Overcoming Marginalisation? Open Access Research Repositories at a South African and a Swedish University.

IGU Commission on Marginalization, Globalization and Regional and Local Responses
International Conference at Universiti Teknologi Mara, Shah Alam, Malaysia 6-8 July 2009.

R.C. Fox, Rhodes University, South Africa (r.fox@ru.ac.za)
E. Wihlborg and **D. Lawrence**, Linköping University, Sweden (elin.wihlborg@liu.se, david.lawrence@liu.se)

Open Access Research Repositories have developed very rapidly since c2000 as a global phenomenon in their number, their location, the number and type of resources available in them. The creation of institutional repositories has been affected by different motives. They can serve as collections of a University's research output with the intention of making it 'freely' available. We ask here, just what patterns of access can be analysed, what trends do we see when examining our two institutions? Alternatively they can be seen as ways to raise the research profile of individuals and institutions and citation records. We do not see Open Access Research as being a neutral, value free technological innovation with clear outcomes. Our perspective draws from the fields of socio-technical and complex adaptive systems and so we anticipate that although the future impacts of Open Access research can be discerned they cannot be mechanistically predicted.

Complex adaptive systems (CAS) is a multi-disciplinary field which embraces the ideas of systems theory and evolution. Everything we experience in our human environment, including eResearch, can be understood as consisting of agents (authors, institutions, publishers etc) connected by processes (uploading and downloading using the internet) to each other and their immediate environment (Dooley 1996). Relatively few rules are usually governing the behaviour of the agents and these give rise, through adaptation and evolution, to emerging regularities. Contradictions and tension are common within CAS and act to produce the adaptations and evolution. The regularities themselves can be discerned by the agents as patterns and so feedback can go back into the system thus producing further cycles of change and adaptation.

This paper focusses on eResearch Repositories as part of a CAS and is, in fact, examining the regularities discerned in the two case studies to discuss patterns which will then produce feedback, following the presentation of this paper, into the publishing and academic systems. We do not view eResearch as a technologically neutral process that simply adds a new rule or process in some value free or independent way. eResearch is also a part of its environment and can be readily considered as a socio-technical system, with technology and information engaged in an interplay with their social, cultural, economic and political environment. As socio-technical systems, eResearch Repositories, consist of the technical components, individual actors and organizations, legal frameworks, and institutional and political structures. The system is interwoven with the societal context, cultural values, and specialized professional know-how by which it operates. The term *sociotechnical* indicates that we are including the social and human as well as technical components in the system and considering them to be interdependent. Socio-technical systems are, from this perspective, regarded as seamless webs of tightly interconnected components. Changes made to one part of the system must be adjusted to the other existing parts to obtain a working whole. According to this view, governing the innovations and policies that affect technology requires both social and technical knowledge of the system (Hughes 1983).

The paper later examines how agents (researchers) can reposition and understand more about the impact of their work through using eResearch Repositories. The repositioning is both geographical and, through inference, economic. We also comment on cultural (language) repositioning possibilities. Typically CAS can be examined at three different levels or scales: the global, at the level of the network itself, and at the local scale. The following sections follow this categorization and present the patterns and regularities which our analyses have discerned.

Global Scale: Examination of two papers

Let us start this paper with an example illustrating the potential for eResearch Repositories to throw up patterns of usage which we, as analysts, have not previously been examining. The software used at Rhodes University Repository allows for detailed examination of the following:

- The number of downloads of the entire document (commonly a pdf file) and abstract views for deposited material,
- The IP addresses which the queries came from, this allows us to see geographical access patterns,
- The patterns of downloads through time.

We have selected two papers placed in the repository at more or less the same time, October 2006 and January 2007, they come from the same discipline, Geography, and they were originally published within one year of each other, 1999 and 2000. The key indicator separating them is that the first paper was published internationally by a commercial publishing house and the second paper was published locally by a professional society.

The first paper was published in Applied Geography in 1999. An international accredited (on the ISI list) Journal published by Elsevier. The research was local, examining farming prospects in the Eastern Cape, South Africa. In October 2006 it was uploaded into the Rhodes University eResearch Repository as part of the University's programme of contacting Journal publishers and negotiating permission to upload material. By February 2009 the article had been downloaded 1221 times. The pattern of downloads by country is revealing.

Table 1 shows that South Africa, with 801 downloads, had almost six times the amount of downloads as the USA, which was the second largest country of origin of download requests. Significantly Kenya, Uganda and Rhodes University were the origins of other downloads from the African continent The UK and Ireland are the other two northern countries with many downloads. The downloading started with an initial surge of interest and then dipped in the next month before rising to a peak six months later.

The second paper was published in the South African Geographical Journal (SAGJ) in 2000, a locally accredited Journal owned by a professional body, the South African Geographical Society. The research focus this time was both methodological and local, concerning the development of a methodology to classify the zonation of South African rivers.

It was uploaded into the eRepository three months after the Applied Geography paper, in January 2007, and by February 2009 it had been downloaded 556 times. Table 1 shows that the geographical pattern of downloads for the SAGJ paper was completely different. The USA dominates with 369 downloads, South Africa, Iran and Rhodes University were the country origins of the next three. The timing of downloading was also quite different. The vast majority of the downloading happened in the first month and then dwindled very

rapidly afterwards rising to a secondary peak 18 months later. Perhaps we can infer that the secondary peak reflects interest from citations in work which followed the first peak.

The download patterns allow us to infer that the international (American) community had been previously excluded from material in the local Journal of the South African Geographical Society whereas the local (South African) research community had been marginalised from material in the internationally published Applied Geography because of cost and access issues (Horton 2000). So we can start to see how eResearch repositories are modifying the ways in which research is accessed.

