a location for transiently unguarded information.

It seems likely that the wider provision of services to encourage data-continuity would be desirable, and that an institutional repository could see it as their role to provide such a service.

In the situation where Sakai is used as a research tool, a considerable quantity of information will be within the project worksite. A simple tool to archive a Sakai site within a DSpace repository, for future examination, would allow the data to be maintained over the vulnerable inter-project period.

It is likely that such data would be “incomplete” in the sense of the previous section, so the use of tools suggested in that section would aid the librarian in performance of this task. As, almost by definition, few individuals will be able to allocate significant quantities of time to this “orphaned” information, such an archival step would need to be simple and easy to use, perhaps at the expense of an increased workload for a subsequent data user.

Sakai/DSpace Integration White-Paper

Such incompleteness in metadata can be accommodated for two reasons. First, context is implicitly encoded in the rich organizational structure of the resources within the resource tool of a large project. Secondly, as time progresses, users of information which is regarded as uncatalogued (as this information would be), already expect to spend considerable proportions of time reconstructing context. A digital repository, in this context, functions not only in a traditional archive or library role, but also at a much more fundamental level, as a store for unsorted documents. Collections Workflow within DSpace (see later sections), would allow material to be migrated, in time, if resources allow, into archival formats.

Menu of Options

Menu Overview

We now present a number of options for technical interventions which may aid integration between Sakai and DSpace in an institutional context, and between VLEs and Repositories in general.

These options are presented in addition to the non-technical 'community' interventions which would be focussed around increasing the involvement between the two communities, particularly in communicating their "world-views", terminology, and underlying aims.

The *Sakaibrary* [Sbry] project potentially covers many of the interventions described below.

We present a broad range of potential solutions, relating to the practice of using VLEs with repositories, and some of these options involve minimal novel interaction between the individual pieces of software, but nevertheless are designed to aid implementation, in practice, of integration of Sakai and DSpace in an institutional context.

Extending Resource Tool

Many of the use-cases described above involve, as a necessary step, the visibility of DSpace collections within Sakai, and some would also benefit from deposit being available through similar mechanisms.

For example, in the conduct of research, in compiling material for small-group teaching, and in compiling personal webpages, the presence of repository items visible within Sakai resources would be intuitive in both user-interface and administrative terms.

Initially, a mapping could be established between a DSpace collection and a folder within the Sakai resources tool, though a later implementation could relax this 1-to-1 correspondence. The resource tool could also be extended to incorporate a form of discovery, by means of criteria inductively derived from the explicit contents of folders. This would work a little like the "*users who liked these also liked*" feature on many e-commerce sites, and could potentially aid researchers in the discovery of literature valuable for the conduct of their research.

In conjunction with negotiated policy manifestation mechanisms (an intervention which is described subsequently), the tool could also provide an implementation of any policy requirements for ingest, such as display of click-through "splashes", the display of available service-levels of the repository, asking for user-selected criteria, checking of copyright status, and similar workflow operations, to be executed within Sakai, as specified by DSpace. Some of these (such as certain metadata elements, or workflow items) could be specified at the folder level to reduce the overall complexity of the ingest process.

From our research, it appears to be important, in order to retain the interest of end-users, that this interface is presented in terms with which the user feels engaged, rather than the abstract and perhaps more precise terminology of Information Science.

Most of this work is in scope of existing LNI/SRW/OAI-PMH DSpace interfaces, with the possible exception of any initial policy manifestation exchange. This work appears to be within the scope of the *Sakaibrary* [Sbry] project, but not is unlikely to be a focus in its initial stages.

Sakai/DSpace Integration White-Paper

Site Archive

Use cases emerging from issues surrounding project termination suggest that the availability of a tool which provided simple one-click archiving of worksites into DSpace would help solve issues surrounding temporary breaks in the chain of responsibility for digital information.

A site archive tool could deposit a complete worksite for preservation ends. It could also be used to disseminate site templates, to create similar sites based upon the template.

Initially, this tool could be a simple modification of the existing archive Sakai tool, to allow it to use LNI to deposit in a repository (possibly into a distinguished individual collection, described in a policy manifestation, described later).

A second phase for this tool would be to migrate from Sakai's entity format, towards a more open format, such as an IMSCP-based scheme, to allow for longer term preservation. Once such a format has been decided upon by the Sakai community, the existing preserved sites would be migrated from the entity format to the more stable standard.

Showcase Tool

For the uses-cases involving librarians and curators producing their own material, the natural application for them to use to do this is an extension to DSpace. With the addition of the ability to add front- and back-matter to contextualise a collection, the librarian may then wish to publish, or "push" this collection to Sakai.

A showcase tool would demonstrate the available collections for integration with Sakai sites, and could take the form of a separate tool, oriented around the Announcements/RSS "push" interaction model, or else a service could place collections into the resource tool of a distinguished Sakai worksite.

The showcase tool could be implemented using existing DSpace protocols, with the addition of a convention indicating the location of the supporting material, which would, themselves, be DSpace items. To achieve this, either an out-of-band agreement or convention could be entered into, or the policy manifestation could be used (described in a subsequent section).

