

# Scoping the Online Visibility of e-Research by Means of e-Research Tools

Robert Ackland<sup>1</sup>, Jenny Fry<sup>2</sup> Ralph Schroeder<sup>3</sup>

<sup>1</sup>Australian National University and Oxford Internet Institute, University of Oxford

<sup>2</sup> Department of Information Science, Loughborough University

<sup>3</sup> Oxford Internet Institute, University of Oxford

Email address of corresponding author: [j.fry@lboro.ac.uk](mailto:j.fry@lboro.ac.uk)

**Abstract.** e-Research has generated a number of expectations about increased collaboration between institutions and disciplines and also the organization of research communities around a central infrastructure. In this paper we use dynamic hyperlink network analysis to map the web presence of funded e-Research in the UK (both e-Science and e-Social Science) between 2002 and 2007. While it is too early to identify directions that would point to changes in science policy, there are indications that UK e-Research has undergone some major shifts while it is still uncertain whether the expectation of cohering around a core infrastructure will be met.

## Origins and Expectations of e-Research in the UK

e-Science in the UK can be considered to have started with the introduction of the e-Science Core Programme in 2001. The programme was initiated with joint funding from the EPSRC (Engineering and Physical Sciences Research Council) and the Department for Trade and Industry to the sum of £250 million. Each of the seven UK Research Councils at that time were responsible for selecting pilot projects under the e-Science Core Programme, which resulted in eleven initial pilot projects across a range of disciplines. There have been a variety of follow-ups to the Core Programme, with the EPSRC continuing to fund a dedicated e-Science Programme, the establishment of the National Centre for e-Social Science funded by the ESRC (Economic and Social Research Council) in 2004, and several initiatives in the humanities by the JISC (Joint Information Systems Committee) and the Arts and Humanities Research Council, such as the JISC funded Arts and Humanities e-Science Support Centre.

As UK e-Science has matured and transitioned from developing proof of concept for grid-enabled tools and infrastructure, a number of expectations have emerged amongst the organizations that govern academic and scientific research. These include increased collaboration between developers of e-Research tools and technologies towards interoperability, increased collaboration between different disciplinary communities, the development of a sustainable infrastructure through interoperability, standardisation and the appropriation of tools and infrastructure beyond initial funding programmes for e-Research, and finally the diffusion of tools, technologies and approaches developed in the UK e-Science Core Programme to the social sciences and the humanities. Put differently, it has been thought that research communities would adopt grid-enabled tools and come to organize their work around a central e-infrastructure.

The extent to which these expectations might be fulfilled or are currently being fulfilled is unclear. Partly this is due to the lack of a systematic approach to evaluate the outcomes and impact of e-Science within and across research communities, but it may also be that new forms of knowledge are difficult to gauge by the standards of conventional research outputs. In this paper we attempt to address this lacuna by presenting a network perspective of how e-Research in the UK has evolved over time – taking the hyperlink structure to indicate influence and impact on development, collaboration and diffusion.

## An e-Research Methodology for Mapping e-Research

Our approach for collecting and analysing e-Science and e-Social Science project web data extends methods for researching web networks outlined in Ackland and Gibson (2006) and Ackland and O’Neil (2007) using the Virtual Observatory for the Study of Online Networks (VOSON) e-Research software tool (see <http://voson.anu.edu.au>).

First, we used the National e-Science Centre (NeSC) and National Centre for e-Social Science (NCeSS) websites to identify projects that were recipients of UK e-Research funding. Our sample consists of 51 projects: 42 projects identified from the NeSC website, 7 projects from the NCeSS website (the NCeSS nodes), and NeSC and NCeSS themselves are also included. For each known project, we tried to identify a project website. However some projects are not represented by a single project site, and in such cases we searched for websites that (in our view) best represent the project (in some cases, this included academic homepages of the researchers most closely identified with the project). In what follows, we refer to the 51 projects as “seed sites” or “seed URLs”, but it is important to note that our approach allows for a given project to be represented by any number of websites.

We then used the Internet Archive’s (IA) Wayback Machine ([www.archive.org](http://www.archive.org)), which has been archiving the web since 1996, to collect historical data on hyperlinks and text content for our sample of e-Science and e-Social Science project web sites. For each project, we collected data at yearly intervals (we attempted to collect the data for the middle of each year, although the exact date of data collection was dependent on the date of the IA crawl). The archived pages had to be manually accessed using web browsers (the IA currently does not allow web crawler access) and in order to limit the data collection process we only accessed the archived top-level (or index) pages from the websites, and those (internal) pages linked to directly from the top-level pages. The data collection resulted in a time series of web networks (from 2002 to 2007) where the nodes represent websites (the projects in our sample and the websites that they hyperlink to) and the edges or arcs represent hyperlinks. Note that not all projects in our sample were present in the web network for each of the 6 years of our study: the first year of inclusion was the year the project was first funded (or 2002, if that was later), while the last year of inclusion was determined using information on duration of grants and also by checking for project activity on the websites.