Table 1: RUeRR: Document Downloads by Country of an International and a Local Journal Article

Applied Geography		South African Geographical Journal	
South Africa	801	USA	369
USA	140	South Africa	76
UK	35	Iran	17
Unknown	34	Rhodes University	12
Kenya	21	Australia	9
Rhodes University	17	Unknown	8
Uganda	16	China	8
Satellite Provider	15	Ireland	7
Ireland	12	UK	6
40 other countries	130	21 other countries	44
Total	1221	Total	556

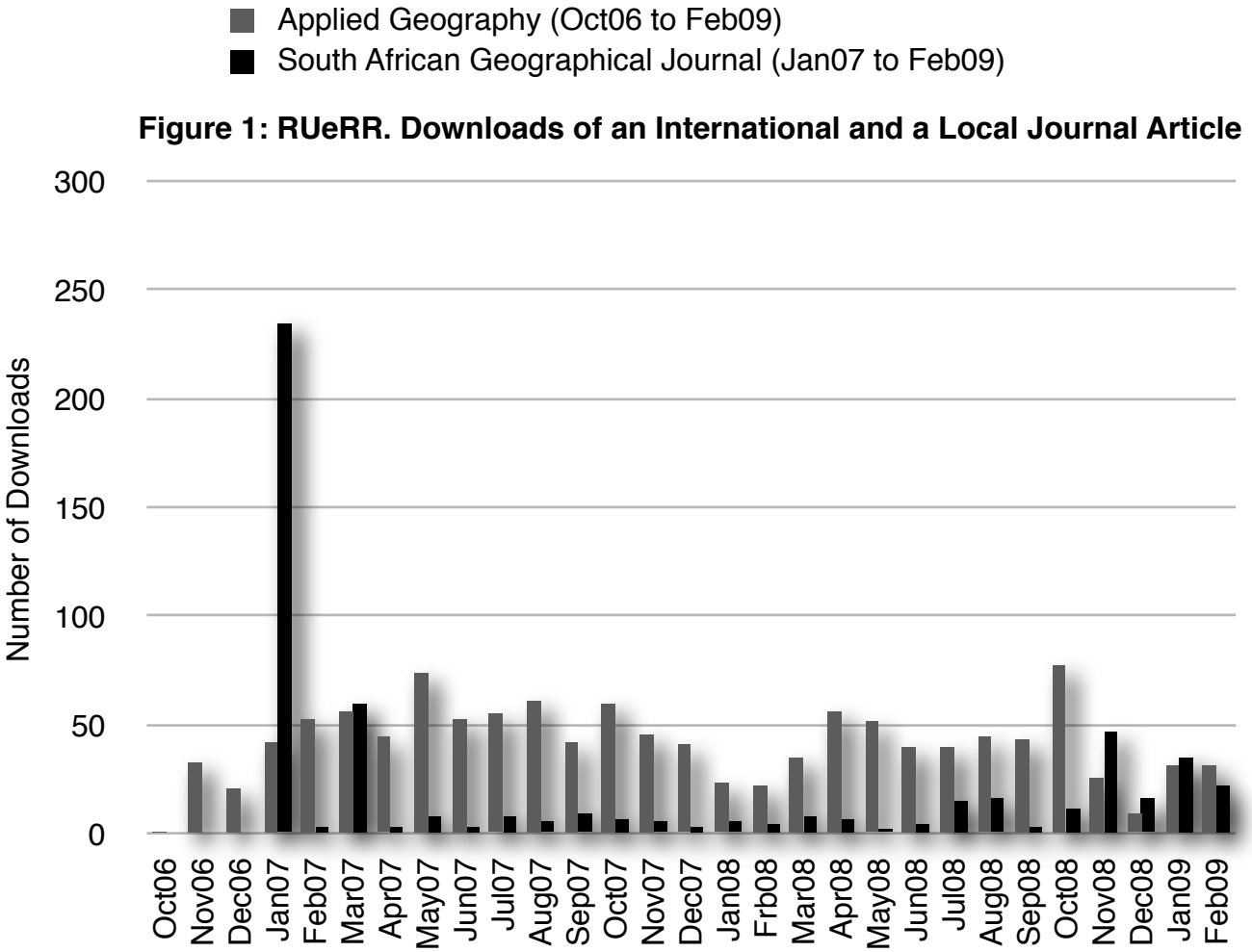


Figure 1: RUeRR. Downloads of an International and a Local Journal Article

Globally

Although open access repositories are a recent phenomenon they have spread quickly and had a very rapid impact. Figure 2 (Registry of Open Access Repositories 2009) shows that there was a very slow adoption through the 1990s, by 1999 there were only 35 repositories globally (Lewis 2009). The USA and Germany had five each, Brazil had three then the UK, France, Canada, Italy, Finland and Turkey had two. What is significant is the group of more peripheral countries that were early adopters: Mexico, Cuba, Peru, Chile, Brazil, Turkey.

Ten years later, by March 2009 the global map is completely different (Figure 3). There are 1225 repositories, many with over 500,000 records in them, spread across 73 countries. The big increase in the number of repositories came around 2005, both Linköping University's ePress and Rhodes University's eResearch Repository are therefore typical of this trend. The focus is now clearly Anglo-American: the USA tops the list with 25% and the UK with 11% of all repositories. Following them are a group of rich northern nations: Germany, Japan, Australia, Netherlands, Italy and Canada. These top eight countries account for two-thirds of the global total. Noticeable absent are repositories in Africa, the Persian Gulf region and the central Asian republics grouped around the Caspian and Aral seas. Later we will see that although these areas have no repositories of their own they still generate significant numbers of queries to open access repositories such as at RU.

Thus strengthening the argument that open access can overcome the barrier of cost that characterizes the global academic publishing industry (Horton 2000).

Figure 2: Repositories Registered by Time and Geographical Distribution 1999.

