Abstract:

Are advances in ICTs enabling positive transformations in academic research practices? This paper explores key themes emerging from a 2008 survey of researchers aimed at identifying use and non-use of advanced ICTs for research. Deterministic perspectives on e-Research suggest that new e-infrastructures will reshape research in predictable ways, such as by fostering more collaboration. In contrast, social shaping perspectives lead to the expectation that existing practices and institutions will shape the ways in which e-Research is employed. This exploratory study found that individual variations across researchers cluster into identifiable groupings of research practices,

* This work is based on the e-Infrastructure Project of the National Centre for e-Social Science (NCeSS) conducted by the Oxford e-Social Science (OeSS) node at the University of Oxford, supported by a grant from the UK’s Economic and Social Research Council (RES-149-25-1022). We thank all the members of the NCeSS Strategy Board, who collectively reacted to early versions of the questionnaire, and advised us on ways to refine the study. Rob Procter and Peter Halfpenny of NCeSS, and Ralph Schroeder and Marina Jirotka of OeSS were more directly involved in supporting the design and conduct of this study. Monica Gerber and Ellen Helsper provided invaluable assistance and feedback with our data analysis. We thank these colleagues.
which help to illuminate involvement in e-Research, and differences in methods and research software used across the disciplines. In line with a social shaping perspective, the findings suggest that, in the social sciences, top-down visions of e-infrastructure development need to be tied to the realities of bottom-up patterns of innovation in approaches to e-Research.

**Keywords:** e-Research “e-Social Science” survey “research software”

**Introduction**

e-Research initiatives have been launched around the world, and across the disciplines, including within the social sciences. However, given their early stages of development, we know relatively little about their use and impact on actual research practices and outcomes. The proponents of e-Research believe that the benefits e-Research, which encompasses the use of advanced Internet, Grid and related technologies for enhanced research, will extend to all of the disciplines, enabling more collaboration, data sharing and work with increasingly larger data sets. From this perspective, there is a deterministic perception that the social science community lacks a sufficient level of awareness of e-Research, and that this has contributed to a low uptake of advances in ICTs as tools for social research. In contrast, social shaping perspectives lead to the expectation that existing practices and institutions, such as disciplinary codes in the social sciences, will shape the ways in which e-Research is employed. These critical perspectives on e-Research view the proponents as overly deterministic, and see individual researchers and their disciplines as more actively shaping the design, up-take and implications of e-Research tools and data sets.

This paper addresses these issues surrounding the use of e-Research, particularly within the social sciences, through an analysis of survey data collected in 2008. An online survey was designed and distributed to gauge the ways in which researchers use software tools and platforms to enable research, and to gauge their attitudes and awareness of developments in e-Research. It
addresses such questions as: Who employs the tools of e-Social Science? Who is interested and aware of e-Research initiatives? Are researchers shaping e-Research, or are their practices being reshaped by the transformation in the technologies of research?

This article begins by describing the respondents, their main research methods, the software tools and technologies they use in their research, and their attitudes towards e-Research. It then focuses on evidence about the uptake and use of e-Research tools, along with levels of interest and awareness, and how these are related to disciplines and methodological approaches. These observations allow us to address some key issues in the debate between the technological deterministic and social shaping schools within the context of e-Research.

The findings suggest a potential mismatch between the visions of developing standard e-Research tools for the social sciences, and very eclectic patterns of bottom-up innovation even within particular methodological schools. Findings of this exploratory study suggest a number of policy implications, but there is a need for more systematic samples of researchers as one means to guide initiatives aimed at supporting the diffusion of e-Research in the social sciences and other allied fields.