## Findings

As shown in the first row of Table 1, the number of seed projects included in the web network rose from 11 in 2002 to 38 in 2006, and then decreased to 33 in 2007. This fluctuation in the number of projects in the network reflects project funding duration, which is a major source of network dynamics in a research field. We coded the seed projects into three groups reflecting their primary focus: domain-specific research, development of generic tools, and provision of services and infrastructure. We coded the research-oriented projects into the broad disciplinary categories of engineering and physical sciences, life sciences and

social sciences. We can see from table 1 that tool development projects account for the highest proportion of seed projects, in 2002 the number was just 6 this peaked in 2005 to 20 projects and by 2007 had dropped to 16. The number of life science seed projects also peaked in 2005 at 11. The services/infrastructure group contained one website in 2002-2003 (NeSC), with NCeSS being added to this group in 2004. We also classified the seed projects as either e-Science (which included research-focused projects from engineering and physical sciences and life sciences, the tool development projects, and NeSC) and e-Social Science (which included the social science research projects and NCeSS). In terms of number of projects, e-Science in the UK reached a peak in 2005 and subsequently tailed off, whilst the e-Social Science effort only took off in 2006.

	2002	2003	2004	2005	2006	2007
<b>Composition of sites</b>						
Seeds	11	22	30	37	38	33
Sites linked to by seeds	249	374	511	676	734	763
All	260	374	541	713	772	796
<b>Composition of seeds</b>						
Engin. / phys. sciences	2	4	4	3	3	3
Life sciences	3	6	7	11	7	5
Social sciences	0	0	0	1	8	7
Tool development	6	11	17	20	18	16
Services/infrastructure	1	1	2	2	2	2
All	11	22	30	37	38	33
e-Science	11	22	29	35	29	25
e-Social Science	0	0	1	2	9	8
All	11	22	30	37	38	33
Network inclusiveness	0.55	0.64	0.73	0.81	0.82	0.97
Network density	0.073	0.048	0.043	0.044	0.038	0.059

**Table 1. Composition of sites in web network 2002-2007**

From Table 1, it is apparent that the number of URLs linked to by seed projects rose steadily between 2002 and 2007 by roughly 100 URLs each year. Even though only 8 new seed projects appear in the web network in 2004, there was a marked jump in the number of sites linked to by the seeds (this is possibly due to NCeSS entering the network). Two network measures (calculated for the network comprising only seed sites) are also presented in Table 1. Network inclusiveness (the proportion of connected network nodes) increased monotonically throughout period studied, reflecting a greater degree of interconnectedness between seed projects mainly through the role of NeSC and NCeSS as sites that both link to, and are linked to by, the seed project sites. Network density (the number of links between nodes expressed as a proportion of the maximum number of possible links) decreased from 0.073 in 2002 to 0.047 in 2003 (reflecting the large increase in the number of projects) and then averaged around 0.04 until 2007 when it jumped to 0.059. The fact that network density was relatively stable during the period of growth in the number of projects between 2003-2006 is also indicative of growing interconnectedness. A visual representation of the growth of e-Research in the UK is presented in Figures 1 and 2, which show the web networks formed by the seed projects in 2002 and 2007, respectively. While space considerations prevent the presentation of the other maps, they revealed that three seed projects start to play a more central role in the web network in 2004: Emap, GridPP and Astrogrid in terms of

receiving outlinks from other seed projects and the OGSA-DAI project also generated a number of outlinks to other seed projects.

Table 2 shows the average number of outbound hyperlinks from the seed websites to other websites (seed and non-seed). The figures in brackets show the average number of inbound hyperlinks to the seed projects from other seed websites. We have broken down the average number of outbound and inbound hyperlinks in the dataset according to our main categories of disciplinary grouping, tool development and services or infrastructure development.

