Supporting the Uptake of Cyberinfrastructure in the Social Sciences and the Challenges Faced.

Paul Dolby[1], Nick Pearce[1], Adrian Fish[1], Ties Van Ark[1], Rob Crouchley[1] and Rob Allan[2]

[1]Lancaster University, Lancaster, England, LA1 4YT
[2]Science and Technology Facilities Council, Daresbury, England, WA4 4AD

Email address of corresponding author: p.dolby@lancaster.ac.uk

Abstract: In this paper we provide arguments as to why the tools and resources that ReDReSS offers provide an effective and ultimately unique online learning environment for social science researchers. Learning content is provided in a variety of formats including audio-visual recordings of lectures and workshops. This is complemented by a range of classification and search tools to facilitate the construction of "learning ladders" for self motivated researchers wishing to acquire new skills. Web 2.0 style collaboration tools are included to give a rich experience and to stimulate community engagement. Statistics indicate the popularity of the resources offered by the project, and demonstrate the importance of the ReDReSS project and the services and resources it has to offer.

Introduction

Social science research is often carried out by small research collaborations or frequently by individual researchers. Datasets are often private and much smaller than in the physical sciences, somewhat diminishing the need for cyber-infrastructure and large scale collaboration as compared to the physical sciences. High-performance computers are only recently becoming considered for some of the larger challenges, for instance in quantitative statistical modelling. The benefits of cyber-infrastructure are therefore not so obvious when considered in the context of the social sciences and researchers have to learn new skills in order to use e-Science technology.

We argue that cyber-infrastructure is of benefit to social science researchers and offer as evidence the large numbers of projects now being undertaken by early adopters of e-science technology in the UK. We believe that a lack of awareness of the benefits of cyber-infrastructure amongst social science researchers is a major factor hindering its uptake.

In order therefore to raise the awareness and to facilitate the discovery of the benefits of cyber-infrastructure in the social sciences, a project was commissioned in 2003 by JISC [2] (The Joint Information Systems Committee) and the ESRC [3] (Economic and Social Research Council) in the UK titled, Resource Discovery for Researchers in e-Social Science" (ReDReSS [4]). This project is carried out by staff at the Centre for e-Science at Lancaster University. It achieves its aim by utilising the knowledge and experiences of social science researchers that have already taken the first steps in implementing and trialling cyber-infrastructure in their research processes, the early adopters. The central part of this project is

its user-centric e-learning based tools aimed at social science researchers like these who are self-motivated to develop new skills.

In this paper we will describe the tools, services and resources that have been developed as part of ReDReSS to aid social science researchers in discovering the benefits of cyber-infrastructure as well as enabling them to learn more when required. In particular we will talk about the Learning Space Catalogue, a specialised repository (database) of training and awareness raising resources (such as use cases from various projects utilising cyber-infrastructure). We will then proceed to describe the collaboration tools developed in conjunction with the JISC funded VRE demonstrator project. The tools provide a collaborative learning environment. We will then cover the resources offered within the Learning Space Catalogue repository, in particular our audio-visual content. In the penultimate section of this paper we will discuss possible future developments of the ReDReSS project, based on experiences so far. This includes the development of Web 2.0 tools. In the final section of the paper we present statistics that convey the popularity of the ReDReSS project and therefore its importance. Throughout this paper we will convey the challenges that we have had to overcome along with the solutions which we have adopted. However there are many changes still to be overcome, these will be discussed.

Learning Space Catalogue and Repository

The Learning Space Catalogue (from now on abbreviated to LSC) is the flagship tool offered by the ReDReSS project. It is essentially a database of learning materials, supported by a custom built interface. This interface is the key to the uniqueness of our repository.

The LSC is designed specifically for a social science audience. This is unique compared to many other e-Science repositories which are often aimed at many disciplines such as the physical sciences where use of high-performance computers is more commonplace. Aiming specifically at a social science audience provided us with a challenge. We needed to provide access to learning material specifically aimed at the social science context which was easy to find and intuitive to access. At the start of the project in 2003 open source repositories such as Fedora [5] were in their infancy and did not provide the specific requirements to meet the need of social scientists. We therefore decided to produce our own repository, the LSC.