By adding copious categorization information (in the style of OAI-PMH sets), the showcase tool could be configured by each user to provide largely information of interest, or to categorize the material into Sakai folders.

This tool could be implemented using only LNI/SRW/OAI-PMH.

Bibliography Tool

In various of the use-cases above, both in research and teaching, the compilation or use of a reading list, bibliography, or similar instrument, forms an essential part of the task. This is functionality which is frequently requested, and could naturally be implemented as a Sakai tool.

Such a bibliography tool would, amongst other functionality, capture bibliographic information from searches performed with a companion search tool (described subsequently). It would integrate with DSpace to the extent that its data model could incorporate persistent URLs (perhaps for multiple formats), and also by allowing metadata to remain "live" with respect to modifications of metadata performed within DSpace.

Some of the references within bibliographic lists are likely to be outputs of projects which were using Sakai as a cooperative research worksite. As such, it is likely that the editing community within such a Sakai site could be said to correspond to the user community who 'own' an item, and so authorized to update and correct metadata. With co-ordination with the repository in terms of permissions/policy users could be enabled to update the metadata for these documents from within the bibliography tool, and this be reflected in the repository (where the paper may be currently stored in a 'dark' collection because of poor metadata quality). This delegates part of the task of improving metadata quality back to the academic users, but in a context which is

Sakai/DSpace Integration White-Paper

seen as academic rather than administrative [IMLS], and allowing it to be done at a time other than an approaching publication deadline.

However, there are more general wins for DSpace, too, when the bibliography tool is used in coordination with the search tool. If search results persist, through a bibliography tool, and if those results originally flow from a DSpace repository, then this metadata is more likely to be quoted correctly, and the correct URL given when this data is subsequently used in papers, or on other sites. With the aid of the tool, bibliographies could also be added to DSpace as first-order items, or as user-specified collections (should that be in scope for the repository).

This tool could be implemented using standard DSpace protocols.

Learning Object Creation Tools

Motivated, primarily, by use cases surrounding small-group teaching, it is likely that there will be value in the development of tools for creating learning objects by practitioners, within Sakai.

The provision of easy-to-use interaction to allow deposit into a Learning Object repository would allow institutional Learning Object repositories to be more easily populated.

Such a Learning Object creation tool could use a repository in two ways. First, for discovery (probably in association with a search or showcase tool) during composition of the learning objects, and secondly for the ingest of these objects into a learning object repository once they have been created (perhaps to a distinguished individual collection determined by policy manifestation, as described in later sections).

These tools could use standard DSpace protocols.

Search Tool

There is already considerable interest in implementing library search within Sakai, both of digital repositories and physical libraries. Already, a simple implementation exists amongst the outputs of the Twin Peaks [TwnP] project. Many of the identified use-cases in the proceeding sections would involve some search component.

A Sakai search tool would allow a user to search a number of repositories and libraries using SRW, Z39.50, [SRU/W, Z39.50] or similar, and then present these results to the user. This would also integrate with any bibliography tool (described in earlier sections) to preserve any valuable results. This presents an opportunity for the Institutional Repository's profile to be raised, by its search results being presented alongside those of the traditional library OPAC.

A policy manifestation could allow a search interface to use various non-standard metadata terms, where such is supported by a particular repository, or collection. This tool would use standard interfaces and protocols.

Researcher Webpages

The Researcher Webpage creation use-case (described fully in [IMLS]) creates opportunities for repository maintainers to secure high-quality metadata directly and willingly from those responsible for the material.

The publication of researcher webpages, with their strong emphasis on bibliography, presents an opportunity for a researcher to deposit a paper within a repository, and to provide accurate metadata, whilst they are producing their personal research webpages.

An editing tool for personal professional webpages would take the form of a graphical editor within Sakai. In addition to the usual services of such editing, a section would add bibliography functionality to allow a user to prepare personal publication lists based upon papers in digital repositories and elsewhere. Upon preparing a page for publication a user would be asked if they wish their documents to be added to the repository (for those not already present, but instead drawn, say, from Sakai site resources or a local machine), and to add these resources to a

Sakai/DSpace Integration White-Paper

particular collection reserved for personal professional papers.

As further inducement, the repository could offer migration services to ingested items to ensure that a paper (or similar item) is available in a wide range of formats (pdf, ps, etc.).

Such a tool would provide unusually high-quality metadata to the repository, as the user will take care that their personal webpage contains accurate details.

Teaching/Research Collections

In some of our use-cases we consider a teacher or researcher preparing a collection from resources within a repository. Were there facilities within Sakai to allow intuitive presentation of these collections, depending on their type, perhaps as a part of a previously discussed Learning Object Creation tool, (and analogous to the current presentation tool), the presentation of many items and collections could be improved.

Through the use of policy manifestation (discussed in a later section) the material can be presented by a suitable enabled presentation tool within Sakai in a user-friendly form. This would involve establishing which parts of a bundle or item perform which particular roles. For example, an image item may incorporate thumbnails, full size images, watermarked images, 'tiles' for zooming, and so on. If these can be identified reliably, a presentation tool within Sakai can present the items in an organised and consistent manner (as is currently being developed for Stellar [StIm]).