Disciplinary grouping	Seeds - average number of hyperlinks to (from) other sites					
	2002	2003	2004	2005	2006	2007
Engin. / phys. sciences	20.0 (0.0)	14.3 (0.0)	18.8 (0.0)	19.3 (0.3)	24.0 (1.0)	31.0 (1.3)
Life sciences	45.5 (0.0)	17.8 (0.7)	18.6 (1.0)	15.1 (1.2)	18.3 (0.9)	24.4 (1.0)
Social sciences	n.a.	n.a.	n.a.	0.0 (0.0)	7.3 (0.1)	17.4 (1.0)
Tool development	23.3 (0.7)	24.2 (0.9)	25.2 (1.1)	30.2 (1.5)	31.9 (1.3)	35.1 (1.6)
Services/infrastructure	25.0 (4.0)	40.0 (8.0)	32.5 (6.0)	48.0 (7.5)	63.0 (10.0)	62.0 (10.0)
All	26.9 (0.7)	21.4 (1.0)	23.3 (1.2)	25.0 (1.6)	25.2 (1.4)	31.0 (1.9)

**Table 2. Average number of hyperlinks to seed projects**

The projects that we categorized as being primarily concerned with the provision of services and infrastructure (NeSC and NCeSS) received on average a higher number of inbound hyperlinks than the research oriented and tool development seed projects. The services or infrastructure projects also had on average a higher number of outbound hyperlinks. Of the research oriented projects those seed projects that we categorized as life sciences had on average the highest proportion of inbound hyperlinks. Table 3 shows the five most highly linked to websites according to year. The top five linked-to sites in the UK are fairly stable and it is notable that they do not represent individual projects (except for NeSC which, for the purposes of this exercise, we term a project), but are sites of the organizations that govern e-Research in the UK and sites that represent the umbrella and standards organizations in e-Science.

Top-5 sites (by number of inbound hyperlinks)		
2002	2003	2004
<a href="http://www.nesc.ac.uk/">http://www.nesc.ac.uk/</a>	<a href="http://www.nesc.ac.uk/">http://www.nesc.ac.uk/</a>	<a href="http://www.nesc.ac.uk/">http://www.nesc.ac.uk/</a>
<a href="http://www.epcc.ed.ac.uk/">http://www.epcc.ed.ac.uk/</a>	<a href="http://www.ogf.org/">http://www.ogf.org/</a>	<a href="http://www.ogf.org/">http://www.ogf.org/</a>
<a href="http://www.scitech.ac.uk/">http://www.scitech.ac.uk/</a>	<a href="http://www.epcc.ed.ac.uk/">http://www.epcc.ed.ac.uk/</a>	<a href="http://www.globus.org/">http://www.globus.org/</a>
<a href="http://www.ogf.org/">http://www.ogf.org/</a>	<a href="http://www.dcs.gla.ac.uk/">http://www.dcs.gla.ac.uk/</a>	<a href="http://www.inf.ed.ac.uk/">http://www.inf.ed.ac.uk/</a>
<a href="http://www.grid-support.ac.uk/">http://www.grid-support.ac.uk/</a>	<a href="http://www.grid-support.ac.uk/">http://www.grid-support.ac.uk/</a>	<a href="http://www.grid-support.ac.uk/">http://www.grid-support.ac.uk/</a>
2005	2006	2007
<a href="http://www.nesc.ac.uk/">http://www.nesc.ac.uk/</a>	<a href="http://www.nesc.ac.uk/">http://www.nesc.ac.uk/</a>	<a href="http://www.nesc.ac.uk/">http://www.nesc.ac.uk/</a>
<a href="http://www.ogf.org/">http://www.ogf.org/</a>	<a href="http://www.ogf.org/">http://www.ogf.org/</a>	<a href="http://www.ogf.org/">http://www.ogf.org/</a>
<a href="http://www.grid-support.ac.uk/">http://www.grid-support.ac.uk/</a>	<a href="http://www.epcc.ed.ac.uk/">http://www.epcc.ed.ac.uk/</a>	<a href="http://www.allhands.org.uk/">http://www.allhands.org.uk/</a>
<a href="http://www.globus.org/">http://www.globus.org/</a>	<a href="http://www.esrc.ac.uk/">http://www.esrc.ac.uk/</a>	<a href="http://www.esrc.ac.uk/">http://www.esrc.ac.uk/</a>
<a href="http://www.inf.ed.ac.uk/">http://www.inf.ed.ac.uk/</a>	<a href="http://www.rcuk.ac.uk/escience/">http://www.rcuk.ac.uk/escience/</a>	<a href="http://www.webstandards.org/">http://www.webstandards.org/</a>

**Table 3. Most highly linked to projects amongst seed websites**

## Discussion

This paper presents a preliminary analysis of the data. We found that a number of projects funded as part of UK e-Science, including some of the initial pilot projects and the later best practice projects, and e-Social Science initiatives were missing from our web networks. This was surprising and could be the result of a number of factors, such as a limitation in the selection criteria for identifying seed projects, a function of changing policies regarding website management, or it may indicate that the objective of the National Centres for e-Science and e-Social Science to act as central portals for e-Research in the UK has not been achieved.