Initially the LSC was based upon an XML file containing metadata for each record in the repository. A simple category based interface was then produced to search and display the records. The aim of the repository was to contain links to learning materials sourced mainly from the Web in order for us to quickly build up a collection of on-line materials. There was no need for content management as the majority of content was hosted externally; therefore the XML format was ideal. However it quickly became clear that using a database to contain the metadata provided for much easier manipulation of the data. Also we could use the advanced features of the database to manipulate the data much more easily. We then started on an interface that could utilise the advanced functionality of the database to provide social scientists with an intuitive and powerful tool to find learning resources customised for their personal requirements.

Each learning resource in the catalogue is described using metadata in Dublin Core format [6]). This metadata format was chosen for its simplicity and to allow us to comply with open standards. A subset of tags was chosen. These include identifier, creator, publisher, format and description. These tags are applied to learning resources found on the web which we

deem suitable for social scientists, timely and of sufficiently good quality. Each resource had an associated description which is either written manually or obtained from the resource itself. This means that all resources within our repository are reviewed. Each resource is then listed in one or more categories to allow the easy discovery of resources. At present the categories are organized into 18 categories. These categories are listed down the right hand side of Figure 1.

These categories provide easy access to the resources at a basic level of navigation. This picture also shows other features mentioned in the paper.

Another powerful feature of the repository is the user interface. This provides easy discovery of resources by providing advanced filter and sorting options, unlike the basic search options offered for many other repositories. It also offers alerting features such as RSS feeds, which can be understood by RSS readers. There is also a book-marking feature described below. The interface offers access to the categorised approach in organising the resources but also offers the flexibility of more general searching and sorting by author, title, subject or any of the metadata items.

Due to the fact that this software is developed in-house, we are able to add new features easily, in particular features related to our audio visual content, which will be discussed later. Web 2.0 tools are also being incorporated to enhance the user's ability to discover, use and share resources or to contribute in various ways. These are aimed at community-building will be discussed later in the paper.

Most content at present is produce in-house or obtained manually from the web. It is important for the continuation of ReDReSS to be able to continually add content to the repository. For this reason work has begun on automatic content harvesting. This will be achieved using a number of techniques. These include the implementation of standards based methods such as OAI-PMH or Z39.50. This will enable us to harvest relevant content from other standards based e-Learning repositories. Using these standards, we can also allow other repositories to harvest our content.

Another technique we plan to implement is automatic harvesting from the web. We will not go into the details here but essentially this technique uses a Bayesian learning algorithm to analyze content which has been manually added and classified into the categories discussed earlier. This then produced a reference for the harvester that will enable it to gather suitable content from the web and place it in the correct category.

One last message we would like to convey concerning the LSC. The software behind the catalogue is constantly being improved with new feature. We do not have release cycles. Once a new feature has been produced it will be made available. We hope to provide an alerting tool to alert users to new features in the near future.

Use of Audio-Visual Content

The initial aim of the project was to populate the LSC with resources manually harvested from the Web. We also aimed to commission original material fro expert authors. The commissioning of material however proved to be very difficult. Authors often did not have enough time available to complete the material, especially due to the upcoming Research

Assessment Exercise in the UK[1]. Any educational content created for the ReDReSS project could not be included; hence there was no incentive to produce material for the project. We therefore adopted an alternative approach to supplement the material that we gathered from the Web.

Since generally it is a requirement for researchers to present research findings, tools and services developed and also to teach, we decided to capture them doing these activities. This would provide us with learning materials at no extra time cost to the authors.

Presentations given generally have three main components, audio from the presenter, visual images of the presenter and the (usually PowerPoint) presentation, which may contain dynamic audio and various visual components. We wanted the ability to be able to capture all of these components and present them in a way which could be played back by the on-line user of our resources as a whole. Each component provides valuable information, which can sometimes be lost when the components are not combined in the original sequence.

There are many software solutions available to achieve what has been described above. At the time of starting, the best solution that we found was supplied by a company called Impatica [7]. The software, Impatica OnCue, allows the easy synchronization of recorded audio-visual content with a PowerPoint presentation which can include animations. The software provides easy navigation between slides. It also provides a powerful feature not found in any other software, the ability to synchronize a transcript to provide subtitles. This is essential in meeting accessibility requirements[2]. This transcript is also fully searchable from within the application. This enables users to search for specific key words and jump direct to them in the presentation. Another advantage of this software is that it is Java-based and therefore platform independent. This enables us to reach as many users as possible.

The process of producing Impatica on-line multi-media presentations is also relatively simple and quick. A one-hour presentation takes approximately one day to process depending on the quality of the raw material. The aim of ReDReSS is to produce reasonable quality material, but this quite often is not broadcast quality due to the environment the presentation is being filmed in, e.g. a noisy lecture room or training lab rather than a TV studio.

