

Collaborative yet independent: Information practices in the physical sciences

DECEMBER 2011



IOP Publishing **IOP** Institute of Physics



Acknowledgements

This report was the result of a collaborative effort between the Research Information Network, the Institute of Physics, Institute of Physics Publishing and the Royal Astronomical Society. They would like to thank the study authors at the 1) Oxford Internet Institute, University of Oxford, 2) Department of Information Systems, London School of Economics, 3) UCL Centre for Digital Humanities and the Department of Information Studies, University College, London, 4) e-Humanities Group, Royal Netherlands Academy of Arts & Sciences (KNAW) and Maastricht University, and 5) Oxford e-Research Centre (OeRC), University of Oxford.

The main authors for this report are: Eric T. Meyer, Monica Bulger, Avgousta Kyriakidou-Zacharoudiou, Lucy Power, Peter Williams, Will Venters, Melissa Terras, Sally Wyatt.

For the full acknowledgements, please see the project website:

www.rin.ac.uk/phys-sci-case

Contents

Executive summary	4	Tools and practices of information	68
Overview	4	Information sources	69
Method	4	Research software	77
Cases	4	Dissemination	78
Glossary	14	Complexity	79
Information in the physical sciences	16	Conclusion and recommendations	84
Background and related literature	16	Information retrieval	84
About the study	18	Information and data management	85
Approach and methodology	18	Data analysis	85
Particle physics	22	Citation practices	86
Astrophysics gamma ray burst	30	Dissemination practices	86
Nuclear physics	38	Collaboration	87
Chemistry	46	Transformations in practice	88
Earth science	50	New questions	90
Nanoscience	56	New technologies	91
Zooniverse and citizen science	62	Recommendations	92
		Endnotes	96
		References cited	97

Executive summary

Overview

In many ways, the physical sciences are at the forefront of using digital tools and methods to work with information and data. However, the fields and disciplines that make up the physical sciences are by no means uniform, and physical scientists find, use, and disseminate information in a variety of ways. This report examines information practices in the physical sciences across seven cases, and demonstrates the richly varied ways in which physical scientists work, collaborate, and share information and data.

This report details seven case studies in the physical sciences. For each case, qualitative interviews and focus groups were used to understand the domain. Quantitative data gathered from a survey of participants highlights different information strategies employed across the cases, and identifies important software used for research. Finally, conclusions from across the cases are drawn, and recommendations are made.

This report is the third in a series commissioned by the Research Information Network (RIN), each looking at information practices in a specific domain (life sciences,

humanities, and physical sciences). The aim is to understand how researchers within a range of disciplines find and use information, and in particular how that has changed with the introduction of new technologies.

Method

The study used seven cases, described briefly below, to understand the range of information practices across the physical sciences. In each case, data was gathered by interviewing scientists who were at various stages of their careers, and following these interviews up with focus groups to explore common themes emerging from the interviews. A total of 78 participants were involved, including 51 interviewees and 35 focus group participants (with 8 participants doing both).

Cases

The following seven cases represent different aspects of the physical sciences, using academic fields as the main way of defining a case boundary, but also including one department, and one case focusing on users of a resource.

Particle physics has a vibrant culture of developing and adapting information resources to suit both their extensive computational needs and large, geographically-dispersed collaborations.

FIELD: Particle physics

Information practices in particle physics are particularly well-studied. This is partly because particle physicists have been at the leading edge of new developments in information technologies for several decades, including the Internet, the World Wide Web, email, and pre-print repositories such as arXiv. The CERN laboratory in particular, where many of our case study participants have worked, has a vibrant culture of developing, using, and adapting information resources such as document servers, wikis, video conference tools, and other information management tools.

Particle physics, particularly as it is practiced at large research facilities such as CERN, requires collaboration, so researchers need to adopt or develop collaboration tools. The Grid is an example of one of the most advanced

collaborative tools in the world, as it allows distributed supercomputers, computing clusters, and data storage facilities from around the world to be linked to the desktop computers of scientists.

In terms of information sources, the particle physics participants in this case use Google heavily, but not Google Scholar. They use email lists and wikis, but rarely use libraries. They rely on the arXiv pre-print server, but do not rely heavily on general databases of articles. They use databases and programming tools to work with their data, write in-house software, and connect to the Grid. They do not, by and large, use software to manage their citations. In short, as with the other cases, they are early adopters of some technologies, but only when the technology meets their scientific needs.



FIELD: Astrophysics gamma ray burst

Gamma ray burst astrophysicists are unusual for a number of reasons, but one of the most interesting is related to the phenomenon they study: gamma ray bursts happen without warning, and usually last only for a few seconds. When a new burst is detected by space-based instruments, scientists are alerted to the event via text message or email so that they can quickly respond to observe the afterglow effects of the burst. The fast-paced, unpredictable pace of this type of research is in contrast to laboratory-based sciences, where experiments are planned long in advance.