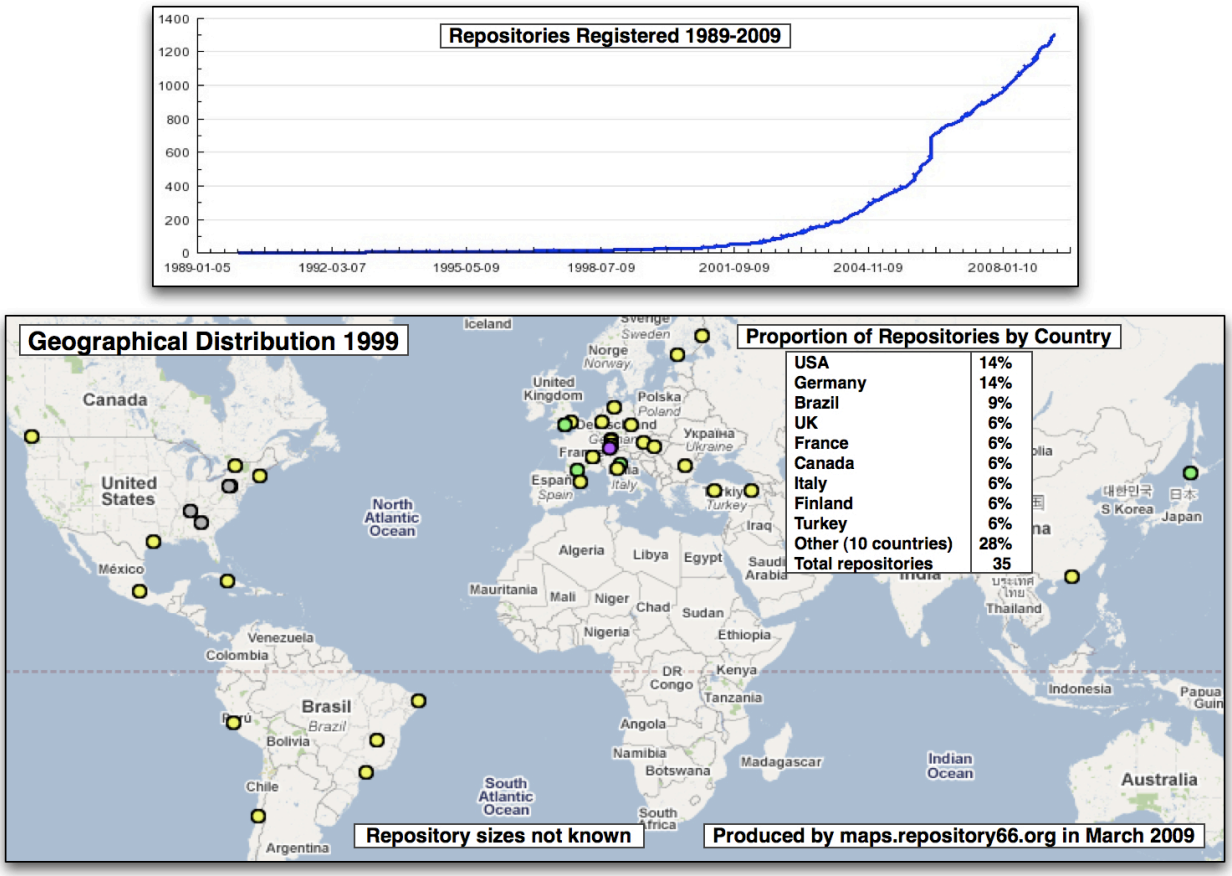
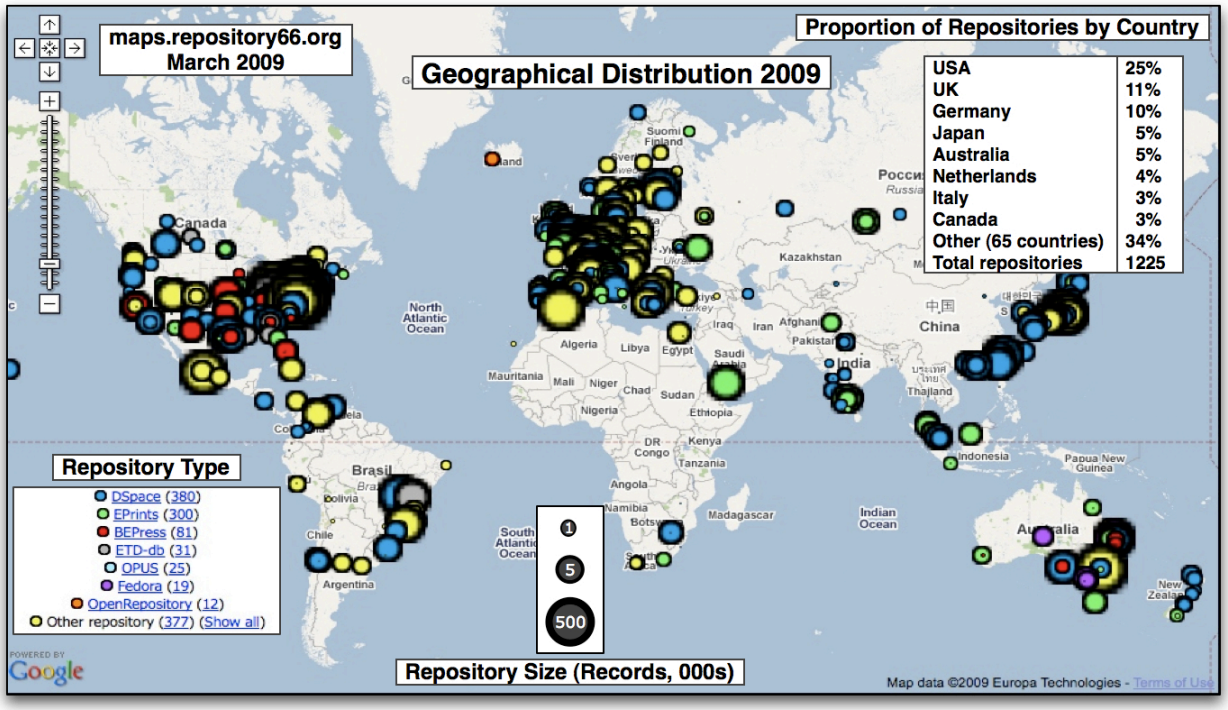
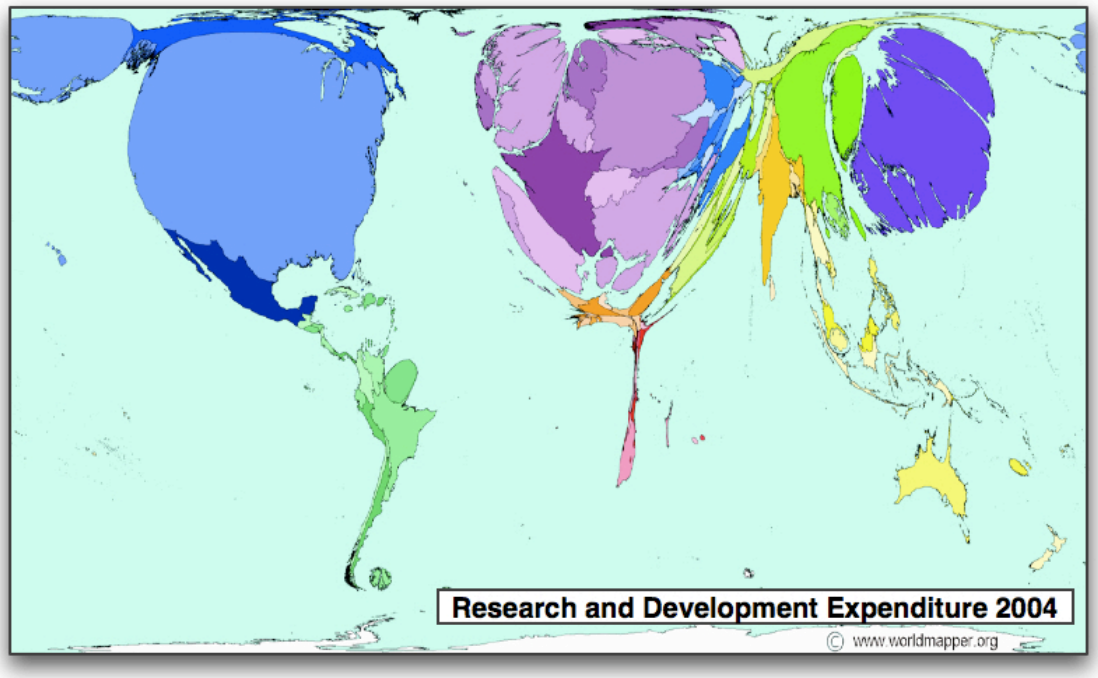


Figure 3: Geographical Distribution eResearch Repositories 2009.