It is also possible that in certain circumstances it would be appropriate for these ad-hoc collections, once assembled, to be resubmitted to DSpace as collections for preservation or, perhaps, as items containing collection member pointers for a personal portfolio style preservation.

In the case of the collection-addition operation, an extension, or convention, would be required to LNI, for other parts standard DSpace protocols can be used.

Realtime Discovery

Many of the use-cases discussed above, and particularly that of conducting research, could benefit from dynamic updates when new content becomes available within repositories, or through library search [Sbry].

A realtime discovery service could be a separate tool, or else a service within Sakai. Such a service would be used for dynamic update operations, in order to extend search and bibliography tools (described above) to allow results to be dynamically updated with any new content. The tool would discover such updates using OAI-PMH, or similar mechanisms. This functionality is being discussed within the Sakaibrary project [Sbry], using an intermediating OAI to RSS server. There is increasing interest, within the DSpace community, at the use of ATOM as a mutually comprehensible realtime update mechanism [Boston].

Work on such a tool would impact on similar areas as any Showcase Tool (discussed above). However, the manner in which information is discovered by realtime discovery tools would differ from that of the showcase tool, in that the showcase tool would be designed to *collect* information into a single place for easy discovery by someone explicitly searching for it, whereas the resource discovery tool would be designed to *distribute* information to places where it may be most relevant for opportunistic discovery by users going about other business. For example, material could be placed as "suggested" material within Sakai folders within the resource tool.

This tool could use existing DSpace protocols throughout, though migration towards a consensus protocol (such as ATOM) and away from RSS and OAI would remove the need for an intermediating protocol-translating application as is being considered, for example by *Sakaibrary*.

Sakai/DSpace Integration White-Paper

Sakai-Delegated Permission

There are a number of concerns surrounding the general ability to automatically access repositories. The use-case concerning system administrators and repository managers monitoring remote submissions is concerned with such matters.

Permission within DSpace is usually granted on a per-EPerson basis. Were a robot, embodying the integration between Sakai and DSpace, given the identity of a single EPerson, the union of permissions granted may give the robot more permissions than it should be granted in any one operation. However, allocating an EPerson to each VLE user could result in a considerable administrative and maintenance overhead in managing the DSpace userbase.

The issue of "robot" access is addressed in a separate section. A part of this provision, though, would allow Sakai (or other remote tool) to be delegated a part of the responsibility in identifying users. A tool could be developed which allowed Sakai administrators and superadministrators to delegate DSpace permissions to individual users and groups of users, which could then be used in interactions with DSpace.

Through the ERobots mechanism (below) this tool could be restricted by repository managers to be a bounded delegation.

The Sakai tool would need to be newly developed, but no DSpace work would be required beyond that described under ERobots (below).

ERobots

Again addressing the use case surrounding retaining administrative and managerial control over repository policies, the ERobots work focuses on managing the access and update permissions for automatic connections to the repository ("robots").

One of the current tensions between library and VLE communities during integration concerns authorization. DSpace, when used as a service, is often required to accept "generic" credentials of a service connecting (for example Sakai) either as a "fake" EPerson, or (even worse) as the name of a user who has been cajoled into letting their identity be used in such a way.

"ERobots" is a name for a set of tools which would move towards addressing these issues within DSpace. A firm footing would be established by first differentiating human users from individual "robots" within the data model, to aid logging and tracking for administrators by representing this essential difference.

In addition to presenting its own identity, a robot could offer an identity to DSpace, with each operation, which indicates the user on whose behalf it is working at that time. This user could either be identified within the DSpace domain (as an EPerson) or only recorded within the domain of the caller. Whilst these details could be forged by a robot (so are no substitute to securing completely off limits collections, based upon ERobot identity), with a trusted ERobot, malicious, confused or lost users using an ERobot from a trusted system could be easily identified, and damage controlled.

Perhaps, with ideal infrastructure, authenticated user-ids within a single authentication domain would be transmitted across network boundaries, allowing a single-signon environment with end-to-end accountability and without any endpoint ever delegating trust over performance of this process. However, it seems likely that, for the foreseeable future, an ERobot type mechanism would be a considerable, low impact, improvement over the present situation.

One way to minimize impact on network protocols would be, for example, to use a rarely required (or currently illegal) punctuation sequence as an identifier and separator within a username. A robot would construct its robot identity, then the separator, and then the user it is acting on behalf of, and then use its single shared secret as a password for identification.

Sakai/DSpace Integration White-Paper

Item Semantics

The DSpace model does not currently record the semantics of its contents in a machine readable manner beyond the basic two-level syntactic structure of bundles and bitstreams within the data-model. A description of an item both structurally, and historically, at a distinguished place (a form of metadata) would enable extra features to be exposed within Sakai which may be of value in integration.

First, this work can be seen as a generalisation of an aspect of the proposed teaching/research collection work, describing an item in sufficient detail that a front-end can display the work in context appropriate to the display tool.

This could be extended, for example, so as to distinguish those bitstreams which represent alternatives (such as differing formats) and those which represent accretions (such as composite datasets). This could then be used by display tools to organise the information.

