HarvestRoad.

ETD 2005: Evolution Through Discovery
8th International Symposium on Electronic Theses & Dissertations

Learning Object Repositories, Content Management, Publishing and Library Systems Merge to form New Generation Institutional Learning Resource Management Systems

Disclaimer: This paper represents the personal views of the author based on industry experience with his company HarvestRoad Limited and has been prepared for a speaking presentation to the Electronic Theses and Dissertations Conference September 2005 seminar. It is not intended as a scholarly publication and the author's observations are not supported by any scientific evidence. On that basis, readers are free to use this content providing reference is made to the author and this disclaimer.

Grame Barty
Managing Director & Founder HarvestRoad Limited

Keywords: Super Discovery, Federation, Super Distribution, Learning Object Repositories. Digital Repository, Content Management System, ILMS, Library Systems

SYNOPSIS

Fundamentally our work and living environments have changed forever and there are three main drivers that continue to push us forward – there's no going back. These drivers include:

- Technology factors

- Societal factors

- Human logic factors

How Does That Impact Information Scientists?

What we previously thought of as the challenges of the internet – and benefits – are only just starting to become obvious now.

We have to create a capability and service that deals with the 'always on' environment. Students and researchers want access to institutional repositories 24/7 and you are measured on your ability to deliver that service.

Until now we have been trying to work with digital capability but using analogue processes. This is why, in some instances, we have a mixed environment of applications that are only doing a number of things half well. We are 'talking the talk' but not 'walking the walk'. We can also occasionally observe that due process for an institution's intellectual capital – such as copyright compliance – is being completely ignored by users, and this is not acceptable.

From our perspective, some of the tool sets currently available have peaked in a technological sense for an 'always on' service. But it can't be a case of just replacing a widget with another widget. We need a milestone change – what Ed Walker (CEO of IMS) refers to as a 'moon shot' in technology and capability.

It is the fundamentals of that *moon shot* requirement that I refer to.

Moon Shot Fundamentals

The moon shot fundamentals that are impacting the research and educational information science environment include:

- What new services are required of us and what are their priorities? The author refers to super discovery and super distribution in this regard.

- Creating new methodologies and processes for managing a wide array of content from multiple sources.

- Understanding the concepts of granularity and reusability.

- Understanding an institution's role within the global publishing exchanges.

- How to operate in a multi-exchange environment.

What's Going to Happen to My Current Environment?

The author claims there is a mismatch in the roles of the different environments that create, manage and deliver content, and what is being espoused today as the 'eLearning solution'. This has stalled progress in distributed information science and a circuit breaking capability is emerging.

The roles of publishing content and managing institutional repositories are in the hands of experts. The next generation deployment of this functionality will require an adaptation of the current perceived roles within institutions and a revision of the appropriate tools to undertake this adaptation over the next 10 years. In particular for information scientists, we can see dedicated enterprise-wide repository roles being assigned. They will deal with copyright, meta data and access across the institution to formulate and enforce policies and methodologies.

Summary

More people want more information, more of the time, in more formats, in more locations, for more uses, and this information is to be managed and maintained globally for a lifetime.

It is possible as a result that Learning Content, Publishing and Library Systems will merge to form new generation institutional repositories called something like **Learning Resource Management Systems**. They will provide very high, continuing levels of enhanced discovery, and the ability to manage and federate granular objects across multiple repository and business environments.

1. INTRODUCTION

It's now late 2005 and nearly 10 years since the digital world truly began with general deployment of internet protocol and digital cellular networks creating 'virtual networks' to publish and share information in an unheralded way. The browser gave us a multimedia webtop to displace private networks and the distinction between an intranet and extranet became blurred. A portal was still a funny medieval Latin word, people had trouble pronouncing web addresses, and workers were just getting the hang of emailing instead of faxing, but the idea of an institution's information being universally accessible – and valuable – was starting to gain ground.

Fundamentally our work and living environments have changed forever and there are three main drivers that continue to push us forward – there's no going back. These drivers are:

- Technology factors

- Societal factors

- Human logic factors

1.1 Technology Factors

Apart from the well-understood Moore's Law and the advance of microprocessor technology, one of the major impacts on information science since 1995 has been peer-to-peer technology. Surely Napster will rank along with Johann Gutenburg's printing press of 1450 as *the* major publishing innovation of the Internet. It has single-handedly changed the way we access content (desktop and mobile media players), how we value content (individual songs on an entire album compiled to our personal taste), and the way we pay for content (per unit electronically).