This rapid-response approach is reflected in the information-seeking and publication patterns of the gamma ray burst community, where scientists read sources such as arXiv daily, and also rely on a centralised database of astrophysics articles (ADS) run by NASA. Many

results are released quickly to the community via short communications and notes, and via frequent conference presentations. A premium is placed on current information in this fast-changing field, and the tools the gamma ray burst scientists use reflect this.

For gamma ray burst scientists in this study, Google is much less important than arXiv and the ADS for discovering new information. Citation chaining, or following citations from one paper to the next, is a key strategy, as is information from peers and experts, often communicated informally. They rely heavily on bespoke software, and work with databases, programming languages, and image processing software. They do not visit libraries, and they do not use social network sites for their professional activities.

When a new gamma ray burst is detected by space-based instruments, scientists are alerted to the event via text message or email so they can quickly respond to observe the afterglow effects of the burst.

FIELD: Nuclear physics

Nuclear physics in the UK has been shaped by an unusual paradox: while nuclear physicists rely on major research facilities to do their scientific work, no facilities of this sort have existed in the UK since 1993. As a result, they must participate in international collaborations and travel to laboratories in other countries to do their experimental work. Nuclear physics is also distinctive from particle physics or astrophysics because a major branch of the field is directly concerned with very practical direct applications of science in the nuclear power, nuclear weapons, and nuclear medicine industries, among others.

Nuclear physics is a relatively small field, both within the UK and globally, and as a result, this case reported the least

concern with information overload, at least in terms of research information. Most of the important developments in the field of nuclear physics are published in just a few journals, and monitoring those journals allows researchers to keep up with developments in the field.

Important information sources for participants in this case reflect this relatively small pool of publications: the most common information source was browsing or reading online journals, followed by searching using Google and searching specialised databases. Because the key resources are so limited, keyword searching of journals was relatively unimportant. As with other cases, participants rely on bespoke software and databases as key software tools.

With a long history of shared document archives and global collaborations, nuclear physicists reported high levels of confidence in staying up-to-date on current research.



FIELD: Chemistry at Oxford

Chemistry as a discipline encompasses a range of fields and sub-fields, ranging from laboratory-based wet chemistry to cheminformatics, which relies on computer models. This case mainly recruited research students at a leading large UK department of chemistry at the University of Oxford. Thus, this case examines a mainstream chemistry department, but also explores new information practices engaged in by younger scientists.

The chemistry students in this case appeared to inhabit, simultaneously, opposite ends of the technological spectrum. Although they were by far the most likely among all the participants to use citation management software to organise information about research articles, they were also the most likely to print papers, and physically annotate them

with highlighter pens while reading. They were sophisticated users of advanced tools such as MATLAB, but also relied heavily on much simpler general tools such as Excel. The students reported that most of their information strategies were learned from peers just ahead of them in their careers (i.e. senior doctoral students and early career post-doctoral researchers). They found this domain-specific knowledge much more valuable than training in general information search strategies provided during their undergraduate training.

The participants in this case rely heavily on reading journal articles, browsing databases, Google, and peers for new research information. They rarely visit libraries, and make little use of Web 2.0, RSS feeds, or social networking sites for discovering new research-related information.

Students reported that most of their information strategies were learned from peers, and that they found this domain-specific knowledge much more valuable than training in general information search strategies.



INTERDISCIPLINARY FIELD: Earth science

The interdisciplinary field of earth science encompasses the study of geologic history, natural hazards, resource availability, and climate change, among other areas. Scientists come from fields including (but not limited to) volcanology, hydrology, seismology, climate science, geology and geophysics.

Unlike the particle physics and astrophysics cases, earth scientists do not rely heavily on pre-print archives. Instead, personal contacts were identified as a key way to keep up with new information. Earth scientists need to monitor a broader collection of journals than participants in some other cases, and thus were more likely to use tools such as the Web of Science or Google Scholar to search for information on a research topic. The participants also reported that computer programming skills are essential

for most earth scientists, since much of the work requires data preparation, processing, statistical analysis, and visualisation. Many of the advances in earth science are tied to technological advances in recent decades, including the widespread and cheap availability of GPS devices, remote sensors, satellite imagery, and weather data.

Participants in this case were among those most likely to see social media tools such as blogging as potentially important, but more as a means of communicating with the public and as a means for reaching out to young people than as research tools. In terms of important information sources, earth science participants relied on online journals, peers and experts, and citation chaining. Earth scientists in the study were the most likely (with the nanoscience participants) to use Google Scholar.

Many of the advances in earth science are tied to technological advances in recent decades, including the widespread and cheap availability of GPS devices, remote sensors, satellite imagery, and weather data.



INTERDISCIPLINARY FIELD: Nanoscience

Nanoscience, like earth science, is an interdisciplinary field, involving domains such as chemistry, engineering, biology, electronics, material science, physics, and medicine. Nanoscience is concerned with advancing science, engineering, and technology related to understanding matter in the 1-100 nanometre range. The resulting nanotechnologies are increasingly being used in commercial products including industrial, medical, and consumer applications such as clothing, food, and cosmetics.