Secondly, historical records of changes and activities could be transmitted to the user, to reinforce their "ownership" of a submitted document, describe how it has been archived and disseminated, and allowing the user to intervene according to their wishes.

Initially, the History Feature of DSpace was identified as a site for such an implementation. However, further discussion with the DSpace team suggests that the history mechanism is probably inappropriate for this work, and that a separate system is appropriate.

The item semantics would need to be exposed through network interfaces, through either a modification to LNI, or through the establishment of a conventional location for a semantics bitstream.

Collections Workflow

Concerning issues around data which is incomplete, or otherwise unacceptable for the library in quality terms, the development of a workflow system could allow digital repositories to accept sub-standard data on a limited release basis, for subsequent improvement by the library, or by the original submitter.

DSpace workflow currently operates only prior to ingest. The addition of item state workflow actions determined by that state, even for ingested items, allows partial or incomplete items first-order status.

This work would be entirely within DSpace, however it would help mitigate the implications of the quality-mismatch that is likely to impact heavily upon real-world institutional integrations of DSpace.

Typically, it is envisaged that poor-quality data would be partitioned into particular collections. Collections workflow would maintain state associated with particular items and collections concerning issues such as dissemination and preservation guarantees, quality requirements, and so on. When certain conditions are met, this would trigger events such as beginning a transfer from one collection to another, the adding of an item to an OAI steam (perhaps as a member of a particular OAI-PMH "set"), trigger a particular migration, add or retract from the public catalogue, and so on. It's envisaged that many DSpace administrators will want to granularize these issues on a collection-by-collection basis (hence the name "collections workflow").

This work would amount to an extension of the DSpace ingest workflow (or a functional equivalent) so that workflow could be used after ingest upon items already within the repository. This change would be motivated by extended times in which data may be present in a repository (benefiting from preservation and migration services, for example) without fully meeting the criteria for public availability of an item within that particular repository.

A tool would present a view for a DSpace item which could represent its state in terms of HTML form elements (for example). An authorized user could alter these elements. A predefined configuration (either through a web UI, or through a static configuration file) would trigger events upon a particular item's conditions being met. These states would be reflected in the item's

Sakai/DSpace Integration White-Paper

representation.

For example, a curator could add a particular piece of missing metadata and this trigger a move to a more public collection. Or the article quality could be reassessed (say from peer-reviewed journal to internal technical report) and so the item again move to another collection. This allows complex collection policy to be separated from individual item and collection properties, and so allows library users without full knowledge of the repository workflow to simply manage collections in repositories with complex quality constraints.

Accompanying this workflow implementation would be the task of adding to each collection (or item) a metadata representation of its state (perhaps in an XML format) to allow such workflows themselves to be archived and retrieved (perhaps along with human-readable notes describing motivation). Such a representation could be associated with Item Semantics work.

Policy Manifestation

A number of the above implementations make use of what is described as 'policy manifestation'. Essentially this amounts to, upon initial connection between Sakai and DSpace, information being shared concerning the organisation of their sites and collections, and the various operational mechanisms required of each piece of software to ensure that policies are adhered to.

Sakai can then (for example) add extra information to the resources tool to indicate the status of an item concerning preservation, dissemination, licence, etc. Sakai can learn of the correct collection to use for specialist tools, if any (for example the webpage and Learning Object creation tools).

The negotiation of Policy Manifestation would constitute a handshake at the initiation stage of a web-services connection. The handshake would take the form of exchanging a semantic representation of pertinent parts of a DSpace and Sakai configuration and policy in machine-readable form. These two representations, the local and remotely-received representations would be "merged" by both parties using a common algorithm so that each derives a common representation of the capabilities of a connection, not only in technical terms but also in terms of manifestations of policies adopted by each party. It is reminiscent of a high-level version of the "handshake" present in many Internet protocols.

Rather than represent policy in an abstract form (which is a complex task) the policy document would contain a number of requirements for concrete tasks which each party expects or allows the other to execute, or lists of members of a class for use in such operations. These could include, for example

- Lists of available collections.
- Lists of restrictions upon particular EPersons, to ensure that only relevant subsets of functionality are displayed.
- Lists of distinguished collections to be used by certain specialist tools (if permitted). Examples may include collections for Learning Objects, for Sakai-user-created collections, or for documents produced using Personal Webpage tools.
- The contents of required notices to be read by a user, and agreements into which they must enter, manifested within Sakai prior to ingest.
- Lists of acceptable policies concerning copyright, licensing, preservation and dissemination.
- Lists of required metadata, in terms of mandatory keys and/or ranges of values for particular keys.
- Lists of available extra metadata searches within particular collections, along with the ranges for values, to allow extended search of DSpace to be implemented within Sakai.
- Extra ancillary metadata, useful for DSpace workflow in focusing the ingest process (for

Sakai/DSpace Integration White-Paper

example, transmission from Sakai to DSpace of self-determined quality assertions, or of submission type).

Synonym relationships between policy documents, as authored by DSpace and Sakai, can be established by means of "fragments". These would be simple, uncontroversial action-based policy snippets which each side could assemble into allowed terms. (For example, one set of fragments might be the four principal constraints within a Creative Commons licence).