In the past we also thought of content creation as a specialist task but new generation laptops and mobile devices have brought content creation to the layperson. Think of the way you now manage photos, send emails instead of letters, manage home movies, create multimedia presentations or even my own creation of this document – there are no more typing pools, just as there are no more telephone switch boards. Combined with high-speed Internet service, or broadband and the elimination of storage as a technology issue, we have also over-ridden any weaknesses in applications so we can do even more amazing things.

Technology has created – and continues to create – a simplified, convenient method of producing content and we have adapted our business and personal processes to that – with surprising dexterity I might add!

1.2 Societal Factors

We are improving our ability to use digital tools in our everyday work environment, and we are also using these same tools to enhance our private lives. As an upside we now demand that

'our' information must be instantly available and served where and how we want it – on the net, as a text message, RSS feed, or CNN style as it happens! Gone are the days when our only information came from the 6:00pm news or the morning newspaper or the local library. We are 24/7 users, unrestricted by location, or method of access and we have an appetite that must be fed!

That's the good news.

The *bad news is we can also be contacted 24/7* and we have to develop personal control mechanisms to 'opt out' or we will end up in the 'crackberry' category of information addicts.

The point here is that communicating and the publishing and sharing of knowledge is a fundamental human activity – we *have to do it.* The only major change we are seeing here from 10 years ago is the rise of immediate, unfettered 'personal content creation' and 'personal broadcasting' through things like discussion groups, podcasts or blogs. In some instances we are amazed with the silliness of these activities but occasionally, for example in a crisis, they can become the most critical methods of communicating and publishing information. For example there were the bloggers in Iraq and Iran and, with no disrespect intended, the mobile phone user who filmed and sent video images of the devastation of the London Underground bombing.

The other interesting aspect of this is an audience's ability to accept this content for what it is regardless of its quality or 'slickness' – *timeliness and accuracy overrides this*. The exception here in my mind is reality TV because this is just entertainment (with my apologies to professional entertainers) but the 'must know' information principle is the same – we want to know 'what happened today' or 'what's new'.

1.3 Human Logic Factors

I can make no claim to understanding the science of human logic but I do have two teenage children and I know what we all know: their generation is wired differently.

From within their environment they multitask, they collaborate, they communicate – all at the same time. They are highly unlikely to sit and watch TV, preferring to be stimulated and challenged with games. (Add the games to a network and you have created seventh heaven for these kids!)

They use digital tools like we used to ride a bike – without thinking. Perversely they can even bully each other digitally! From compiling songs, to bus routes, to concert tickets to homework material, they are the 'now' generation of information science.

In his paper 'Beyond the Horseless Carriage'[1] Gerry White CEO Educationau refers to J.Boscos observation that:

The technology of the book, in disseminating knowledge, has obscured the disorderly process of developing knowledge (sic). Whereas information technology is closer to the way that knowledge is created.

This begs the question: Are we approaching pedagody and its use of ICT from a fundamentally flawed basis and therefore if we haven't created an educational environment to satisfy this market's requirement, how and what are we teaching them or preparing them for? Learners after all are going to determine which services are going to be the most successful but nothing can change if the logic associated with new generations thionkling patterns is mismatched.

---

[1] http://www.educationau.edu.au/papers/horseless_carriages_GW.pdf

2. HOW DOES THAT IMPACT INFORMATION SCIENTISTS?

I believe that the real challenges of the internet – and benefits – are only just starting to become obvious now.

This means we have to create a capability and service that deals with the 'always on' environment. Students and researchers want global access to institutional repositories 24/7 and you should be measured on your ability to deliver that service.

The basis of my presentation today then – and in fact my business – is that up till now we have been trying to work with digital capabilities but using analogue processes. This is why in some instances we have a mixed environment of applications that do a number of things only half as well as they need to. We are 'talking the talk' but not 'walking the walk'. We can also occasionally observe that due process for an institution's intellectual capital – such as copyright compliance – is being completely ignored by users and this is not acceptable.