Given the predominance of the developed world in internet usage, research and development expenditure etc it is no surprise that countries such as the USA, UK, France, Australia and Canada have the most eResearch Repositories. Compare Figure 3 with Figure 4 which shows country size proportional to their research and development expenditure (Worldmapper 2009).

Figure 4: Research and Development Expenditure 2004.



Sweden had 31 repositories in 2008 whereas South Africa only had 15 (University of Nottingham 2008). The Ranking of Web Repositories (Consejo Superior de Investigaciones Científicas (CSIC) 2009) shows that the Swedish repositories were more 'visible'. Göteborg University was ranked 28th in the world, Uppsala University was 69th, Mälmö was 114th but LiU was not ranked in the world's top 300. The only three South African repositories in the world's top 300 were the University of Pretoria at 238th, Council for Scientific and Industrial Research at 241st and Rhodes University at 263rd.

The OpenDOAR database (University of Nottingham 2008) shows that English is the most frequent language encountered in the repositories (Table 2). This is not surprising, nor is the fact that German is the second and Spanish the third most frequent language. Ten percent of all repositories are in Germany and there are many more throughout the Spanish speaking world in central and southern America. Swedish is the dominant Scandinavian language found in 34 repositories.

We examined the IP addresses of requests for a Swedish language report held at LiU's ePress and found that it was only accessed by addresses registered in Sweden. A fraction of these came from various Swedish Universities, but the majority was from private network suppliers. This could indicate that there is a significant usage from the general population, or it could just be that researchers often work at home. Here we can see that Open Access facilitates dissemination of knowledge to an audience which is not English speaking and to whom the report holds interest and presumably significance.

Table 2: The Most Frequent Languages Used in Repositories belonging to the OpenDOAR Database.

Language	Repositories
English	1164
German	155
Spanish	100
French	88
Japanese	75
Dutch	46
Italian	43
Portuguese	37
Swedish	34
Others (43 languages)	184

If we scrutinize who uses the RUeRR we find that the queries for research material do not follow the pattern that could be expected given the information shown in Figures 3 and 4. India lies third, China is eighth, Indonesia is 12th, Iran 13th, Malaysia 14th, Thailand 19th, Brazil 20th. Table 3 shows that Rhodes University is the 10th highest origin of requests for downloads. Over 30% of all requests come from South Africa (including RU) and India combined which gives a clear indication of how an eResearch Repository can facilitate access in the global south just as it does to minority language speakers in the global north. Evans and Reimer (2009) have recently shown that the greatest impact of Open Access is on researchers in the developing world. They found that this trend does not apply to the poorest countries, where electronic infrastructural is still limited, but we will show later that this does not necessarily follow.

Table 3: RUeRR: Document Downloads by IP address of query

Downloads by top 25 countries (IP address)	
USA	256,792
South Africa	236,518
India	50,997
UK	45,089
Unknown	36,675
France	26,791

Downloads by top 25 countries (IP address)	
Australia	23,869
China	17,540
Canada	17,362
Rhodes University	15,008
Germany	13,236
Indonesia	11,393
Iran	10,252
Malaysia	8,605
Netherlands	8,183
Ireland	7,508
Italy	6,891
Switzerland	6,679
Thailand	6,531
Brazil	6,095
Sub-total	812,014
186 other countries	169,106
Total	**981,120**