**Methods and Sample**

In order to explore awareness of, and attitudes toward, e-Research, we designed a Web-based survey instrument, targeted at social scientists, but open to researchers from any discipline. The survey instrument measured demographics of researchers, attitudes towards e-Research, and, in the portion that will be reported here, research computing tool usage. These data allow us to examine the ways in which social scientists interested in e-Research are currently working with research software, and to compare what our results tell us compared to what we have observed developing within the e-Research community. These observations are based on our role as researchers at the Oxford e-Social Science (OeSS) node of the UK National Centre for e-Social Science (NCeSS)
programme. The OeSS node, currently funded from 2005-2011, has focused on understanding the social, legal, institutional, and ethical issues relating to e-Research. As such, we have been both participants and observers of developments in the e-Research area, particularly within the UK but also have been part of projects at the EU and global levels. The OeSS project has collected data and reported on numerous case studies, including cases in the sciences, medicine, social sciences, and humanities. This ongoing work informs our conclusions here.

The survey was pre-tested in November 2007, revised, and distributed from January to March 2008 using two mechanisms. The first was the use of targeted invitations to mailing lists obtained from the National Centre for e-Social Science (NCeSS) and the Oxford Internet Institute (OII). The second distribution mechanism was a generic version of the same survey that allowed anyone to take the survey via a weblink, which was distributed to a number of targeted mailing lists, including the ESRC National Centre for Research Methods newsletter, the NCeSS weekly and monthly newsletters, the Cybersociety Live mailing list, the Association of Internet Researchers (AoIR) mailing list, and the American Sociological Association Communication and Information Technologies (CITASA) mailing list. Recipients were also asked to forward this request to other appropriate lists.

The survey was completed by 526 respondents; the response rate for the most closely targeted e-mail list (NCeSS) was 23% (n=615), which is in line with other online surveys. For the more generic e-mail list (of OII academic contacts), the response rate was only 10% (n=1761), supporting the notion that many academics aren’t engaged with this topic at all. The final sample consisted of 526 valid responses originating from all sources. The data were collected with DatStat Illume and analysed with SAS and SPSS.

Given self-selection biases of online questionnaires, the sample was expected to over-represent those interested in e-Research. This is the case. The sample is skewed towards the interested: across the sample, when asked ‘How would you describe your interest in e-Social
Science initiatives?’ only 7 percent said they were ‘not interested at all’, while 57 percent reported that they were either ‘interested’ or ‘very interested’. While this bias should be taken into account when interpreting the findings, it can work in our favour, as our primary interest in this paper is to understand who is interested in e-Research, and how they use e-Research. Given the early stages of e-Social Science diffusion, it would take a very large sample of social scientists to obtain a truly representative sample.

Demographically, the sample was predominately composed of respondents from the UK (47%), PhD holders (56%), men (57%), and social scientists (55%). With respect to age and cohorts, the sample is fairly evenly distributed across age groups, with about one quarter of respondents in each 10-year span, but is skewed in terms of the date they earned their highest degree. 43% of the respondents received their highest degree after 2001; compared to 25% of the sample earning their highest degree in the 1990s, or the 13% from the decade of the eighties. The relative prominence of those with recent degrees indicates that a cohort of early researchers (regardless of age) might have more interest in e-Social Science. There is some support for this in the data (see Table 1), as those with highest degrees earned in 2001 or later are somewhat more likely to be enthusiasts (24% versus 16% of those with earlier highest degrees) and less likely to be sceptics (3% versus 8% of those with earlier degrees).

For the aims of this study, even with the limits discussed here, the sample allows us to understand the research activities, particularly with respect to research tool use, of those interested or engaged in e-Social Science and e-Research. Therefore, while this sample is not a representative random sample of researchers, it does reflect a large number of researchers with an expressed interest in e-Research.
Results

Much of the original e-Science and e-Social Science agenda was focused on the provision of top-down infrastructure to support large-scale computing, such as the Grid. Although more recently some e-Research proponents have been advocating a more bottom-up, lightweight approach\textsuperscript{11}, there is still a tension between those who feel that the best way to provide e-Infrastructure is via top-down, centralized systems and those who feel that more flexible but potentially less focused bottom-up approaches are more likely to advance research. Our survey raises some questions about whether a top-down approach is congruent with the actual practices of academic researchers, or alternately, to what extent researchers would be required to alter their research practices if they are to adopt e-Research methods.