From our perspective, some of the tool sets that are currently available have peaked in a technological sense for an 'always on' service. But it can't be a case of just replacing a widget with another widget. We need a milestone change – what Ed Walker CEO of IMS refers to as a 'moon shot' in technology and capability.[2] It is the fundamentals of that *moon shot* requirement that I would like to refer to today.

2.1 Moon Shot Fundamentals

The moon shot fundamentals that impact on the research and educational information science environment include:

- Determining which new services are required of us, and their priorities

- Creating new methodologies and processes for managing a wide array of content from multiple sources

- Understanding the concepts of granularity and reusability

- Understanding an institution's role within the global publishing exchanges

- How to operate in a multi-exchange environment

2.2 New Services

The two most important service features in Generation 2 Internet access relate to 'super discovery' and 'super distribution'.

Super discovery is really based around the fundamentals of federated search and not just through meta data harvesting of repositories like Merlot but the highly advanced functionality being promoted through developments like the US Department of Defence CORDRA standard for the registering and handling of objects across multiple repositories using a common services infrastructure, as shown in the diagram below.

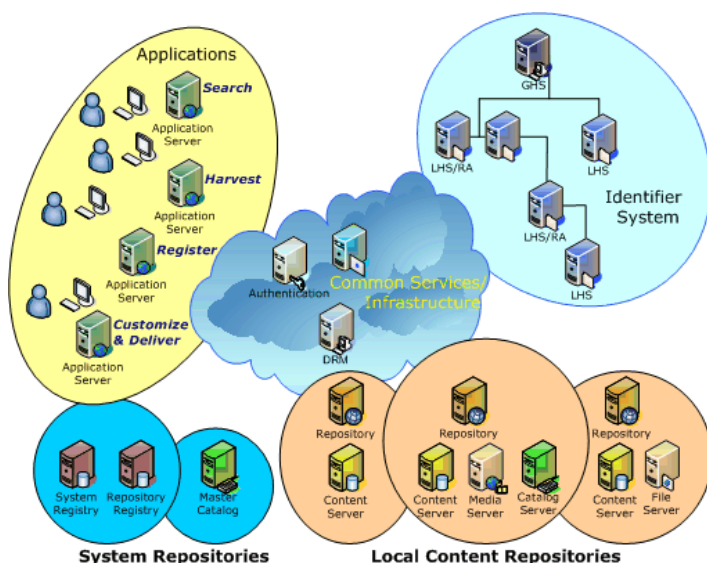**http://hive.harvestroad.com.au/hive/quickfetch.html?VIEWTYPE=Original&IDTYPE=Alias&ID=24828**

Figure 1: CORDRA, Content Object Repository Discovery and Registration/Resolution Architecture, ©Learning Systems Architecture Laboratory.

Another example is the Open Knowledge Initiative (O.K.I.™) which is led by the Massachusetts Institute of Technology. It is a community effort to improve interoperability among applications and the enterprise system services on which they depend. This was recently displayed at IMS AltiLab in Sheffield UK with the main parties being MIT[3], Apple Education, HarvestRoad and Guinti Labs and is the basis of Ed Walkers 'moon shot' comments.

These developments are important because they create a virtual plug-and-play between applications and repositories that can only benefit the many 'compete and co-operate' agreements being developed by and between institutions globally. They give users universal access to prescribed content, on a permissions basis, and rendered in a manner appropriate to the inter-institution sharing agreement by including copyright or branding rights. This goes well beyond the 'general' discovery functionality of a Google-style search engine because content is qualified by repository and access rules, and can be repurposed.
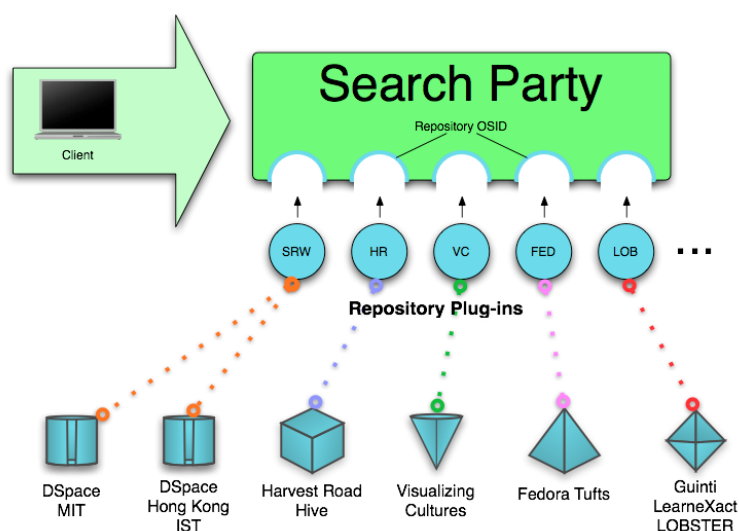