It would also make sense, within such a "policy" to describe the structure of collections and/or of their contents, if any, beyond the DSpace data model: for example, including a semantically readable indication of contact details for a collection's maintainer; machine readable details of programmes under which a collection is managed, front- and back-matter, and so on. This per-collection information would be augmented by the per-item information available through any Item Semantics implementation (discussed above).

The manifestation could also include a representation of the repository software status in order that community-prevalent patches and changes which are understood by both Sakai and DSpace (for example, concerning special issues surrounding image galleries) could be used when available in a particular implementation.

This policy document could be developed within RDF, which could make it simple to extend. The policy terms would be developed within the context of the DSpace community, and would be DSpace led. The changes to DSpace, however, would be minimal, amounting to the installation of the simple reasoning engine within DSpace (which could be configured off-by-default) and the establishment of the communication protocols necessary to communicate this document (in its simplest form it could be a single static URL at the DSpace end).

Within Sakai, large changes are required. These changes would concern the implementation of resource integration tools. This asymmetric workload, the simple implementation within DSpace and more complex work within Sakai, emerges from the symmetric protocol because it is largely Sakai providing the user-interaction, which would need to be customized according to any negotiated policy. For example, the following features would be desirable within Sakai, derived from the initial policy negotiation, (in some cases in conjunction with per-item metadata),

- The ability to display (repository-defined) notices and brands to the user at ingest.
- The ability to present (negotiated) drop-downs and other user selectors at ingest to guide ingest workflow (perhaps in terms of destination collection) and additional metadata.
- The ability to check existing Sakai properties (such as IPR properties) are consistent with repository policies prior to ingest.
- The ability to modify search to include other metadata key terms for a particular collection (for example, geographical information, or reprographic technical metadata) within any search tool.
- The ability to associate, with items pulled from a repository, various clarifying notices which persist along with data and metadata throughout their use in Sakai, for example, in a bibliography tool.
- The ability to fulfil licence requirements (for example, attribution in creative commons licences), through identifying those metadata fields within the repository which contain this information, and to make this information available to Sakai.

To reduce the potential for a cluttered and confusing user-interface, at the Sakai end it is likely that many of these requirements could be fulfilled on a *sticky* per-folder basis. This feature would entail implicit or default values being taken for some of the requirements (such as user-asserted quality, IPR requirements, and some metadata fields) merely by the presence of an item in a particular Sakai folder which has been configured appropriately.

This work is related to that of the PLEDGE project being carried out at MIT, which is focussing on repository policy issues, particularly in a distributed data grid context [PLEDGE].

Sakai/DSpace Integration White-Paper

Migration Services

Many libraries offer format migration facilities for their physical media to end users (such as reprographics and book-binding), available partially because they consider it within their service remit to do so, but also because of the need to perform such migrations for the purposes preservation.

As preservation services come to be built into repositories such as DSpace, it could make sense to make (some of) these services available as a web-service so that they can be performed, for example, by Sakai tools, at the user's instigation. In order to keep costs under control, the repository would probably want to rate-limit these migrations, though many automatic migrations are probably extremely cheap.

Two services could support this functionality. The first is a conceptually synchronous service where a Sakai requests a particular migration and DSpace obliges (or otherwise). The second is an asynchronous protocol where Sakai registers an interest in an item (for example, because it deposited that item and retains a view upon it within resources), and DSpace alerts Sakai to any newly available formats.

This work fits closely with the Item Semantics which would be used, amongst other purposes, to describe migrations. Where such a feature has been implemented, the asynchronous notification could amount to little more than an update of the Item Semantics bitstream which could be discovered through OAI-PMH/ATOM/RSS.

This service would involve *either* the implementation of a new protocol for DSpace (or separate migration tool), to allow the instigation of migrations; *or else* the considerable repurposing of some functionality within an existing protocol, probably LNI.

Summary

A number of tools have been presented, motivated by use-cases following conversations with practitioners, and by examination of available studies. Many of the changes which would aid institutional integration between Sakai *as an institutional VLE*, and DSpace *as an institutional repository*, are not directly concerned with the technical interconnection of the software, but with providing tools within each of DSpace and Sakai separately, to allow integration to proceed in practice. These tools would be concerned, amongst other areas,

- with issues surrounding the use of information within repositories in an “activity” (or “tool”) based manner, as is typical of VLE paradigms, rather than the “object” based models typical of repositories. This led to the specification of a number of *tools* within Sakai, designed to facilitate use of a repository by VLE users, according to end-user use-cases.
- With the handling of differing policies, attitudes, and quality standards in the separate communities. Given the differing strategic focuses of libraries and research and learning environments, particularly concerning risk, relevant timescales, and counts of items, differing policies concerning the quality, for example, of metadata, necessarily emerge. Technical facilities, such as workflow, machine-readable policies, and item descriptions, can aid reaching an accommodation in these respects.
- with the minimisation of extra systems maintenance work required when an integration is in-service. Work such as partial delegation of trust, and explicit support for automated authorization (“ERobots”), can simplify these tasks.
- and with creating separate control over, and representations of, repository services (preservation, dissemination, metadata-rich representation, etc.) within end-user tools, allowing repository facilities to be used, where appropriate, beyond the traditional “preprint” use-case.