Because of the transient mature of much of the content (itself a challenge as e-Science technology is rapidly changing), we do not aim to provide a TV professional services, but due to the editing process we do produce resources that are of better quality than live Web casting and are retained in their original context (this has been found to be extremely important for learning). The processing time does however increase if transcripts are produced as we have so far found no reliable automatic transcripting tools. We produce transcripts at present as an additional feature when appropriate.

Transcripts are very important when putting video content on the web. They are not only valuable for disability purposes but they are important for search engines. Video content has no text associated with it; therefore search engines are unable to determine the content of the video so will not index it. By providing the transcript and an associated link into the Impatica material, search engines will be able to present users with the content of the presentation. We are now making the transcript available to search engines and users when they have been

[1] RAE is a university department rating exercise based on research outputs.

[2] As enforced by SENDA: the Special Educational Needs and Disability Act (2001)

produced. We also hope to provide features in the LSC that take full advantage of the transcripts.

The ReDReSS Future: Using Web 2.0 to enhance the ReDReSS Learning Experience

Tools similar in nature to Web 2.0 services are being added to ReDReSS to further enhance the learning experience. These provide a means for users to share information about ReDReSS content and to contribute to it. The tools being deployed and/ or developed include:

- Threaded discussion Forum
- Blog
- Wiki
- On-line Chat
- Resource repository for new contributions
- Feedback and rating tools
- Hardness rating
- Book-marking and annotation tool
- Acronyms

Several of these tools, namely Discussion, Blog, Wiki, Chat and Resources are built in as collaboration tools from a Sakai portal framework. Sakai [8] was developed as an open-source Collaborative Learning Environment and has the additional feature of providing worksites for group learning or tutoring and dedicated e-Learning tools. Other collaboration tools (such as Agora [9]) are also available. The other tools mentioned above act directly on the ReDReSS repository content and thus can be viewed via the Learning Space Catalogue. We provide more details of two of these: the book-marking tool and the hardness rating tool.

Book-marking

Each entry in the LSC now has the capability of being added to a list of bookmarks by clicking on a button. Items can be added to or removed from the list of bookmarks. This enables the student to build up a preferred set of learning material to work through or download, this is what might be used to compose the “learning ladder” if it is sequenced. A tab in the LSC then enables to “view content” giving assess to the full resource list, or “view bookmarks” giving access to just the personal content selection. Bookmarks can be saved and shared, but this currently simply works with the HTML. An XML listing and tools to export and import bookmark lists is being developed.

Hardness Rating

Many items in the LSC have been rated in terms of hardness or complexity. The complexity is determined by natural language processing using a Bayesian learning algorithm which determines the number of long or unusual words. This is however only possible for repository content which is in a machine-readable format, such as PDF. We have not found any tools for rating PPT or DOC files. Interestingly the Introductory section in the LSC has both the easiest and hardest material, the latter being identified because it introduces a number of rather complex concepts.

As well as introducing these web 2.0 tools, we are also aware of the increasing usage of mobile devices. We are aiming to produce "Lite" versions of our audio-video content. These will be cut down (possibly video only) versions of our Impatica presentations, which are suitable for distribution on mobile devices. We are also aiming to make these available via Podcasts.

ReDReSS Popularity

In this final section we would briefly like to present some statistics that demonstrate the increasing popularity of the ReDReSS project's resources. This we hope will demonstrate the importance of the resources the project has to offer. Although as of yet we still need to ascertain the make-up of our audience, it is reasonable to assume that most of our audience will be social scientists or people from other disciplines working on a social science project. We assume this since nearly all of our resources are social science related.

In figure 2, we show the results of an analyses of the web statistics, which have been collected since February 2006, show compared with the same months in 2007. These statistics show a significant increase in the number of monthly unique visitors from 2006 to 2007. These are the unique IP addresses registered in the server logs each month.

A further analyses of hits, pages and bandwidth on the ReDReSS website during the period from the 1st of January to the 19th of August 2007 (figure 3) showed that the majority of users are of the service are located in the United Kingdom (most of which originate from IP addresses related to a Higher or Further Education institution) and a strong interest from other European countries. Although the number of hits from other continents including the US is significant, the use of bandwidth is small, which leaves us to believe that there is a limited interest from these countries in the presentations. Increases in bandwidth usage provide a good indication of increased usage of our audio-visual material.