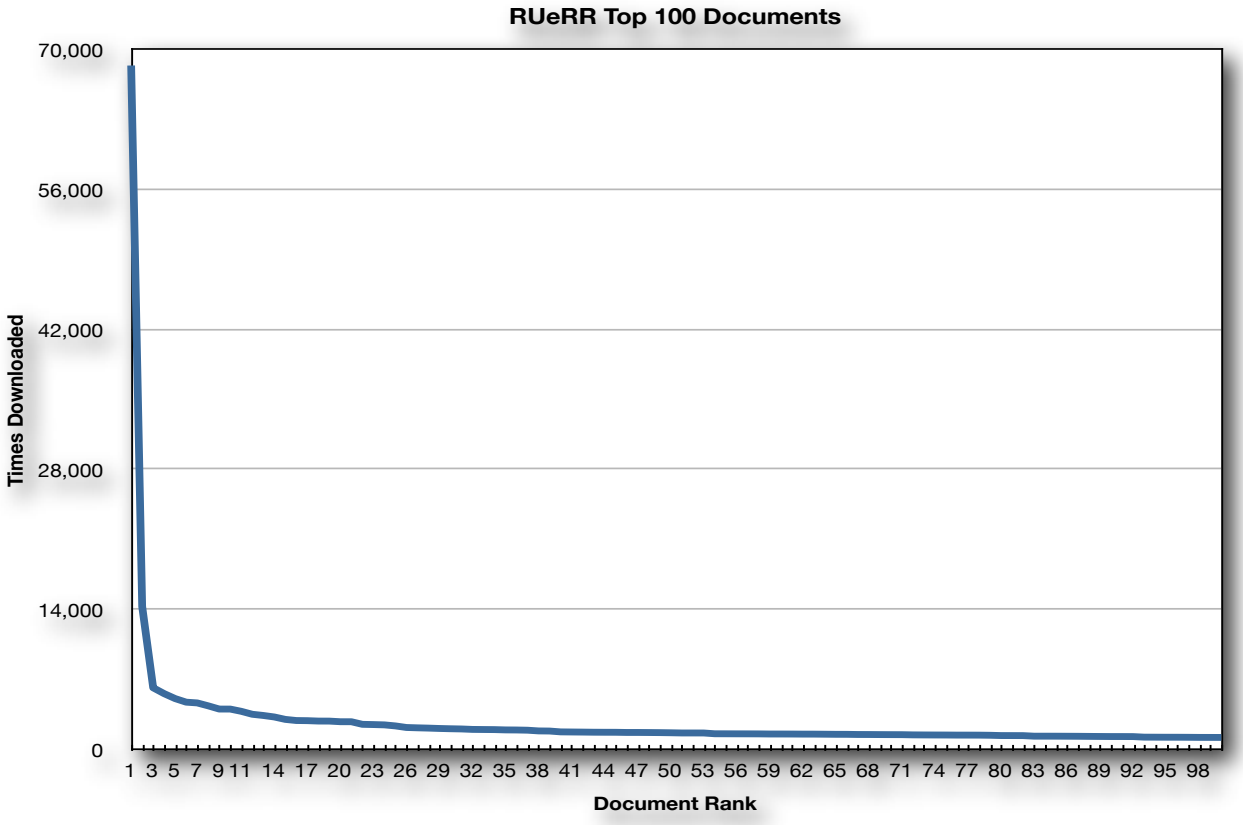
Network Scale

Citation indices (Antelman 2004) and internet connectivity are two socio-technical fields which have already received a good deal of scrutiny. Newman's major review (2003) of network systems leads us to expect certain types of regularities in our eResearch Repositories, for example in how many times documents are accessed and the sources of downloads. The regularity we can expect is a power-law distribution, also known as a scale-free network. This states very simply that there will be a large number of, for example, documents in an eResearch Repository which few people have downloaded, a few papers that many people have downloaded and one paper with a very large number of downloads. The relationship between the papers can be thought of as exponential. Alternatively, there will be a large number of countries from which there have been very few requests for downloads and only a few from which most of the requests have originated. Table 3, above, has already indicated something of this type of relationship though it does not follow a power law relationship, Perhaps we can speculate that this is

because the RUeRR is still a very recent phenomenon and there have not yet been enough requests to generate the pattern predicted?

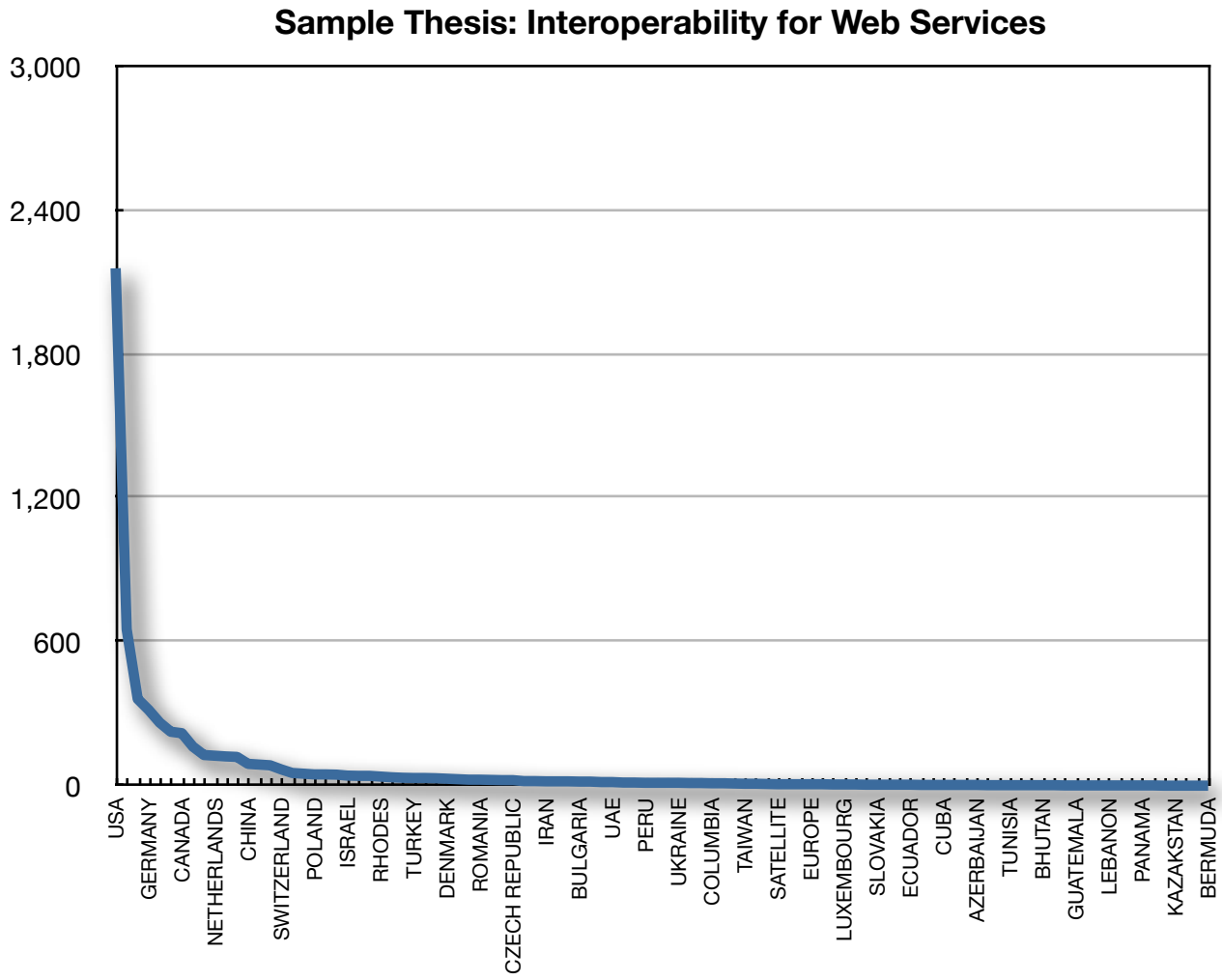
If we examine the top 100 documents lodged in RUeRR you find the power-law distribution. Figure 5 shows the rank of the documents on the horizontal scale and the number of times they were downloaded on the vertical scale. At Rhodes University one document was accessed almost 70,000 times whilst most documents formed a long 'tail' to the right of the graphic, a tail that continues to the right for a a considerable distance. The total number of downloads for all of the documents was approximately one million at the time when these data were extracted in February 2009. Later we will examine what the authors of the most downloaded documents felt about their prominence.