DSpace already supports a suite of web-service protocols, LNI [LNI], OAI, SRU/W, and so on,

Sakai/DSpace Integration White-Paper

which provide much of the functionality typical of repository interfaces (such as the OKI DR OSID [OSID]). When conjoined with the packaging and crosswalking services now provided within DSpace, further major extensions to these protocols seems largely unnecessary.

Perhaps the most important underlying work, upon which many of the proposed enhancements are based (or through which they could be enhanced), is the implementation of Policy Manifestation: a representation of state and policy in a machine-readable manner, analogous to the RDF-type content representations typical of the semantic web.

Discussion

Primary Tasks

The proposals above were discussed with MacKenzie Smith and the DSpace team at MIT. Together, a number of priorities were identified to ease integration between DSpace and Sakai.

- The Extension of the Resource Tool to aid research and teaching practice use-cases.
- Using ERobots to manage the presence of automated users.
- The implementation of Policy Manifestation to allow machine-understanding of capabilities and configuration.
- The development of a Site Archive tool for maintaining 'neglected' content, and distributing site templates and exemplars.

Consultation with colleagues within the University of Cambridge suggest that additional primary tasks from the above list may include

- Bibliography tool within Sakai.
- Search tool within Sakai.

Research [for example, CREE] and private correspondence with other UK institutions suggests similar requirements exist at other institutions.

Classification of Tasks and Roadmap

The tasks identified fall into a number of broad categories in terms of the tools and expertise required, and in terms of the various stakeholders involved.

A large proportion of the tasks identified above are, essentially, Sakai tools, the only technical DSpace involvement being the existence of a service (via OKI DR OSID, or otherwise [OSID]), within Sakai by which DSpace may be invoked (via LNI, OAI-PMH, SRU/W, etc.). These APIs and tools are available for development within Sakai, particularly with the emergence of a strong library development group emerging within the Sakai project. These tasks include bibliography and search tools, site archive and learning object creation tools, and realtime discovery mechanisms.

Other tasks are primarily DSpace work, mainly for the purpose of addressing issues relating to information quality, and with the integration of automated clients. These tasks include, ERobots, Policy Manifestation, Collection Workflow, and Item Semantics. This work would be useful in DSpace integrations with a broader range of teaching software, not only with Sakai.

A few tasks involve more closely integrated development within both Sakai and DSpace. In particular, Sakai-delegated permissions involve *Sakai* development to ease the workload of *Repository* administrators. Personal Webpages also fall into a similar category: the work is primarily Sakai-based, but many of the benefits are for repositories.

The Realtime discovery, Migration Services and Showcase Tool proposals have been enthusiastically received, but they are perhaps seen more as "neat-possibilities" than demand-

Sakai/DSpace Integration White-Paper

led requirements at the moment, and are therefore perhaps better left until opportunities arise where they can be developed in a more immediate context.

Those tools which primarily relate only to DSpace or Sakai can be implemented within those development communities with high-level consultation between projects. Some of the tasks, in particular Sakai Delegated Permissions and Site Archive may need to be "sold" to Sakai by the repository community, the former from the perspective of minimising increases to administrative workload, and the latter from the repository focus on information preservation.

A number of pieces of work would be valuable to support other tasks identified here, and so would benefit from early implementation. These include Item Semantics, and Policy Manifestation.

Recommendations

Both the **Item Semantics** and the **Policy Manifestation** work will be valuable services which can be used by the other tools which have been described in this paper. Early implementation of these two services would help minimise redundant effort which might otherwise be undertaken in tool implementations. However, both these (closely related) tasks are complex, and require extensive planning and consultation. It would therefore be highly desirable to attempt to secure funding for the development of these areas.

Projects which would benefit from **ERobots** work are already being undertaken at various Universities, and there is considerable negotiation at present to accommodate needs for interaction with a repository's requirements for integrity and security, in the absence of an ERobots-type solution (for example [Stlm]). This is already a practical hurdle in integration between VLEs and DSpace. The ERobots work could help with many of these issues. It is therefore recommended that this DSpace work be carried out as soon as is practical. The next hurdle in this area is anticipated to relate to authorisation management, for which **Sakai-Delegated Permissions** would help reduce the additional workload.

The Sakai site archive tool already provides much of the functionality required of the **Site Archive** task identified in preceding sections (it currently writes to local disc). Given that Sakai is already being used on high-status projects as a collaborative research tool which contains material for which preservation may be in order, the risk of information-loss is already present. Therefore, since Site Archive is associated with a time-increasing risk, and much of the implementation is already present, this intervention should be a priority. Also, as the conduct of teaching and research within Sakai is an immature field, the ability conferred by this tool to create and make available template sites would be invaluable for new adopters.

The tool which would most enable research and teaching to take place within Sakai with an integrated digital repository would be **Resource Tool Extension**, even if initially this takes place with read-only DSpace access.

Bibliography and **Search** are the library-related tools which end-users most frequently mention as being absent from Sakai. These do not specifically relate to digital repositories, though it presents institutional repositories with an opportunity to present themselves alongside traditional libraries. Therefore, perhaps these tools should be priorities for the Sakai library communities.