Conclusions

We have provided details of an e-learning approach being developed in the ReDReSS Project: Resource Discovery for Researchers in e-Social Science. This will significantly enhance a user's ability to learn from the resources in the Learning Space Catalogue and associated repository. This "Learning Ladder Approach" allows resources in the Learning Space Catalogue to be structured using various semantic and Web 2.0 style tools to allow a stepping stone approach to learning. This, coupled with efficient content harvesting tools and collaboration tools, provides a powerful cyber-infrastructure e-learning platform aimed at self-motivated social science researchers wishing to acquire new e-Science skills.

We have also shown that the ReDReSS project provides increasingly popular resources as indicated by the increasing usage of the ReDReSS site. This is important as it shows the need for projects such as ReDReSS.

Figures

Category	Externally Hosted Resources**	ReDReSS Hosted Resources**	Total Resources**	Oldest Added Date	Latest Added Date	Lowest Difficulty	Highest Difficulty
Introductory Material and Support	38	50	88	2005-12-20	2007-09-03		
Motivation and Background	9	8	17	2005-12-20	2007-06-11		
Examples from e-Social Science Projects	22	27	49	2005-12-23	2007-06-26		
Agenda Setting Workshops	1	16	17	2005-12-22	2007-03-09		
UK e-Science Programme	12	19	31	2005-12-20	2007-08-24		
Collaboration	21	23	44	2005-12-23	2007-09-12		
Data Management	14	22	36	2005-12-23	2007-08-15		
Semantics, Knowledge and Information Management	13	28	41	2005-12-20	2007-08-20		
Databases	9	8	17	2005-12-23	2007-05-15		
Social Shaping	12	8	20	2005-12-23	2007-07-23		
Visualisation	7	5	12	2005-12-23	2007-08-12		
Software Development	6	13	19	2005-12-23	2007-09-12		
Portals and Related Technology	5	43	48	2005-12-23	2007-08-14		
High Performance Computing	11	16	27	2005-12-22	2007-09-12		
Security and Confidentiality	10	12	22	2006-01-18	2007-08-15		
Grid and Web Services	44	61	105	2005-12-23	2007-09-12		
Network Technology	5	5	10	2006-01-19	2007-09-12		
Social Science with the Grid	17	15	32	2006-05-09	2007-09-03		

Figure 1. The ReDReSS content categories.

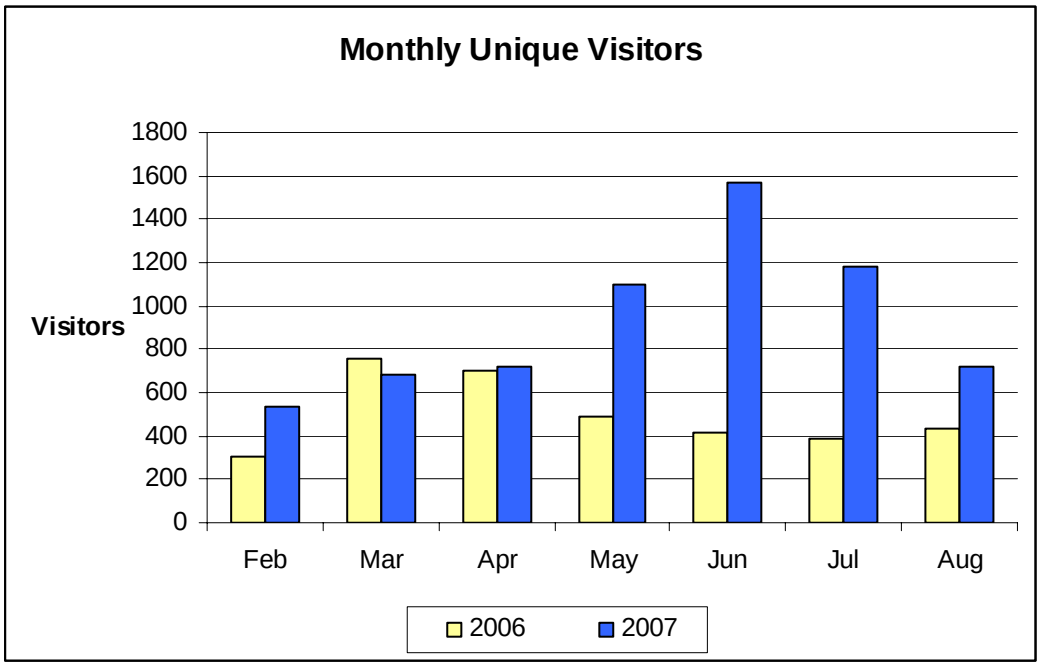

Figure 2. Monthly Unique Visitors to the ReDReSS Web Site.