The multidisciplinary nature of nanoscience is reflected in the diversity of information practices among participants. This case also highlighted the difference between academic scientists, who are rewarded for publishing influential papers, and scientists working in industry, where publications are not a major concern. In fact, industrial scientists have to protect the intellectual property claims

of their companies, so often avoid publishing their results. However, for both academic and industrial scientists, public outreach was seen as an important activity, whether it involved speaking to schools or setting up websites with educational content available.

The nanoscientists in this case all reported using Google and Google Scholar as an important source for research information. However, they also highlighted the frustration of finding useful articles that are not available via their institutional subscriptions. Searching databases, consulting peers, and following citation chains were all important strategies identified by participants. Libraries were seen as relatively unimportant resources, although there was awareness that subscriptions to journals were facilitated by university libraries. As with other fields, nanoscientists rely heavily on in-house, bespoke software tools.

Nanoscience spans several disciplines, bridging research and industry, resulting in diverse information practices among its participants.



USERS OF A RESOURCE: Zooniverse

The Zooniverse platform was set up to solve a particular problem: some scientific data requires human brains to process it in ways that are not currently possible using only computers and algorithms. The first Zooniverse project, Galaxy Zoo, enlisted the help of thousands of citizen scientists to help classify photographs of galaxies. The project has succeeded beyond all early expectations, resulting in the ability to classify objects at a scale one to two orders of magnitude higher than was previously possible.

Unlike researchers from other cases in this study, the scientists working with data from this project must deal with the general public on a sustained and regular basis. Interactions are important for prolonging the data-creation work of existing citizen scientists and for recruiting new ones. But they are also important from the point of view of

data analysis, as several new discoveries have been made by citizen scientists, who went on to become collaborators with the researchers. As a result, results are disseminated via traditional routes such as journal publication, but also on blogs and Twitter and other tools which can reach a wider audience of professional and citizen scientists.

The participants in this case were the least likely to use Google as an important tool for finding research information. Instead, they relied heavily on peers and experts, they browsed relevant databases, and were the only case to report a heavy reliance on Web 2.0 services. They were unlikely, on the other hand, to use Google Scholar, library materials, or wikis. Across all seven cases, the Zooniverse participants reported the highest use of in-house, bespoke software.

The Zooniverse platform was set up to solve a particular problem: some scientific data requires human brains to process it in ways that are not currently possible using only computers and algorithms.

Key findings

The physical sciences are a diverse set of fields, and the cases presented in this report illustrate the wide variety of information practices, research strategies, collaboration patterns, and dissemination methods used across the cases. Selected findings include:

- While general tools such as Google search are important, each field or sub-field also relies on specialised information sources unique to their field or discipline.
- Peers and experts are important sources of new research information.
- Information overload is neither uniform nor universal. While some fields are deluged with new information and express the need for better search and management tools, others find the pace of new research manageable, and report that current information tools are adequate.
- Data analysis takes up the majority of research time in many of the cases.
- Tools for data analysis vary widely, but all cases report a heavy reliance on bespoke software and tools built to serve a very particular set of research needs.
- Programming skills and the ability to work with data are increasingly a prerequisite for physical scientists.
- While citation credit is important for measuring productivity and impact, there is still little agreement on how to cite (or otherwise assign credit to) databases and the scientists and technicians who created them.
- Peer review remains important, but some fields are moving too fast for formal publication outlets to keep up. In these fields, various mechanisms have been developed to allow scientific results to be disseminated more quickly.
- Scientists are increasingly collaborative, although the size of collaborations varies widely by field and scientific topic.
- Computation is growing more complex as scientists generate larger and more complex datasets.
- General information practices in the physical sciences remain relatively simple.
- Disciplinary and field differences are evident throughout the data presented in this report.
- While few scientists report that technology has enabled them to ask completely new scientific questions, the cumulative effect of years and decades of advancing technologies has been that some scientific questions which would have been impossible to answer in the past can now be addressed.
- New technologies will certainly develop, but what effect these will have on science, information practices, and collaboration practices is unclear.

Recommendations

The report concludes with a number of recommendations, including the following.

- Several main barriers exist to better information practices including:
 - Lack of funding that supports the development of new field- or discipline-specific information tools
 - Lack of open access to scientific publications and data
 - Lack of methods for dealing with information overload
 - Inadequate annotation tools
 - Lack of funding for new tools for experimentation and data analysis
- Funders should prioritise increased efforts to share and link data.
- Funders and professional bodies should target postgraduate students and postdoctoral researchers with training in best practices for finding, managing, and disseminating information. This training will be most effective if it demonstrates concretely how their peers (scientists working in the same field) use these practices.
- New publication models need to be developed that expand access to published results and data, but which also support quality and long-term maintenance of resources.
- Publishers need to move beyond understanding their customers from a top-level disciplinary perspective, and begin to understand their audiences with more granularity and build tools and offerings that fit into the information practices of fields and sub-disciplines.
- Libraries need to be proactive in seamlessly providing access to information resources on- and off-campus, while educating their users on the role they play in negotiating and maintaining access to resources.
- There is a pressing need for all stakeholders to work more closely together as partners to build a more effective information ecosystem that serves the needs of scientists.