First, we expected that interest in e-Research would be dominated by those whose methods were most in line with the early tool sets. However, we found that, at least within our sample, interest in e-Social Science sprang from across a wide range of methodological approaches to research, as shown in Figure 1.

[INSERT FIG 1 ABOUT HERE]

In Figure 1, two related sets of data are presented. The first, in the darker uppermost bars, represents the major research methods used by the respondents in this sample. Each respondent could indicate multiple methods; the mean number of methods across all respondents in the sample (n=526) is 3.9 methods per respondent (s.d. 2.3, range 0-12). The methods are arranged from top to bottom in order of the frequency with which they are used, with qualitative interviews the most common (57%). Other very common methods include desk research (56%), case studies (56%), and surveys (55%). The next plateau of usage is considerably lower, with participant observation,
ethnography and historical research all falling in the 20-35% range. Simulation, experiments, formal modelling and webmetrics are less common, with usage rates in the teens, and clinical methods are the least common in this sample, reported by only 3% of respondents.

In the lower bars of each set, we see the proportion of those who use each type of method who report being either interested or very interested in e-Social Science initiatives in the survey. Overall, the interest in e-Social Science is relatively high in this sample (interested or very interested = 56%, somewhat interested = 31%, not interested = 13%). There is not a lot of variation in support, as within each methodological approach there is a high degree of interest in e-Social Science initiatives, ranging from a low of 53% among clinical researchers to a high of 73% among those who do webmetrics. The only significant difference from the overall mean is among researchers who use webmetrics, which is not surprising since webmetrics focuses on using computers to uncover the connections and networks on the web, and is thus well-positioned to be engaged with collaborative and distributed research. So, although relatively few respondents report using these methods (14%), interest in e-Research among that small group is high.

The data above represent research methods, but we also asked respondents about their use of specific research software tools. Among our sample, which is generally interested in e-Research, we find a wide range of tools being employed. While many use tools for quantitative analysis, a wide array of other tools are also used by significant proportions of our sample (Table 2 and Figure 2).

[INSERT TABLE 2 ABOUT HERE]

Table 2 shows the categories of software tools that respondents use in the course of their research as the percentage of the overall sample reporting software tool use in a given category of tools, from most common to least commonly used. Here we note several differences with the previous figure, which reported general research methods. Whereas 57% of respondents in Figure 1
reported doing qualitative interviews, only 18% in Table 2 report using qualitative research software, indicating that the majority of researchers doing qualitative interviews are not analysing or managing their interview data using the qualitative software packages available. The most commonly used research tools in Table 2 are quantitative software programs such as SPSS, Stata or SAS (39%), and database software which includes Access, Excel and MySQL (34%). Note that Excel is included in this category even though it is arguably not always classified as a database. This is because many scientists utilize it as a storehouse for their research data. We let practice shape our categorization. In fact, Excel was the most commonly indicated tool in this category, with 67% of those reporting the use of database software indicating that they used Excel for this purpose. Other categories of research software are less common, with only 13% of respondents reporting using visualization software and dropping to the low of only 4% reporting using software to do video analysis.

In the survey, respondents were asked a series of questions first about the categories of research software they currently use. They also had the option for each category to select whether they have used it in the past but no longer do so, that they plan to use it in the future, that they have considered using it, or have never used it. For those who reported that they are current users of a software category, they were then asked to identify specific software tools within that category that they have used in the last two years. Respondents were only asked for detailed lists of software used within those categories that they reported being current users.

[INSERT FIG 2 ABOUT HERE]

Thus, within each category, we can get a rough measure of the total number of software tools that respondents have used in the last two years based on the number of specific tools they select after indicating they used tools within a category. The mean number of total software tools (in any
category) reported by those who use tools within a given category are shown in Figure 2. Here we can see that, consistent with the high level of interest in e-Research indicated in Figure 1, webmetrics researchers also are heavily engaged in using research software in general, using an average of 8.67 software tools (across all categories of tools) in their research.