Figure 5: RUeRR Power-Law Distribution of Downloads



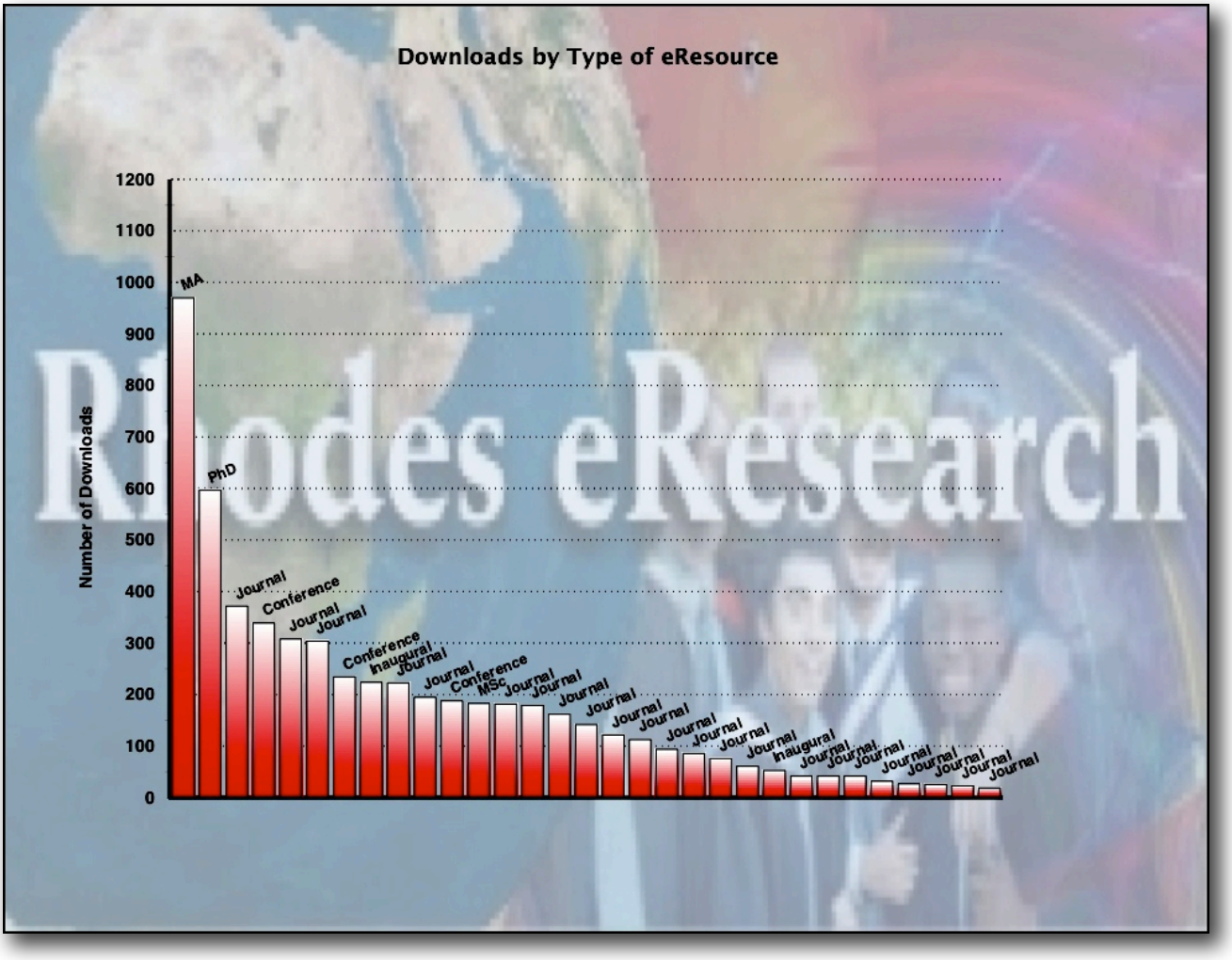
If we take one of the top ten downloaded documents at random then a similar trend line is revealed for the country of origin of download requests. We selected the eighth most popular download: a 2006 Masters thesis in Computer Science. Figure 6 shows the 100 most popular countries on the horizontal axis and the number of downloads on the vertical axis. Once again there is clearly a power-law distribution. Over 2000 download requests came from the USA, 650 from India, 360 from the UK, Germany with 320, South Africa with 260 and so on down to a dozen countries with only 1 request. Geographically it is pertinent that such a high proportion of requests originated from India and South Africa since these are two countries which do not belong to the big spending Research and Development countries (Figure 4).

Figure 6: Power-Law Showing Countries of Origin for a Sampled Thesis



We do not always experience the power-law distribution. In June 2007, some two years after the RUeRR was fully operational, there were 31 documents categorized as 'Geography'. Most of these were post-print Journal Articles with a few Masters and PhD theses, inaugural lectures etc. Figure 8 shows us that these documents have a distribution which may be moving towards a power-law but which is not exhibiting it yet. The two most popular downloads were the recent Masters and Doctoral theses that had been lodged in RUeRR. This was before it became University policy to do so. Perhaps we can speculate that the Masters thesis with the most downloads will increase in downloads once it becomes widely cited by researchers accessing it?

Figure 7: Geography Downloads from RUeRR, June 2007



Local scale

If we move down to the agents themselves, we need to examine how the documents have got on to the RUeRR itself, what is the policy instrument? Who are the authors of the Geography papers talking to? Has their been any evidence that eResearch repositories can overcome the digital divide? Evans and Reimer (2009) recently concluded that Open Access - unrestricted online access - has its greatest impact on access from the developing world, but not the very poorest countries since they have poor electronic infrastructure. So we would expect the RUeRR to be revealing access patterns from other developing African, Asian and Latin American countries.

The pattern of countries of origin reveals strong but not unexpected relationships. Once again we will examine a selection of the Geography documents from June 2007. If we return to the two papers examined at the start of the paper we immediately see some support for this contention. The paper published in Applied Geography was overwhelmingly accessed from South Africa (Table 1) - perhaps not surprisingly given the focus of the paper on rural development in South Africa. Kenya, Uganda and Rhodes University were three of the other developing world locations from which download requests originated. For the SAGJ paper, however, requests for downloads were dominated by the USA, for reasons elaborated on earlier.

The two most accessed Geography documents by June 2007 were a Masters and a PhD thesis, both with strong local focus. Their access patterns are strikingly similar as Figures 8 and 9 reveal but different to the Applied Geography paper (presumably the Americans had already accessed this paper?). Both theses had South Africa as the dominant country with just under 50% of the download requests followed by the USA with around 33%. Rhodes University, China, India and the Russian Federation also appear with the cluster of developed countries such as the UK, Ireland and Australia.

Figure 8: Geographical Origin of Download Requests, Geography Masters Thesis.