Glossary

The following terms appear in this report:

arXiv is an online preprint repository where authors can upload drafts of articles that have been submitted to, or recently accepted by, a journal. ArXiv currently has over 6,000 submissions each month, with a focus on physics, mathematics, and several other fields. Abstracts are archived and searchable by keyword, author, and date, with files of the entire article available as html links or as downloadable files, generally in Acrobat PDF, PostScript, and other specialised formats.

BibTeX is a bibliography tool designed to work with LaTeX.

Citizen science is the practice of engaging the general public in doing science, by contributing time or resources. Examples include not only the Zooniverse case discussed in this report, but also the BOINC distributed computing platform (<http://boinc.berkeley.edu/>), and non-technological citizen science projects such as the Audubon Society's Christmas Bird Count, which started in 1900 (<http://birds.audubon.org/christmas-bird-count>).

CERN Document Server (CDS) is a gateway to particle physics information which indexes the content of major journals in the field and harvests full-text articles from many pre-print servers, with most of the content coming from arXiv. The CDS's scope is more limited than that of the SPIRES database.

CERN INDICO (INtegrated DIGital CONference) **server**, which provides information about meetings together with the PowerPoint slides and minutes of those meetings.

DOI (Document Object Identifier) is an international system for persistent identification of objects located on digital networks. More detail at <http://www.doi.org>.

EndNote is a bibliography tool for Windows and Apple that works with Word, OpenOffice, and several other applications.

Gold Open-Access Journals are those which provide immediate open access to all articles.

The **Grid** is a globally-distributed system of computers (including supercomputers), data storage facilities, and high-speed network links that allows distributed computation and storage. In the UK, the National Grid Service (<http://www.ngs.ac.uk/>) provides core services and access to the global Grid.

h-index is a measure of the impact and productivity of a scholar. It is calculated as the total number of articles published that have been cited at least h times. In other words, if a scientist has published 25 papers, ten of which have been cited ten or more times and the remaining have been cited fewer than ten times, their h-index = 10. To increase their h-index by 1, 11 of their papers would all have to have been cited at least 11 times, and so forth.

LaTeX (pronounced LAY-tek) is a system for document preparation that has features for high-quality typesetting using document markup.

SPIRES is a search engine providing access to literature including journal articles, pre-prints, technical articles, theses, and conference proceedings. SPIRES and arXiv could be considered as a single system since SPIRES provides a front-end interface, as well as giving further context to the arXiv submissions by matching them with published literature and adding citations, keywords and other data.

TWikis are interactive wiki pages. TWikis are particularly important for particle physicists at CERN within the cases here.

Zotero is a free-to-use, web-based tool for collecting and organising citations.

The following resources are mentioned in this report:

ACS: American Chemical Society	http://portal.acs.org/	European Virtual Observatory	http://www.euro-vo.org/pub	Microsoft Sharepoint	http://sharepoint.microsoft.com
ADS: Astrophysical Data System	http://adsabs.harvard.edu/index.html	EVO: Virtual Organisations	http://evo.caltech.edu/EnablingevoGate	NASA SWIFT	http://www.nasa.gov/mission_pages/swift/main/index.html
arXiv	http://www.arxiv.org	Exoplanet Orbit Database	http://exoplanets.org	NASA: National Aeronautics and Space Administration	http://www.nasa.gov
arXiv astro-ph	http://arxiv.org/archive/astro-ph	Extrasolar Planets Encyclopedia	http://exoplanet.eu	National Nuclear Data Center	http://www.nndc.bnl.gov
ATELS: The Astronomer's Telegram	http://www.astronomerstelegam.org	GCN: Gamma-ray burst Coordinates Network	http://gcn.gsfc.nasa.gov/gcn	National Snow and Ice Data Center	http://www.nsidc.org
ATLAS (A Toroidal LHC ApparatuS)	http://atlas.ch	Gemini Observatory	http://www.gemini.edu	ROOT	http://root.cern.ch
BibTex	http://www.bibtex.org	Gemini Science Archive	http://cadwww.dao.nrc.ca/gsa	SAO/NASA Astrophysics Data System	http://adswww.harvard.edu
Brookhaven National Laboratory	http://www.bnl.gov	Global Volcanism Program	http://www.volcano.si.edu/index.cfm	ScienceDirect	http://www.sciencedirect.com
CERN CDS (CERN Document Server)	http://web.lib.cern.ch/	Hubble Space Telescope Data Archive	http://archive.eso.org/Sciencearchive/hst	SciFinder	https://scifinder.cas.org
CERN INDICO (INtegrated DIgital COncference)	http://indico.cern.ch	Huddle	http://www.huddle.com	Sixty Symbols	http://www.sixtysymbols.com
CiteULike	http://www.citeulike.org	HyperChem	http://www.hyper.com	SKA: Square Kilometre Array	http://www.skatelescope.org
Compact Muon Solenoid Experiment	http://cmsinfo.web.cern.ch/cmsinfo	IEEE Xplore	http://ieeexplore.ieee.org/Xplore/dynhome.jsp	Sloan Digital Sky Survey	http://www.sdss.org
Dropbox	http://www.dropbox.com	INSPIRE, which is replacing SPIRES in 2011	http://inspirebeta.net	Spinach MATLAB simulation algorithms	http://spindynamics.org
ECMWF: European Centre for Medium Weather Forecasts	http://www.ecmwf.int/	JAXA: Japan Aerospace Exploration Agency	http://www.jaxa.jp	SPIRES: Stanford Public Information Retrieval System	http://slac.stanford.edu/spires
ESA: European Space Agency	http://www.esa.int	Kavli Institute for Theoretical Physics Podcasts	http://online.kitp.ucsb.edu/online/si-podcast.html	T2K experiment (Tokai to Kamioka)	http://jnusrv01.kek.jp/public/t2k
European Southern Observatory Science Archive Facility	http://archive.eso.org/cms	LaTeX	http://www.latex-project.org/	TWiki	http://twiki.org/
		LHCb: Large Hadron Collider beauty experiment	http://lhcb.web.cern.ch/lhcb	UCSF Chimera, an Extensible Molecular Modeling System	http://plato.cgl.ucsf.edu/chimera
		Met Office	http://www.metoffice.gov.uk	Zooniverse	http://www.zooniverse.org
				Zotero	http://www.zotero.org