None-the-less, a number of emergent patterns can be seen from the findings described above. In general, the UK e-Science effort has been transformed considerably over the course of time. This is evident from the fact that only 10 sites from the original seed set are still present in 2007. In 2002, UK e-Science was represented by the NeSC hub and a few projects in early adopting disciplines such as physics and astronomy. By 2007, this disciplinary core had not changed, despite the addition of many projects and a sizable proportion of life science projects. A high proportion of projects within the engineering and physical sciences are concerned with the development of generic tools and services, whereas a greater number of the life science projects are concerned with developing domain-specific tools within the context of research problems.

The web network of e-Research projects has grown steadily since the Core e-Science Programme started in 2001. In-link metrics for academic websites are typically low (Thelwall, 2006), therefore although the average in-link metrics shown in Table 2 seem low they are indicative of an evolving social structure in the dataset (and remember that they are inlinks from other seeds, not the entire web). This study does not include the type of data that would give insights into linking motivations, therefore, we cannot extrapolate as to why the inlink metric is so high for services and infrastructure, but we can hypothesize that these developments are being influential in some way within UK e-Research.

In 2007 we can see that the e-Science and e-Social Science programmes form two separate clusters meaning that the diffusion of generic tools and infrastructure developed under the e-Science programme have not yet diffused to the social sciences. It is unclear as to how this may occur, to what extent it may occur and what factors may be barriers to this expectation of diffusion.

## Conclusion

This paper is a first attempt at a dynamic hyperlink analysis of a research field. As science communication increasingly comes to rely on the Web, it will be important to gauge how well this type of analysis can map emerging trends in research. It remains to be seen whether a closer inter-penetration between e-Science and e-Social Science (and arts and humanities) will emerge in the future when e-Research in social sciences and humanities has moved beyond its early phase. For the period covered in this paper, however, there are several distinctive patterns, namely, the steady growth of the network from 2002-2006 both in terms of the number of projects and the links between them, and a decline in the number of seed projects in 2007 (perhaps reflecting of the short-term nature of UK e-Research funding initiatives), and the continued separate evolution of e-Science and (arriving later) e-Social Science. Interestingly, the central role of NeSC and NCeSS in the network, suggests the central role that these organizations have played - rather than any other form of 'infrastructure'

that might have done so - in the evolution of e-Research. How the structure of the web network will change when we include other seed sites outside of the funded projects analyzed here (the central role of the two umbrella organizations is in part due to sample selection), how the structure of the network will respond to any major new initiatives, and how the e-Research Web presence has evolved in comparison with other e-Research efforts such as 'cyberinfrastructure' and rapidly growing initiatives in the rest of world - these will in our subsequent analyses provide interesting clues about future e-Research directions.

## Acknowledgments

The research of Ackland on this paper was supported by a National Centre for e-Social Science (NCeSS) Visiting Fellowship, a James Martin Visiting Fellowship and Australian Research Council grant DP0452051. The research of Fry and Schroeder on this paper was supported by ESRC grant RES-149-25-1022 for the Oxford e-Social Science (OeSS) project: Ethical, Legal and Institutional Dynamics of Grid-enabled e-Sciences. We thank Jenny Ackland for assistance in the data collection.

## References

- Ackland, R. and R. Gibson (2006), "Hyperlinks and horizontal political communication on the WWW: The untold story of parties online," under review.
- Ackland, R. and M. O'Neil (2007), "Nanotechnology and Online Environmental Activism," under review.
- Thelwall, M. (2006). "Interpreting social science link analysis research: A theoretical framework". *Journal of the American Society for Information Science and Technology*, 57(1), 60-68.