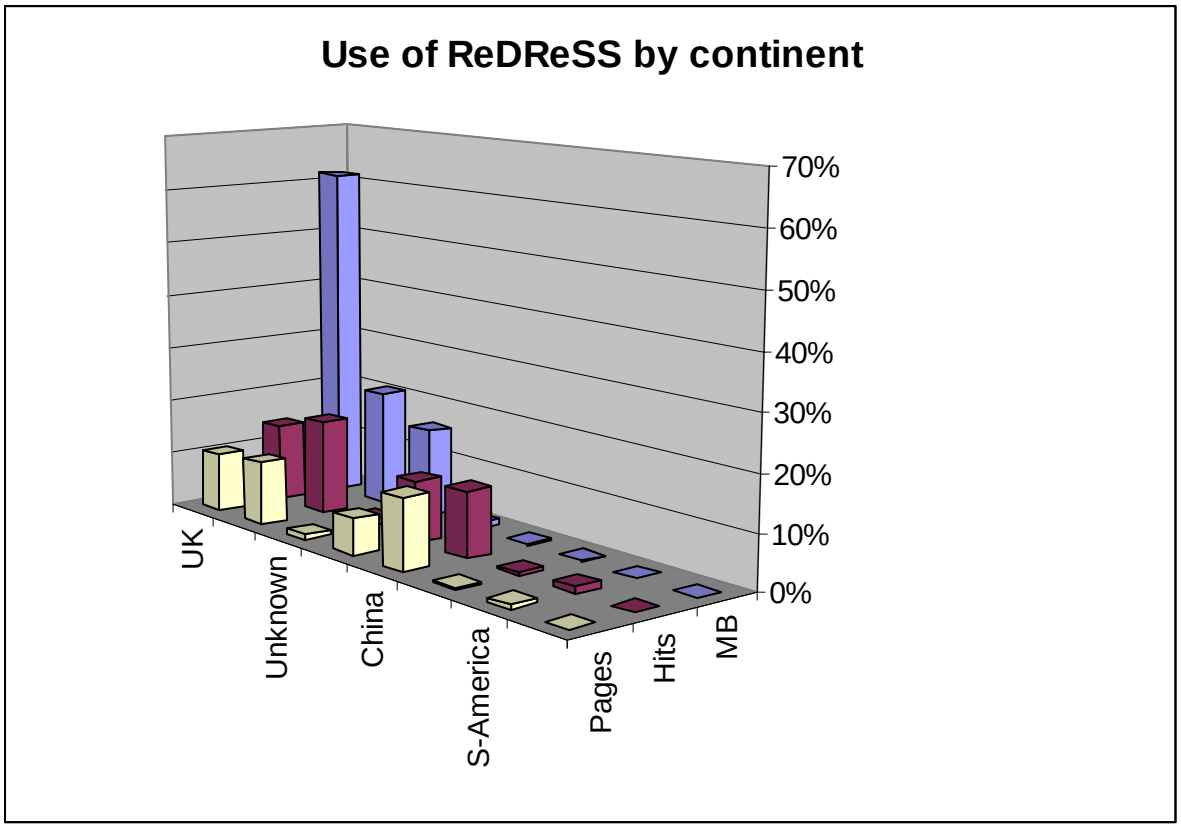

Figure 3. Percentage Use of ReDReSS by Continent.

Acknowledgements

This project is funded by the UK JISC and ESRC.

References

[1] http://lhc.web.cern.ch/lhc/ : LHC – The Large Hadron Collider
[2] http://www.jisc.ac.uk/ : The Joint Information Systems Committee
[3] http://www.esrc.ac.uk/ : The Economic and Social Research Council
[4] http://redress.lancs.ac.uk/ : Resource Discovery for Researchers in e-Social Science
[5] http://www.fedora.info/ : Fedora Digital Repository System
[6] http://dublincore.org/ : The Dublin Core Metadata Initiative
[7] http://www.impatica.com/ : Impatica Inc
[8] http://www.sakaiproject.org/ : Collaboration and Learning Environment for Education
[9] http://agora.lancs.ac.uk/ : The Online Meeting Tool