Information in the physical sciences

How do physical scientists find, use and disseminate information? How does this vary across fields and disciplines, and how are they similar and different from other types of researchers? How do the ways scientists arrange themselves, collaborate, interact, and work influence the kinds of information they use? The answers to these questions are complex and multi-layered, but this report begins to explore the ways that physical scientists are engaging with information in their research.

Background and related literature

This report examines the information practices of scientists across a sample of cases in the physical sciences. Recent innovations in the public understanding of science are also highlighted with an examination of the scientists who collaborate with public ‘citizen scientists’ in the internet-based Zooniverse group of projects. Information use in some areas of the physical sciences has been extensively researched, particularly with regard to their publication practices, which are unusual amongst scientists (Moed, 2007). Fields such as particle physics have been both early adopters and enthusiastic advocates of innovations such

as pre-print repositories (such as the Stanford Public Information Retrieval System (SPIRES) and arXiv) which serve as digital libraries for many physics fields (Nentwich, 2003). Physics is also making use of distributed (Grid) computing for tackling massive amounts of data (Pearce & Venters, 2012).

Other relevant research focuses on the information practices of a broader scientific community. Tenopir and King, in research spanning the last four decades, have looked extensively at the effect of digital technologies on information seeking and publishing. They conclude that the digital environment “has had a dramatic impact on information seeking and reading patterns in science.” The authors, based on survey evidence from U.S. science faculty at various universities, conclude that scientists read more articles, from a broader range of sources, found using search and citation chaining (Tenopir & King, 2008).

Much research, however (e.g. Nicholas, Huntington, Jamali, & Dobrowolski, 2007; Palmer, 2001; Palmer & Neumann, 2002) found that although the amount of material consulted may be increasing, the comprehensive (i.e. full-text) reading

of documents is declining in the sciences. This is because of the facility to keyword search within electronic documents and to quickly move from document to document via hyperlinks. Scientists are thus adopting more of a skimming or – when moving from document to document online – a ‘bouncing’ behaviour. Evans (2008) asserts that “searching online is more efficient [than browsing printed papers] and following hyperlinks quickly puts researchers in touch with prevailing opinion.” However, Evans adds a cautionary note, claiming – somewhat counter-intuitively – that “this may accelerate consensus and narrow the range of findings and ideas built upon” (p. 395).

Online availability of scholarly information has also transformed the way that scholars search, as researchers increasingly use a single interface to scan several resources, and most information retrieval happens at the researcher’s desktop (Hemminger, Lu, Vaughan, & Adams, 2007). Haines et al. (2010) suggest that this development means that researchers do not begin their search by limiting their results to what the institutional library has to offer. Google and Google Scholar are both popular tools for searching across the entire web, while other researchers use subject-specific search engines rather than publisher or library solutions (RIN, 2006).

One reason for this growing use of general search engines by scientists is that a greater selection of material, including grey literature such as conference or working papers, is openly posted on the Internet. Much supplementary data is now available in online repositories and often accompanies the electronic versions of journal articles in the sciences. UCL’s CIBER group found that physicists were now using the open access repository arXiv extensively to access working papers and pre-prints (RIN, 2010). ArXiv now hosts data sets, and the study also found that physical and life science researchers now value access to raw datasets as much as academic papers. Some (e.g. Attwood, et al., 2009) have argued that, for scientists, the use of supplementary material in journal articles will redefine what academic literature means in sciences and have implications for the reporting of scientific studies.