Other interventions described above are perhaps best considered a present lower priority, than these identified tools.

Outcomes

CARET are in the process of designing an (initially read-only) Resource Tool Extension for Sakai, and are investigating funding opportunities for Policy Manifestation and Item Semantics work. The provision of LNI [LNI] interfaces within Sakai by this initial Resource Tool Extension work will greatly simplify the adaptation of the Site Archive tool to use repositories, after which it will be possible to implement this tool comparatively simply.

Sakai/DSpace Integration White-Paper

CARET remain in consultation with interested parties on other identified priorities, including ERobot work, and are tracking “grass-roots” and funded developments in Search and Bibliography.

An outcome of this work, separate to this White Paper, is a document describing, in more detail, work-plans for four of the identified priorities: Policy Manifestation, Resource Tool Extension, ERobots and Site Archive.

William Reilly, of DLRG, MIT, has presented a paper [Reilly] giving an overview of repository practice to a recent Sakai user-group conference, to further the aims of engagement between the two communities, including sections based upon an early draft of this report, and an overview of the role of repositories in e-learning, as indicated as desirable in this white-paper. [Reilly] is a valuable source for integrators wishing to further investigate the issues of this paper.

We hope that this paper will encourage other institutions to engage in some of the identified areas.

Acknowledgements

Thanks are due to Patrick Carmichael, Richard Proctor, Raymond Chan, Ian Boston, John Norman, Harriet Truscott, Raad al-Rawi, Jim Downing, Richard Rogers, William Reilly, Larry Stone, and Mackenzie Smith for their contributions towards this report. The following projects, products, and standards were drawn upon at various points, we hereby acknowledge their contribution: LEADIRS, Stellar Images, Molstruc, Twin Peaks, OCW, TLRP, AERS, CREE, Sakaibrary, IMLS@Rochester ResearcherPages, Sakaibrary, ZeeRex, LNI, SRW, OKI DR OSID.

Glossary and Further Information

- **[TLRP]**. Teaching and Learning Research Programme, <http://www.tlrp.org/>. A live project pioneering use of both Sakai as a VRE, and DSpace as a repository for a distributed research body. Repository: <http://www.tlrp.org/dspace>. Contact patrick@caret.cam.ac.uk or richardp@caret.cam.ac.uk for technical details.
 - The TLRP Digital Repository: Interoperability, Exchange and User Engagement. R Proctor, P Carmichael. Presented to the Annual Conference of National Centre for e-Social Science, Manchester, June 2006.
 - Teaching and Learning: Research Briefing, Number 16. IT for Education Research: using new technology to enhance a complex research programme. TLRP.
 - Collaboration, Coherence and Capacity-Building: Using DSpace to Support a Major Social Science Research Programme in the UK. Patrick Carmichael, Richard Proctor. June 2005. <http://www.tlrp.org/dspace/handle/123456789/267>
- **[AERS]**. Applied Educational Research Scheme, <http://www.aers.ac.uk/aers/>. As TLRP, a live distributed research body using Sakai as a VRE, and DSpace as a repository. Repository: <http://www.aers.org.uk/dspace/>. Contact patrick@caret.cam.ac.uk or richardp@caret.cam.ac.uk for technical details.
- **[Boston]**. Meetings between DSpace@MIT, DSpace developers, and CARET teams, as part of CMI project, along with other interested parties. Spring 2006.
- **[CREE]** (Contextual Resource Evaluation Environment), <http://www.hull.ac.uk/esig/cree/>. A project concerned with search in an academic context. Includes valuable focus group research on user attitudes to search.
- **[IMLS]** IMLS@Rochester Researcher Pages. A project which pioneered the personal research page concept, emerging out of anthropological research: <http://tinyurl.com/oqe5e>. This initial anthropological work strongly influenced the earlier

Sakai/DSpace Integration White-Paper

sections of this report.