The fewest software tools overall are used by qualitative researchers, who use an average of 5.09 tools in their research. Interestingly, researchers who report using quantitative software (\( \bar{x} = 5.5 \)) and databases (\( \bar{x} = 6.3 \)) are also among the lowest categories, in terms of overall quantity of tool use. One possible explanation for this is that several general purpose computing packages dominate these fields (such as SPSS and Excel), so researchers require fewer specialized software packages.

In the other research domains which are generally without a dominant computing package, researchers use a longer list of tools to accomplish their work. This would indicate that more specialized software is currently used in these research domains, which may present challenges to those building e-infrastructure to support e-Research: the more research software with more potentially incompatible technical and data format requirements that is needed to support research, the more difficult it may be to build a top-down research infrastructure. Of course, the counter argument to this is that a top-down approach that is focussed on building interoperability and ensuring standards may make the number of tools irrelevant, if the necessary tools can be combined efficiently via the infrastructure.

It is also worth noting that those fields currently using many tools are also ones that are currently quite active within the e-Social Science domain in the UK: geographic software, visualization software, and simulation software are all foci of NCeSS research nodes. The likelihood that researchers in these areas rely on a large number of tools suggests that efforts to expand e-Research in these areas will need to take this issue into account.
Another way to look at research software tool use is by the type of researcher. The complete method for the construction of the clusters is described elsewhere in this issue\textsuperscript{10}, but in short, a typology for the researchers in this sample was developed through the use of cluster analysis in SPSS that considered methods, skill sets such as coder or user, and collaborative styles such as being a sole researcher or one of a team. This cluster analysis found four types of researchers reflected in the data:

1) Lone e-Researchers, who are often the sole investigator, often or always coding or designing applications, and employing a mix of quantitative and qualitative techniques;

2) Team Players, who usually work as members of a team, develop and use e-Research, and use a mix of quantitative and qualitative methods;

3) Quals, who are primarily users of e-Research, and identify themselves as qualitative researchers, most often as a sole investigator; and

4) Quants, who usually work as members of a team, often coding or designing their own applications, and relying more on quantitative than qualitative research.

Figure 3 compares these researchers by the average number of software tools they use for research, and by their likelihood of using any software tools at all for research. Note that the average number of tools used is lower here than in Figure 2 because these clusters of types of researchers include individuals who reported no use of software tools, as opposed to Figure 2, which only included those who reported using tools in at least one of the categories.

Looking first at the average number of software tools used for research (shown in the bars on the graph), we see distinct differences between clusters and when comparing clusters to the overall mean. All groups are significantly different from the overall mean of 3.1 (± 0.3) software tools used for research. The Lone e-Researchers and Quants both use a greater number of research software tools than the other groups.
tools, on average, than the other categories ($\bar{x} = 4.4$). The Team Players cluster also uses more tools ($\bar{x} = 3.7$) than average.

The Quals, on the other hand, are likely to use far fewer tools than average and than the other clusters, using on average only 1.9 research software tools. This is also reflected somewhat in the other numbers reported on this graph, the likelihood of the respondent using any software tools for research. The Quals have a 60% likelihood that they will use software to support their research, at least in terms of data collection and analysis. (Word processing and activities such as e-mail were not considered as using research software since their use is nearly ubiquitous, as identified elsewhere in the survey.) While this is not significantly different from the overall mean (66% likelihood of using research software tools), it is lower than the likelihood that researchers in the other clusters used any software tools for research.

The team players have the greatest likelihood of using any research software (84%), followed by the lone e-Researchers (78%). Again, this points to the challenge that the builders of e-Research infrastructure face: the researchers most likely to be using computer tools for research fall into two very different types: the Team Players and the Lone e-Researchers. These researchers who are already using research software are likely prime candidates for engaging in e-Research; building an infrastructure flexible enough to support both teams and individuals, however, presents difficulties, particularly when change is coming from the top-down.