Issues in distributed collaborative work within science, such as communication and coordination difficulties, have also been extensively studied. Sonnenwald (2007) identified four stages of scientific collaboration from the existing literature: foundation, formulation, sustainment and conclusion. Within those stages, many inhibiting and facilitating factors were identified from the research literature, beginning with scientific, political, socioeconomic, resource accessibility, and social and personal networks. These factors remain

in place as the collaboration progresses, while new factors also emerge in further stages, such as the use of information and communication technologies and intellectual property considerations in the formulation stage.

The question of how to allocate publication credit appropriately is still open as Birnholtz (2006) shows in an examination of the problems around authorship and obtaining credit for large physics collaborations which may involve hundreds or even thousands of collaborators, impossible to list within a paper. He concluded that most physicists still think that informal attribution will not be enough to support career advancement, but that the problem of how to attribute credit properly has not yet been resolved.

Internet technology provides a number of services that are essential for collaboration at a distance (Nentwich, 2008). In particular, fast communication, resource sharing, version control and other groupware functions can sustain cooperation without face-to-face meetings. As a result, multi-disciplinary collaboration is increasing, and collaborative patterns themselves are changing (Nentwich, 2008). The number of individuals with whom a researcher can interact has expanded, providing greater access to potential collaborators and pathways for diffusing ideas. The new scientific tools available, such as the Grid, can foster an

environment which can organise collaboration among a much larger group of researchers (Nentwich, 2008). Emails, and other tools such as Skype, EVO and instant messenger, facilitate continuity of collaboration, increase the frequency of communication and can help sustain the sense of community among researchers.

About this study

This study is the third in a series of disciplinary case studies commissioned by the Research Information Network (RIN). Previous reports covered the life sciences (RIN & British Library, 2009) and humanities (Bulger, et al., 2011). The previous case studies highlighted some similarities that span fields and disciplines, but also a number of field-specific and discipline-specific practices. Life scientists, the first study found, were engaging in 'big science' at a conceptual level, but much of the day-to-day interactions still take place at a relatively small scale, at the level of the laboratory. In the humanities case studies, by comparison, collaboration is much less concrete: humanities scholars are part of large collaborative networks, but their collaboration is done via conferences, workshops and seminars, commenting on each other's work as part of an extended community of practice.

Approach and methodology

A series of seven targeted case studies were chosen to represent a range of fields within the physical sciences. In each of this series of studies, a slightly different approach to bounding the cases has been used because of the strong differences in how fields and disciplines organise themselves. In the life sciences report (RIN & British Library, 2009), the *laboratory* was the primary means of identifying cases, which fits well with the practices of the life scientists themselves, who are frequently organised into laboratories focused on particular streams of research. In the humanities report (Bulger, et al., 2011), a mixed approach was used which focused on *resources*, *departments*, and *fields*, to reflect the way that humanities scholars organise themselves.

In the current study, we aimed to cover a broad range of research practices in the physical sciences and therefore sought cases that would represent different aspects of scholarship within the physical sciences. The cases in this report are:

- Field: Particle physics
- Field: Gamma ray burst (subfield of astrophysics)
- Field: Nuclear physics

- Department: Chemistry
- Interdisciplinary field: Earth science
- Interdisciplinary field: Nanoscience
- Users of a resource: Zooniverse

As with previous studies, these cases do not exhaust the types of science or scientists in the physical sciences. But they do offer a rich picture of the range of information practices which are necessary to advance work in the physical sciences, and show the importance of field and discipline in understanding how science works.

Participants

Similar to recent exploratory studies of scholarship practices (Harley, Acord, Earl-Novell, Lawrence, & King, 2010; Meyer, Eccles, Thelwall, & Madsen, 2009; RIN & British Library, 2009; RIN & NESTA, 2009), we relied upon a combination of convenience and snowball sampling. The *convenience* aspect of our sampling involved contacting colleagues recommended by a known contact in the beginning of our study. *Snowball sampling* was used to identify other potential participants within a hard-to-reach group. Typically, in snowball sampling, one contact is asked to suggest additional contacts, who are also asked to recommend contacts. In particular, respondents were asked to identify other researchers with higher and lower levels

of familiarity and skill with computational resources as a way of broadening the sample. While these methods potentially introduce bias because they are not random, they allowed us to explore behaviours within relatively small academic communities.

We conducted 51 semi-structured interviews and five focus group discussions (with 35 participants, 8 of whom also participated in interviews), resulting in a total of 78 participants. As well as interviewing senior academics, junior researchers, and students, we identified database developers (3), project managers (1), and citizen science contributors (1). Some scientists acted in dual roles as faculty members and database developers or programme managers. To provide a broad perspective of scholarly resource use in the physical sciences, we also included graduate students (16) and postdoctoral scholars (8). In total, scholars from 32 institutions in 9 countries participated in our study. Table 1 provides a description of participants within each case and case type.