# Appendix I –Selected Web Networks of UK e-Research Projects

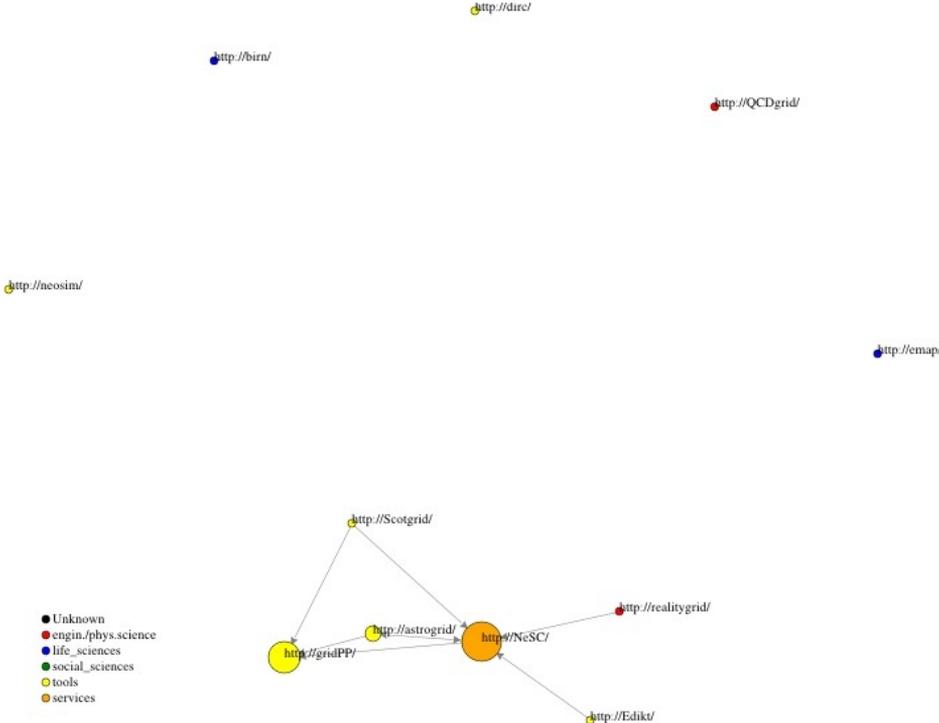


Figure 1. UK e-Research Web Network 2002: Seed Projects plus URLs

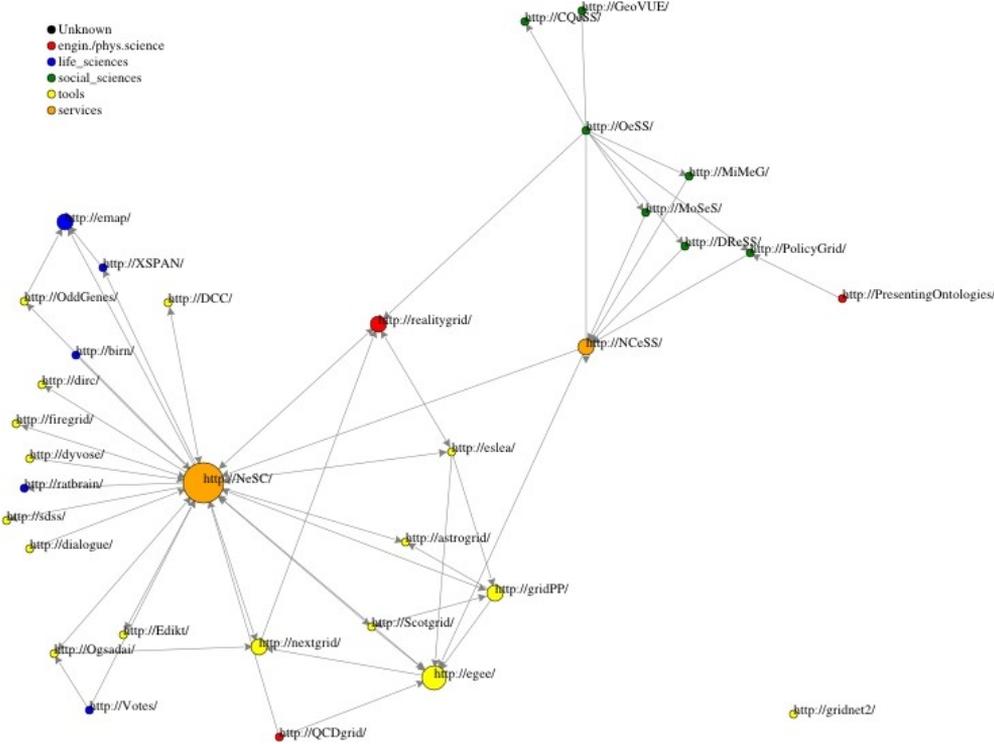


Figure 2. UK e-Research Web Network 2007: Seed Projects plus URLs