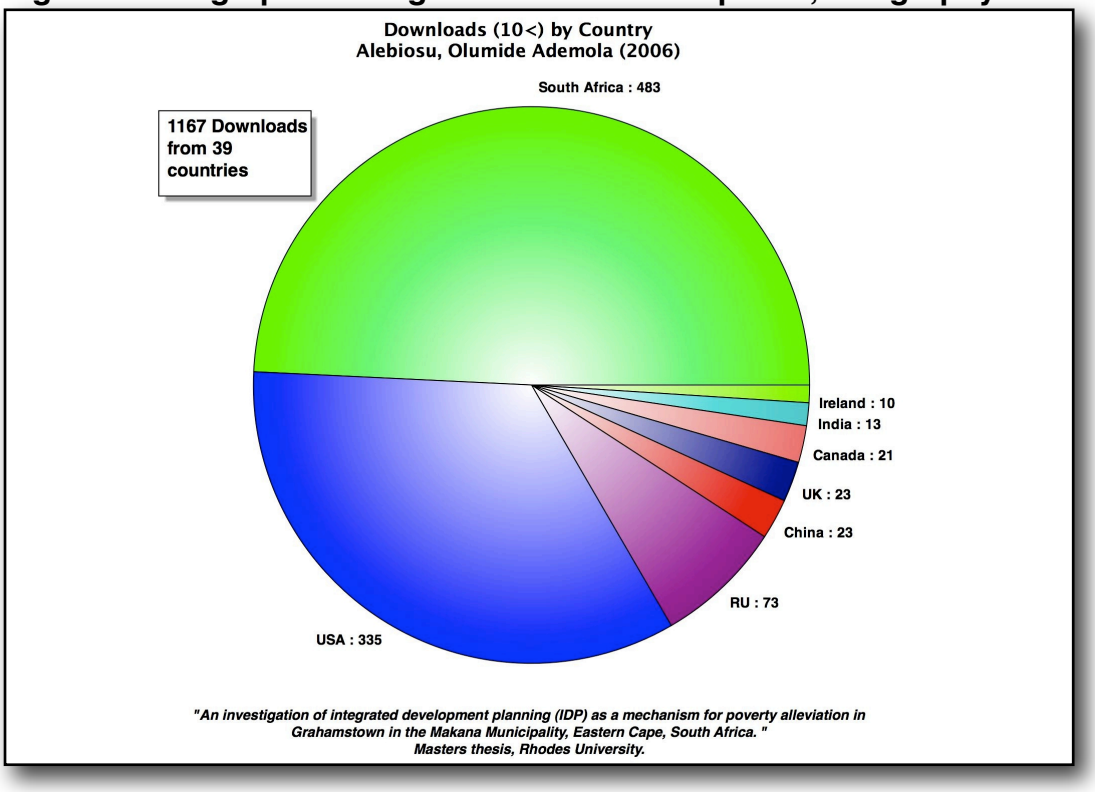
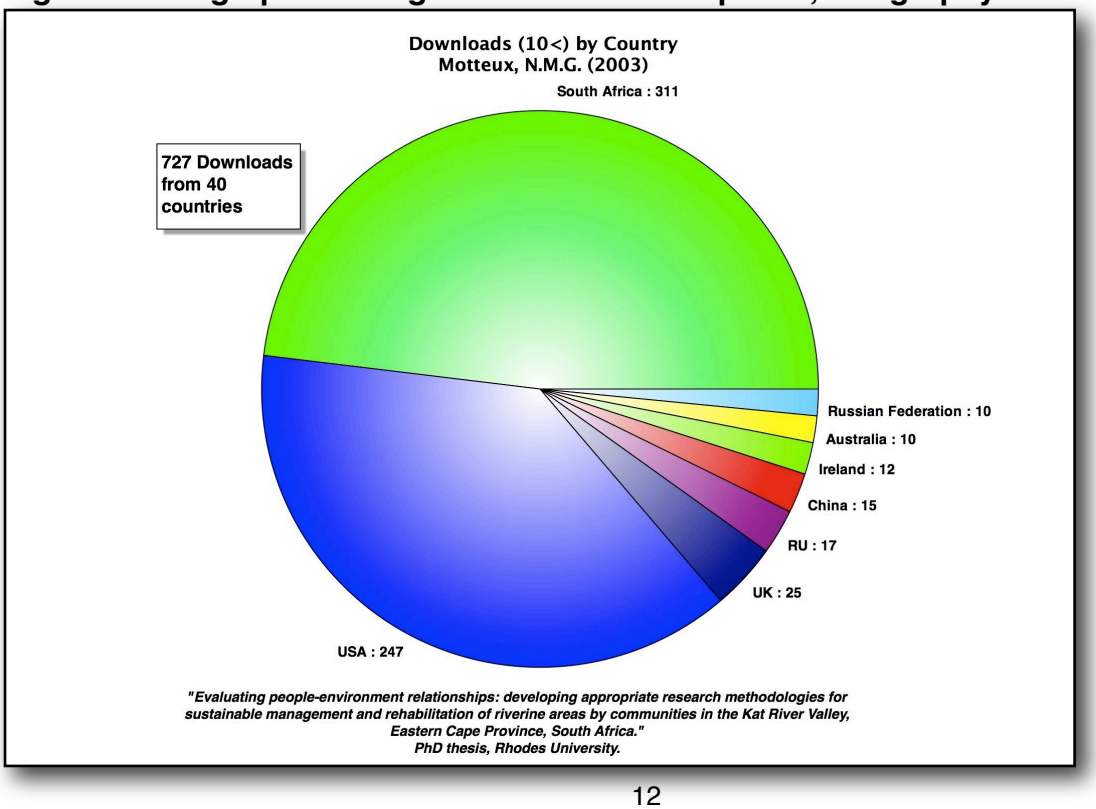
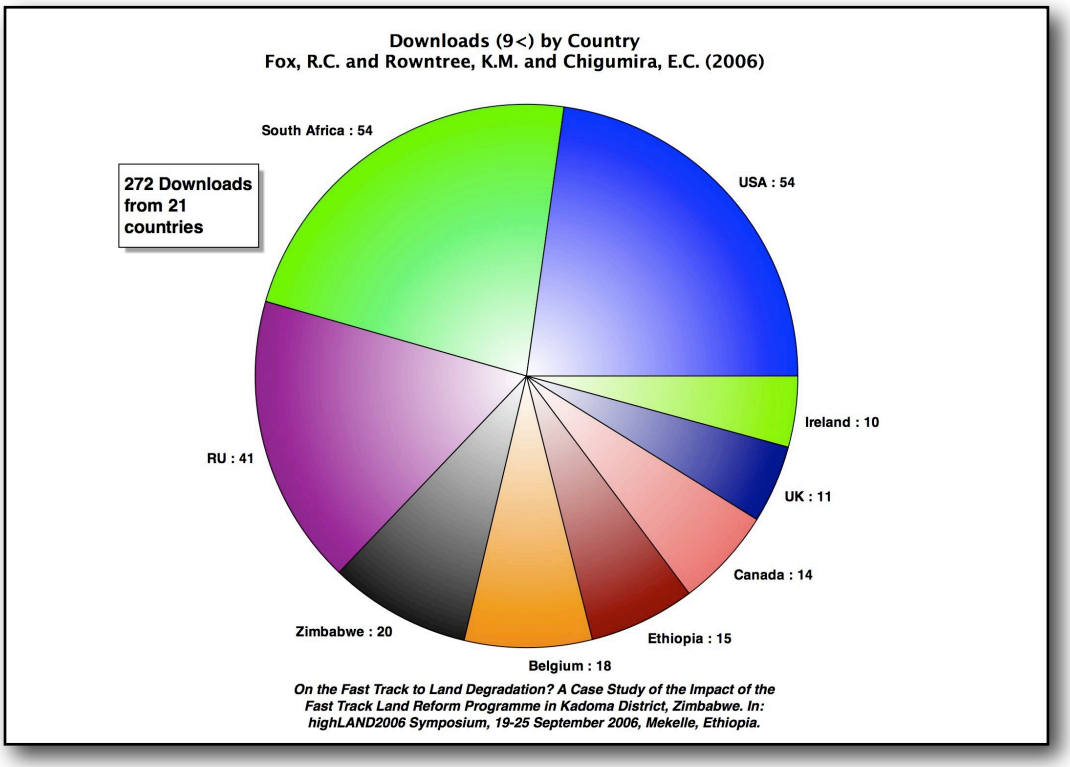