Table 1: Participants by case

Cases	Interview Participants	Focus Group Participants (n interviewed individually)
Particle physics	10	4
Nuclear physics	7	0
Astrophysics: gamma ray burst	7	9 (2)
Chemistry graduate students at University of Oxford	6	7 (4)
Earth science	6	4
Nanoscience	8	9 (2)
Zooniverse	7	n/a
TOTAL	51	35 (8)

Process

Scientists were invited via email to participate in the study. Whenever possible, we conducted in-depth interviews face-to-face, but when distance or timing precluded this option, we used Skype, often with video conferencing enabled. Interviews usually lasted one hour, though we allowed additional time for elaboration and discussion. We also asked participants to complete an online survey.

Once interviews were complete, we conducted focus group discussions. Focus groups allowed us to explore themes emerging from the interviews in more depth. The focus groups also provided an opportunity to speak, for example, to graduate students after interviewing faculty, or citizen scientists after interviewing developers.

Analysis

Interviews were transcribed and themes were identified via qualitative coding; the cases were then written up using these themes for structure. Survey responses were analysed quantitatively using SPSS and Microsoft Excel. We performed frequency analysis of resource use and communication practices. Additionally, we conducted cross-tabulations to explore relationships among groups.

The multiple methods employed involved collection of information behaviours through personal interviews, focus groups, and surveys. By triangulating these different data sources, we have secured an understanding of the information practices of the physical scientists who participated in the study. But our findings should not be taken as being representative of all physical scientists. The aim was to conduct a short exploratory study in order to identify a range of practices and so wider generalisations across these communities might be premature. Nevertheless, we are confident that the report provides relevant insights into transformations in research practice, and their implications for researchers, institutions, and funders.

Particle physics

For decades, experimental particle physicists have worked as a globally distributed collaborative community that thrives on democratic debates and discussions (Knorr Cetina, 1999). Their collaboration has been described as ‘exceptional’ (Chompalov, Genuth, & Shrum, 2002) and the way they work is unorthodox compared to other sciences (Zheng, Venters, & Cornford, 2011). Members of this community are highly technically competent and operate within a culture which accepts the ‘good enough’ (Kyriakidou-Zacharoudiou, 2011) – using solutions which might be a bit messy around the edges but are very innovative. The community has always been at the frontier of computing and electronics, with the development of the Web being a notable example.

Experimental particle physics was selected for this study because it has pioneered innovative solutions in the field of information management and dissemination (Gentil-Beccot, Mele, Holtkamp, O’Connell, & Brooks, 2009). Almost half a century ago, faced with the slow process of peer-review journal publication, particle physicists began mailing their colleagues copies of their manuscripts (Goldschmidt-Clermont, 1965/2002). This led to the creation of the first electronic database for grey literature, which evolved into

a database of the entire subject literature, called SPIRES. In the last two decades, critical innovation in scholarly communication emerged from this community, from the invention of the web (Berners-Lee, 1996), to the inception of arXiv, the first online pre-print repository (Ginsparg, 1994).

We interviewed ten researchers from CERN’s three largest Large Hadron Collider (LHC) experiments, ATLAS, CMS and LHCb, as well as from the T2K, a non-LHC experiment. We interviewed three senior academics, two lecturers, three postdoctoral researchers and two PhD students. Following the interviews, we conducted a focus group discussion with four of the original interview participants.

Information retrieval

Most of the information resources used by the particle physicists were specific to their field or to the broader discipline of physics. During our interviews and focus group discussion, the most frequently-mentioned included general resources such as Google, email, and learning from peers. In addition, the resources which are common in many areas of physics were identified as key resources, in particular arXiv and SPIRES, plus the resources hosted by CERN, including

the CERN Document Server (CDS) and the CERN INDICO server (for meeting-related information).

Recent work suggests that particle physicists begin their searches with these more specialised tools (Gentil-Beccot, et al., 2009). But this study found that most participants began their data collection process with a web search on Google, which they believe is a “stepping stone to everywhere.” Most argued that it is quicker to use Google than any established resource in the field, especially as it comes up with suggestions. But since arXiv and SPIRES are indexed by Google and partly organised in Google Scholar, Google tools are simply an overlay on more established sources of information. Most participants never accessed SPIRES, apart from updating their publications, as they find the interface complicated. A few indicated that they would only use arXiv for scientific paper searches if they knew the exact reference beforehand, while others accessed arXiv frequently and some used it as their first choice of resource.

Google and, to a lesser degree, Google Scholar, are used as a starting point to locate relevant research, with most interviewees reporting using Google for almost everything (general searches, paper searches, code searches, etc.) and all the time. One said, somewhat jokingly: “If it’s not on the first two pages of Google, I’d probably never find what I’m looking for.”