Figure 2: Integration between the MacLearning Environments Search Party Application and HarvestRoad's Hive® - Scott Kahn, Verbena Consulting and Scott Morris Apple Computer - 13 July 2005

---

[3] http://www.okiproject.org/library/doc_33.html

The difference between CORDRA and OKI can summarised as:

- OKI provides a federated search on the existing metadata at the repository site.

- CORDRA provides federated harvesting of metadata into a central repository where it is searched. The user is 'connected' to the item via a URI - unique identifier.

You could say that with OKI you have no idea of the metadata that is available for searching. In CORDRA this will probably be part of the policy information that is stored centrally. The really interesting thing about this is that the community decides on its own metadata schema and usage. As I understand it, this is the opposite of the Library of Congress controlled space to which all publishers comply through ISBNs etc. The question then is how will you maintain an authoritative catalogue that harmonises the usage of different communities? Pretty tricky!

Super distribution is the effect of creating objects for point-to-multiple-point distribution – perhaps millions of copies of the same object that are perfectly reproduced and instantly available. The use of objects in a granular form allows simplified assembly and reassembly. In the case of course packages using a standard like SCORM and assembly packager like JISC's open source RELOAD tool, objects and their versions and rules remain within the repository and only the links to the objects are distributed. In this case they are distributed in the form of a standards-based Manifest File.

The editing and re-editing of courses then can be simply achieved by modifying the Manifest Files or permissions associated with an object's access. This gives you:

- Greater continued control over the object and its versions from within its 'home' repository

- The ability to rapidly reassemble courses to meet new requirements

This is much smarter than trying to manage inflexible zipped course packages. Control of the content can be quickly lost when reassembling courses, and maintaining content becomes impossible because the most useful objects are duplicated into many different places. Changing a single word on a page can become an endless hunt for the tens of hundreds of duplicated instances of that single page.

To be most effective, objects must be held in institutional repositories that are independent of delivery systems such as Learning Management Systems. This way the rules for managing content can be enforced and as shown in the previous slides, content can be integrated in wider inter-repository hook ups.

2.3 Creating New Methodologies

The increasing dilemma that you have as information scientists is: what to capture, why to capture it, and how and who pays for it!

Technology is now giving you unparalleled flexibility in super discovery and super distribution. Do you have the processes to cope with this? Even worse, what if the rate of change in the next ten years is even greater than the last ten years? How will you keep pace?

So what does that mean for your environment? Quite a lot actually….

- How will I categorise information I never worried about before– a blog, a podcast?

- How will I decide if information is important enough to categorise?

- How will I catalogue volatile materials or materials whose value is extraordinarily high but only for a very short period? For example, activity in an election period or voter polls has a different value after the election is decided.

- How will I store and manage information? When should it be stored as compound or granular objects?

- Where will I store information? In an LOR, an LCMS, a library system, a shared open source repository, an EDMS? What do I have in place now? Can I access legacy content across internal repositories?

- How will I make information available for reuse - who can access it? Which version and when? Who helps me decide this? Do I need to consider cross-repository requirements or discovery on a federated basis, and what arrangements are in place for that?

- How will I ensure the integrity of the information and its use - do I have digital right management tools in place? Can I protect against plagiarism? Is the information being used for its intended purposes? Can I resell it if I own it?

- How will I ensure preservation of material for current and future audiences? What roles do content creators have in this?

3. WHAT'S GOING TO HAPPEN TO MY CURRENT ENVIRONMENT?

I believe there is a mismatch in the roles of the different environments that are used for creating, managing and delivering content and what is being espoused today as the 'eLearning solution'.

When referring to course packages, eLearning is simply the packaging and delivery of digital objects in a sequence that is assessed. Scheduling and packaging of courses are pedagogical issues that exist regardless of whether it is e-learning, blended learning or face-to-face learning.