Figure 9: Geographical Origin of Download Requests, Geography PhD Thesis.



Lastly, we examine a third type of document, in this case the pre-print version of a paper on Land Reform in Zimbabwe which was presented in September 2006 at a conference in Ethiopia, one of the world's poorest countries. Following the conference it was uploaded into the RUeRR in October and then submitted to Geography, the Journal of the British Geographical Association, in December 2006. By June 2007 the paper was still in the review process but had been downloaded 272 times with requests coming from 21 different countries (Figure 10). Once more the USA and South Africa headed the list but then came Rhodes, Zimbabwe, Belgium (the conference was co-organised by a Belgian University) and Ethiopia. So there are particular reasons for the access from these specific developing countries. In September 2007 the paper was published in Geography and by then it had already been downloaded 356 times, subsequently the number of downloads did not begin to diminish until the end of 2008 and the beginning of 2009. It is perhaps pertinent that the largest proportional increase in requests for downloads came from the UK once the Geography article had been published.

Figure 10: Geographical Origin of Download Requests, Pre-print of a Paper on Zimbabwe presented at a conference in Ethiopia.



We questioned the authors of the three most downloaded papers in RUeRR to establish what they could tell us about the popularity of their documents and the distribution of the downloads. The most popular download is a 2003 Masters thesis in Computer Science which was deposited in July 2005. The author speculated that one reason for its popularity is its format: the only one in RUeRR in html which means it is very easily indexed by search engines. Furthermore it is a set of html files, one for each chapter of the thesis. He also maintains that the topic is very 'internet-centric' and so it is no surprise that his work is looked for online rather than using other information sources. The geographic location of download requests matches what he feels to be the major IT centres of the world (excluding Rhodes University) though he was surprised at the high ranking of France and lower ranking of Ireland.

The second most popular download lies at the opposite end of the higher degree spectrum: a 2006 DSc degree in Chemistry uploaded in July of that year. The author was surprised at how many download requests there have been and speculated 'people may think it is a free copy of an expensive Elsevier book that I co-authored … reactions of solids are important in many applications and there is very little published, but again I am surprised.' The third most popular document was a 2004 PhD thesis in Law uploaded into RUeRR in April 2006 and one used for initial demonstrations at Rhodes University - which the author feels is one reason for its popularity. He also feels that the topic is very popular 'the law of unjustified enrichment is a big thing in the US, Continental Europe and the UK' which means that the thesis would be found by search engines. The South African interest in the thesis is self explanatory as the thesis is about South African law, as are download requests from countries which he used for comparative information: USA, Canada, UK, Netherlands and Germany. He too is perplexed by some sources of interest: for example, Ukraine and Mauritius.

Clearly there is an interplay between the topic of the thesis, its disciplinary background and the patterns of access. There is indeed a broadening of participation from scholars with Ukraine, Mauritius, India, Ethiopia, Zimbabwe, China all featuring in the discussion in this section.

Concluding Remarks

This paper has focussed mainly on material found in RUeRR with only supplementary material from LiU ePress. Nevertheless we have outlined some trends showing how this new form of publishing research findings has altered patterns of access. Communities of scholars in the developed and developing worlds can now access material from which they were previously marginalised. This applies to the USA and Sweden every bit as much as it does to researchers in Ethiopia, Iran or Zimbabwe. Systemically, eResearch repositories have network characteristics that could be expected. What remains to be seen is how these will impact on conventional, for profit, publication platforms or on the publications of Professional Societies. The other agents in the system are our institutions that have implemented policies making mandatory the depositing of theses. What will be the impact on our Universities and the scholars researching in them of these new patterns and potentialities?

References

Antelman, K. 2004. Do Open-Access Articles Have a Greater Research Impact? *College & Research Libraries News,* 65 (5): 372-382.

Consejo Superior de Investigaciones Científicas (CSIC) 2009. Ranking of Web Repositories. [online] http://repositories.webometrics.info/ [Available: April 9 2009].

Dooley, K. 1996. Complex Adaptive Systems: a Nominal Definition. [Online] http://www.eas.asu.edu/~kdooley/caseopdef.html [Available: 15th April 2009]

Evans, J.A., and J. Reimer 2009. Open Access and Global Participation in Science. *Science* 323, 20 February 2009, 1025.

Horton, R. 2000. North and South: bridging the information gap. *The Lancet*, 355, 2231-6.

Hughes, Thomas P. 1983. *Networks of Power: Electrification in Western Society 1880-1930.* Baltimore: Johns Hopkins University Press.

Lewis, S. 2009. Repository Maps [online] http://maps.repository66.org/ [Available: April 9 2009].

Newman, M.E.J. 2003. The structure and function of complex networks. *SIAM Review* 45, 167-256 [online] http://arxiv.org/abs/cond-mat/0303516 [Available: June 28 2009].

Registry of Open Access Repositories (ROAR) 2009. Repositories registered Over Time. [online] http://roar.eprints.org/index.php [Available: April 9 2009].

University of Nottingham 2008. Directory of Open Access Resources. [online] http://www.opendoar.org/ [Available: April 9 2009].