[INSERT FIG 4 ABOUT HERE]

Figure 4 shows the results when data similar to that in Figure 3 is grouped instead by the perspectives researchers have on e-Social Science initiatives, based on self-report. While both opponents and observers of e-Research are not significantly different from the overall mean, there are very telling differences in the other two categories. Those who consider themselves to be
promoters of e-Social Science are the group most likely to use a higher number of research tools, using on average 4.4 research software tools, and having a 77% likelihood of using any research software tools at all. At the other end of the scale, the disengaged use the lowest average number of research software tools ($\bar{x} = 1.8$) and have the lowest likelihood of using any research software tools (60% likelihood).

This is very telling, particularly to those interested in understanding why some scholars remain disengaged from e-Social Science initiatives. Not only are these researchers disengaged from e-Social Science, but they appear to be generally disengaged from the use of computer-based research tools in any form. While it is possible that these researchers will leap-frog from non-technological methods of research directly to distributed, collaborative, ICT-enabled research, we would contend that this is unlikely. Instead, it seems more likely that focusing efforts on engaging the observers, who represent a greater proportion of the sample (46%) than the disengaged (11%) anyway, will be more likely to have a chance of success, since they are less different from the promoters in the first place.

Figure 5 illustrates another challenge for top-down builders of e-infrastructures, and that is the prevalence of self-built, in-house software throughout categories of software. Prior to analysing the data, we had decided to include in our analysis any individual common software tools used by at least 10% of the users within any given category of software. When we analysed this data, we were surprised to note that in-house bespoke software exceeded the 10% cut-off in every category of software.

Figure 5 shows the top two software packages (which could be commercial or free/open source) per category, plus the frequency of use of in-house bespoke software. Even in categories
with dominant commercial packages, such as SPSS for quantitative software and Excel and Access for database software, in-house software is used by a surprisingly large proportion of researchers (17% and 15%, respectively). In other categories, such as integration software (56%), simulation software (49%), and content analysis (37%), in-house software is the dominant type of software within the category. In fact, for integration software, in-house software is the only type that exceeded the 10% cut-off point. That in-house software is used by more than 10% of researchers who use research software in any individual category, and is used in some form by 19% of respondents overall, suggests that supporting this sort of bottom-up development is crucial if e-Social Science is to expand.

Researchers are currently building software to support their particular research tasks. These surprisingly high numbers suggest several possibilities: one is that researchers are simply unable to find off-the-shelf software adequate to their needs, possibly when no existing software solutions are sufficiently flexible or suited to purpose, but would do so if such software were made available. Another is that research is diverse enough that it will always require task specific research tools. If the latter is true, this has implications for the amount of flexibility that must be built into research infrastructures if they are to support research innovation.

Finally, we examined whether age or the year of the highest degree predicted whether a researcher would use more research tools. One oft-heard comment is that younger researchers are more likely to be so-called ‘digital natives’, and thus more comfortable with using computer software as a research tool. However, at least in terms of the number of different software packages used for academic research, we did not find any difference between younger (≤35 years old) and older (> 35 years old) respondents.

Likewise, when using the ‘year of degree’ variable, those who have earned their highest degree in 2001 or later showed no significant difference in terms of number of software packages used for research from those who earned their highest degree earlier. When comparing the
likelihood that respondents reported using *any* software tools for research, again differences in age were not significant. In this case, however, year of degree was significant: of those who have earned a degree in 2001 or later, 71% (n=160) reported using software tools for research, compared with 61% (n=184) of those who earned their highest degree prior to 2001. This indicates that while age doesn’t appear to correlate with the use of software tools for research in this sample, having earned a degree more recently does have some correlation with the use of research software. Thus, having been trained in the finer points of digitally enhanced research, regardless of chronological age, might be more central than having grown up as a digital native in an electronic society.