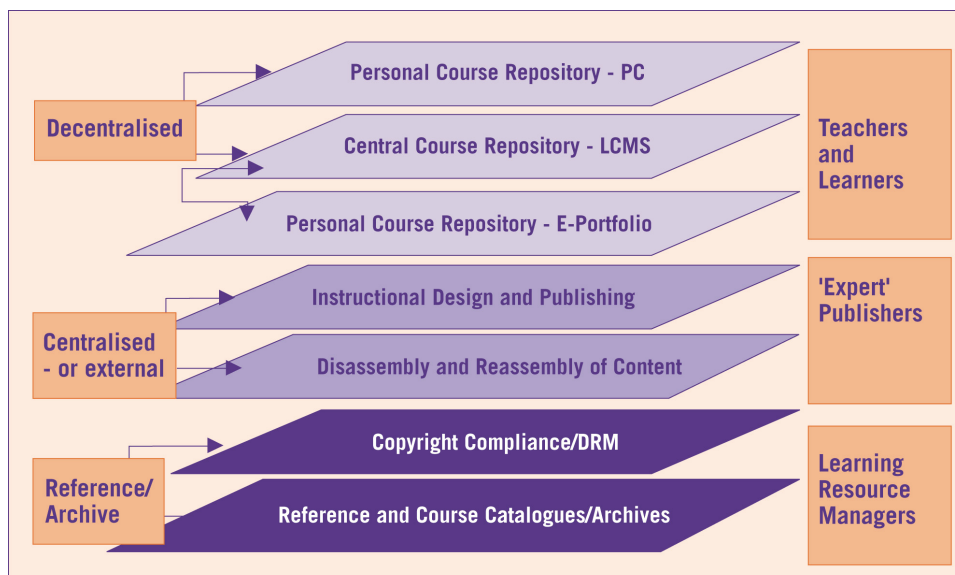The current educational environment in most institutions looks like this:

Figure 3: Educational Institution User Sectors

In the case of eLearning, most responsibility today is passed to the academic who creates and publishes online courses. This is fundamentally flawed. We are asking the wrong people to do this job. For example I know how to use an Excel spreadsheet but I wouldn't dream of creating the financials for the HarvestRoad Annual report nor do I personally make changes to the HarvestRoad website.

The role of professional institutional publishing belongs to publishing experts. The role of managing institutional digital resources also belongs to experts: you as information scientists.

The 'e' bit doesn't have some magical formula – it is just a convenient means of delivery in some instances. The 'e' bit offers the potential to extend the boundaries of the course to discover and use resources that would not have been possible 10 years ago.

Therefore we need to distinguish between a minor drag-and-drop capability for simplified course delivery by teachers, and the sophisticated management and control functions required by expert publishers and learning resource managers in creating institutional learning material.

This group will also appreciate that not all institutional intellectual property is about course content. It can also be processes, exam papers, research , conference proceedings, theses and so on.

The area of copyright or more broadly intellectual property is a complex and difficult area because the law is modelled on a print based world and is having trouble coming to grips with a digital world. The means of production and distribution between the two are quite different and the current punitive method of enforcement may longer be appropriate. An alternative may be to focus on principles of fairness, access and distribution. In an 'always on' world current copyright laws will not be applicable vis. Creative Commons, and new business models will emerge.

We should not accept any system that restricts the type of intellectual property that has value to an institution.

I believe that the current trend of integrating LCMS functionality into an LMS is misguided because of its erroneous assumption that all teachers or academics want to create course content. It also places limits on editing, access, reusability and digital rights and assumes that only course content represents intellectual capital. And of course openness is not necessarily a benefit for a vendor's business model because vendors often prefer to offer a one-stop solution rather than integrate with best-of-breed.

I also believe that the e-Portfolio requirements widely touted these days are actually a function of student administration, not an LMS or LCMS. The ePortfolio should also be offered as an independent service to ensure its lifelong learning validity and usefulness to an individual – not only within a single institution.

The principle function required for the teacher and learner environments is super discovery and not publishing.