- **[InstRep]** *Institutional Repositories: Is There Anything Left to Say?*. (http://www.oclc.org/research/dss/pdf/conway_presentation.pdf) Paul Conway Director, Information Technology Services. Perkins Library. Duke University.
- **[LEADIRS]**. The LEADIRS project provides material for managers to aid the development of institutional repositories. Issues identified in the LEADIRS manual helped identify priorities for this document. <http://www.dspace.org/implement/leadirs.pdf>
- **[LNI]**. A WebDAV-based DSpace protocol, available as a patch, primarily focussed on ingest and deposit. <http://wiki.dspace.org/LightweightNetworkInterface>. Developed as part of CWSpace [OCW].
- **[Lynch]** *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*. (<http://www.arl.org/newsltr/226/ir.html>) Clifford Lynch. ARL Bimonthly Report 226. February 2003.
- **[OCW]**. Pioneering project exposing much of MIT's teaching material online for public consumption. The CWSpace project focusses on preserving OCW content in DSpace. CWSpace motivated the development of LNI, and many other technologies, strongly influenced thinking in Sakaibrary. <http://ocw.mit.edu/>
- **[OSID]** OKI DR OSID. One of a series of abstract APIs (OSIDs) developed to encourage interoperation of academic software, the DR OSID concerns integration with digital repositories. <http://www.okiproject.org/>
- **[PLEDGE]** PoLicy Enforcement in Data Grid Environments. This project focusses on repository policy issues, particularly in a distributed data grid context, and strongly influenced the Policy Manifestation and Item Semantics work.
- **[Reilly]** DSpace and E-Learning, William Reilly. Sakai User Conference, Vancouver, Canada. June 2006. Available from wreilly@mit.edu.
- **[Repo]** RepoMMan. The RepoMMan project focuses on the automatic population of metadata, and uses workflow prominently in solution of these problems. Their approach strongly influenced policy manifestation and collections workflow tools, aimed at reducing the metadata-entry bottlenecks for submission. In addition, strong user requirements work described the research process in a formal context, influencing discussion throughout this paper. <http://www.hull.ac.uk/esig/repomman/documents/index.html>.
- **[Sbry]** Sakaibrary. Lead Sakai project concerned with repository integration, emerging from the Twin Peaks project [TwnP]. <http://www.dlib.indiana.edu/projects/sakai/>
- **[StIm]** Stellar Images. An implementation of an image gallery and presentation tool within the Stellar Content Management System: <http://stellar.mit.edu/>. Contact Craig Counterman (<http://web.mit.edu/ccount/www/index.html>) for details.
- **[SRU/W]**. A more recent search protocol than Z39.50. <http://www.loc.gov/standards/sru/>
- **[TwnP]** Twin Peaks. A predecessor to Sakaibrary, a project which addressed integration of search results and references into Sakai authoring tools. This project's clear elucidation of the separate VLE and Repository communities was pivotal in guiding discussions. <https://twinpeaks.dev.java.net/>.
- **[Z39.50]**. A widely adopted search protocol. <http://www.loc.gov/z3950/agency/>

Appendix: Example Interventions

We describe in further detail possible implementations of some of the tools described above, in

Sakai/DSpace Integration White-Paper

the cases where this may be helpful in understanding the the tools suggested. These examples should not be seen as precluding alternative implementations of the identified tools.

Resource Tool Extension

Resource tool extensions are likely to take the form of a number of extra pages within the resource tool, along with modifications to existing pages. These could include:

- The modification of existing property pages to add "send to library" functionality, including selecting collections, supplying/verifying metadata, and similar elements, to the user interface.
- The addition of "status-line" information to items on the default resource-item page, indicating permanent URL, status in repository, and so on.
- The addition of a configuration user interface to folder views, and data items and relations to the data model for folders, to allow default repository actions to take place upon addition of an item to prepared folders.
- The periodic polling of any item semantics information associated with items visible within Sakai, to allow the representation of alternative and composite formats for items, within the resource tool.

Policy manifestation

One realisation of this work would be as follows. Before opening a connection to the repository, Sakai requests the configuration of DSpace through a preconfigured URL. This URL describes the other URLs of the DSpace service but, importantly, also other configuration and policy information, perhaps in the form of RDF and possibly OWL.

Sakai then:

- Modifies the options presented to the user (for example during ingest or search) to reflect the policy manifestation. For example, if a user selects an item to ingest into a particular collection, Sakai could display a warning derived from data contained therein. Or else, a user could be asked to select a particular self-assessed article quality, which would then be sent along with the ingest, to potentially affect DSpace workflow.
- Ensures that only users explicitly authorized (particular EPerson-equivalents) or implicitly authorized (for example, collections where Sakai user-management has been deemed acceptable permission), deposit into a particular collection through a particular ERobot, therefore providing a realisation of Sakai-delegated permissions.
- For folders mounted in the Resource Tool into DSpace collections, alerts the user in the folder view to that collection's policies concerning licensing, dissemination and preservation, derived from the policy document.
- May take advantage of any DSpace extensions or conventions which may be present, knowledge of the presence of which being derived through the policy manifestation document. For example, Sakai could use this knowledge to enhance the display of resources in specialist tools (for example, image-scaling or the appropriate use of thumbnails in a gallery tool [StIm]).
- Configures specialist item generation tools, such as learning object generators or web-page editors to, automatically or by default, use identified distinguished collections within a repository for their specialist output.

Sakai/DSpace Integration White-Paper

ERobots

The ERobot work is oriented around changes to the DSpace data model. First, each ERobot would be a valid EPerson. In order for authorization to be given, it would be necessary that the corresponding permissions be present within the EPerson record for the ERobot. A second identity, that of an individual on whose behalf the robot is operating could be explicitly indicated and managed by DSpace, through a mapping between remote identities and DSpace EPersons.

The robot provides the mapping either from within the agent connecting to DSpace (be it Sakai, or otherwise), or within a configuration-time mapping within DSpace. Once this mapping has been established, DSpace manages authorization on behalf of the robot, requiring permission to be granted to both *Robot* and *User* EPersons in order to proceed. Through parsing policy manifestation documents, delegated permissions managed from within Sakai can be mapped to EPersons representing classes of individuals, to ease authorization management within DSpace.

It could be established as a configuration-time option within DSpace whether an unknown EPerson presented by the robot to DSpace is created, or if permission is denied to these previously unseen users.