**Discussion**

Interest in e-Research is not widespread, but a sizeable and growing number of researchers across the disciplines are interested and involved in e-Social Science. Interest spans a wide variety of methodological approaches and an eclectic range of research interests appear to be driving involvement in e-Research. We have found evidence of great diversity in the range of tools employed across our sample of 526 researchers, suggesting that there is unlikely to be any killer application of e-infrastructures beyond technologies and policies that support bottom-up innovation, such as evidenced in the prominence of in-house software development. There is little support for deterministic perspectives on e-Research which suggest that new e-infrastructures will reshape research in predictable ways, such as by fostering more collaboration regardless of institutional arrangements. In contrast, social shaping perspectives which posit that existing practices and institutions, such as disciplinary codes and research habits in the social sciences, will shape the ways in which e-Research is employed, have resonance in the results reported here. Thus, the development of top-down policies and infrastructures to support bottom-up research could be of significance for the strategic advance of e-Social Science. However, future research needs to utilize a more systematic probability sample to reliably gauge the diffusion of e-Social Science, such as to
know how widespread e-Research practices have become. The case for doing so is supported by this online exploratory survey.

References

5. Peter Halfpenny, Commenting in introduction to NCeSS Showcase, Manchester, 2008
8. Original survey, including variable names and descriptions, is available at http://www.oii.ox.ac.uk/microsites/oess/survey/
9. The methods for this project have been described in detail elsewhere in this issue, so we will provide a slightly abbreviated version here.
Captions for figures and tables:

Figure 1. Research methods and interest in e-Research

Figure 2. Mean overall research software tools used, based on use of tools in a given category of research software

Figure 3. Types of researchers and research software tool use

Figure 4. Researcher perspectives and research software tool use

Figure 5. Frequently used software tools

Table 1. Perspectives on e-Social Science by year of highest degree earned (n=524)

Table 2. Research software used by respondents (n=526)
Qualitative Interviews: 57% using, 59% strong interest in e-Research
Desk research: 56% using, 58% strong interest
Case study: 56% using, 62% strong interest
Surveys: 55% using, 59% strong interest
Participant Observation: 33% using, 64% strong interest
Ethnography: 25% using, 62% strong interest
Historical: 21% using, 62% strong interest
Simulation: 19% using, 60% strong interest
Experiments: 18% using, 60% strong interest
Formal modelling: 16% using, 56% strong interest
Webmetrics: 14% using, 73% strong interest
Clinical: 3% using, 53% strong interest
Other: 12% using, 62% strong interest
Mean overall research software tools used, based on use of tools in a given category of research software (n=526)

- Webmetrics: 8.67
- Geographic: 8.21
- Visualizing: 8.18
- Simulation: 7.93
- Content Analysis: 7.93
- Integrating: 7.84
- Video Analysis: 7.67
- Database: 6.32
- Quantitative: 5.53
- Qualitative: 5.09

Software tool category

Mean
Clusters of types of researchers by research software tool use (n=526)

- Lone e-Researchers (25.4%)
- Team Players (29.0%)
- Quals (32.3%)
- Quants (13.3%)
- Overall Sample

Number of tools

Likelihood of using any software tools for research

Mean number of tools used

- Lone e-Researchers: 4.4
- Team Players: 3.7
- Quals: 1.9
- Quants: 4.4
- Overall Sample: 3.1
<table>
<thead>
<tr>
<th>Perspective on e-Social Science</th>
<th>Year of degree</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before 2001 (n=300)</td>
<td>2001 or later (n=224)</td>
<td></td>
</tr>
<tr>
<td>Critic</td>
<td>1%</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>Skeptic</td>
<td>8%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>Observer</td>
<td>42%</td>
<td>46%</td>
<td></td>
</tr>
<tr>
<td>Enthusiast</td>
<td>16%</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td>Advocate</td>
<td>14%</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>No opinion</td>
<td>11%</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>Research software</td>
<td>N</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>205</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Database</td>
<td>179</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Qualitative</td>
<td>96</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Visualization</td>
<td>69</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Integration tools</td>
<td>68</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Geographic</td>
<td>62</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td>59</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Content analysis</td>
<td>59</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Webmetrics</td>
<td>44</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Video analysis</td>
<td>22</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>