Institutions should consider moving to an independent institutional repository that is capable of federating internally and externally and delivering objects to a number of systems on a controlled basis. This diagram from OCLC is a nice representation of the central management and distribution role of an institutional repository.
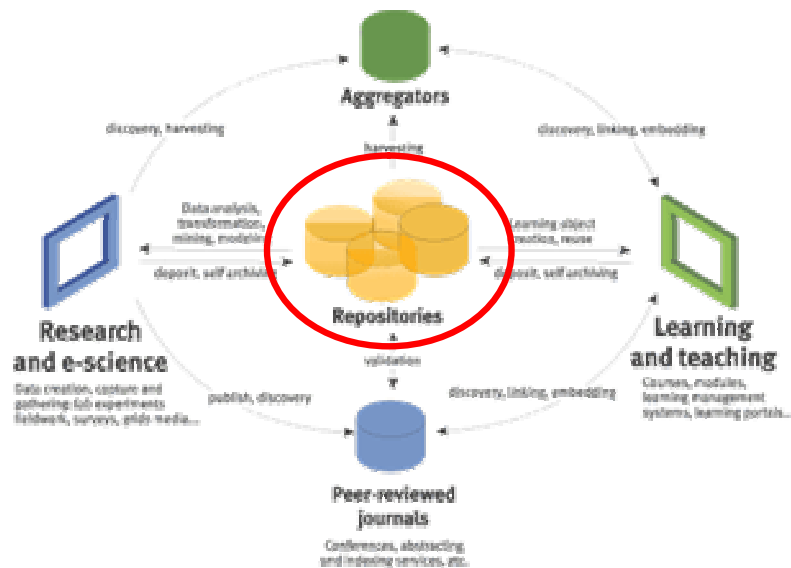
The repository model

Figure 4:"Pattern Recognition: The 2003 OCLC Environmental Scan"
http://www.oclc.org/membership/escan/toc.htm

4. SOME MORE TRENDS

Apart from institutional repositories we see these main trends:

- Continued growth and value of open source solutions and their integration with best-of-breed commercial off-the-shelf solutions – but this is not an either/or scenario. HarvestRoad's recent announcement of its Commercial Partner Affiliation with the SAKAI community is a good example of this trend.

- Emergence of the concept of a 'learning supply chain' with content or objects flowing down the channel and data on results and reporting flowing up the channel. This will include:

  - Mainstream publishers are looking for a Napster style model to directly reach students and teachers by object level – there is a lot of tinkering going on behind the scenes but they are still conservative in their approach.

  - The emergence of global learning exchanges and flexible voluntary digital rights licenses. New services may even offer discounts for 'out of season stock' or book and journal substitution incentives. Global exchanges are worth watching because eventually your institution will have to decide what role you want to play with one or more of them, and at what cost or benefit.

  - The introduction of repositories to seamlessly manage technical content as well as learning content. This is largely being driven by Defence (15% of aircraft failures are a result of out-of-date maintenance documentation[4]), but will eventually move into education publishing circles as higher education and further education is outsourced to help deliver this content.

---

[4] **http://web.nps.navy.mil/~gazolla/files/dallas.pdf**

- Learner-driven learning where the role of instructors is more about learning outcomes and directing students to course resources. Instructors allow learners to learn what they want, how they want, when they want. They provide accelerated or delayed learning depending on the needs of the student rather than on a course timetable dictated by an annual calendar of events.

- Business institutes now see themselves as Distance Educators with lots of IP and a value proposition that they own and can present uniquely to their members.

5. SUMMARY

More people want more information, more often, in more formats, in more locations, for more uses, and this information is to be managed and maintained globally for a lifetime.

It is possible that Learning Content, Publishing and Library Systems will merge to form new generation institutional repositories called Learning Resource Management Systems. These repositories will provide very high, continuing levels of enhanced discovery and the ability to manage and federate granular objects across multiple repository and business environments.

Finally the role of publishing content and managing institutional repositories is in the hands of experts. The next generation deployment of this functionality will require an adaptation of the current perceived roles within institutions and a revision of the appropriate tools to undertake this over the next 10 years. In particular we can see dedicated enterprise-wide repository roles being assigned to information scientists. The information scientists will deal with copyright, meta data and access across the institution, and formulate and enforce policies and methodologies.

That concludes my presentation and thank you for your attention.

29th July 2005

gbarty@harvestroad.com.au

http://www.harvestroad.com.au

Bio Details:

http://www.harvestroad.com.au/investor/board.cfm#grame