While most participants were aware of other information resources provided by their universities (such as library catalogues or online access to the Web of Science), these were generally seen as inflexible search tools and limited in terms of content and so were rarely used. Some scientists occasionally accessed Web of Science to update their publications (as required by their universities) but only one regularly accessed it when writing academic publications. Most participants reported that they never access journal websites directly. Several said that by the time a paper gets to a journal, it is almost out of date and so they rely more heavily on pre-prints than journal publications. Most also reported accessing books very rarely and only when searching for historical information.

Most also accessed publicly-created tools such as Wikipedia to gather information when they start working on a new subject area, saying that “for science and technology Wikipedia seems very good.” All participants used their experiments’ TWiki pages on a daily basis. These provide a wide range of content, including technical details such as how to undertake analysis to information about approved publications, references and guidelines to prepare talks. As an ATLAS interviewee explained:

The TWiki provides information about the day-to-day practice of a particle physicist working in ATLAS... it’s a set of pages that we ATLAS users can create and alter. It provides information that we wouldn’t be able to find from somewhere else. The TWiki’s updates are to keep in line with whatever the latest changes are.

These TWiki pages provide links to the frequently accessed CERN INDICO and CDS servers. Several participants reported accessing CDS through their experiments’ TWiki pages because “the TWiki provides already-filtered information and therefore displays only the things that are useful.”

Learning from peers and experts is valued, and all participants reported frequent communication as well as formal and informal discussions with colleagues when faced with problems. One stated: “The way we learn is through the word of mouth. I mean, the stuff I hear over in the common room, it’s just amazing, that’s an important source of information.” This reinforces the importance of personal relationships as a source of professional information, and also demonstrates that co-location remains important in a digital world.

Information management

When participants find a useful paper, most read the abstract and skim-read the full paper online, bookmarking them on their browser, since “it’s easier to search for things on the web.” Many also save the relevant papers on their personal computers, filing them by paper theme. One used CiteUlike – a website that stores references one finds online. It was more common to use emails as a personal archive. Participants emailed themselves with interesting papers, URLs, or any other useful information they found online. One participant stated:

Every document you receive or send comes by email unless you pull it off the web. And therefore email is the filing system. I tend to email myself to tell me where I found specific information – I try to keep my folder structure simple so I know roughly where to look. So, for example, if I was doing some work and I looked up a paper, probably more likely, I might save the paper on my computer, but I would actually just email myself and say, ok, I’m working on looking at how to do something, and here are the sources which I found on the web and I will paste the URLs into an email to myself.

The participants subscribe to mailing lists and use email as one of their first or second resources for acquiring information. Most did not delete any of their email, and

some saved their email on their hard disks where they perform frequent searches. A few participants received more than 300 emails per day.

Some interviewees printed a document if it was a key resource or was difficult to read on screen, then reading it on paper and making annotations. Others mostly read on their laptop or other electronic devices: one participant sometimes used an Amazon Kindle and another sometimes used an iPod Touch. To annotate texts within a digital file, participants used Adobe Acrobat Reader tools, copied and pasted relevant passages, or wrote notes in a digital document such as Microsoft PowerPoint, that they later reorganised or copied and pasted into their article drafts or talks. They explained that articles are usually prompted by people’s comments on a document – called a note – that leads to discussion and is eventually uploaded to the experiment’s TWiki page, or to the CDS for further debate.

For word processing, most interviewees used LaTeX and were strong supporters of the tool, with only one senior particle physicist using Microsoft Word.

Data analysis

Most estimated that, when faced with a new problem, about 20% of their time was spent searching, 70% was spent analysing and resolving the problem, and 10% was spent

checking their solution (e.g. if their code works). Information search and analysis are not always distinct topics and are often performed simultaneously. Information acquired through search is used to construct codes, resolve any errors and make work more efficient. Most participants used the programming language C++.

Analysis is a complex process involving intense programming, and requires collaboration. Most participants argued that the complex problems they face mean that it is impossible for one person to do the analysis work on their own. Analysing collisions is extremely complex and understanding the effect of the detector is difficult, particularly in distinguishing new physics results from the general messiness of the data using simulation data for comparison. This requires extensive parallel computing power using the Grid and bespoke software, as well as tools such as Root, a statistical analysis tool which generates documentation from comments in their code as well as graphs and plots in a way which respondents argue is similar to Matlab but more sophisticated.

All participants said the Grid was their most valued technical resource for performing any scientific analysis of their data. The Grid, in brief, is a distributed network of computational resources which can be used to divide up large data processing and analysis jobs. This connects supercomputers

and computing clusters to the desktop computers of scientists. Participants argued that without this Grid, full-scale processing of their data would be difficult. One participant said: “For full-scale analysis I just go to the Grid. It’s usually a pain at the beginning because you have to adapt all your code to go to the Grid. But when it works, when you get it to work, it’s just amazing. It’s so fast and you can process millions and millions of events in a couple of hours.”

Citation practices

For most participants, citation practices depend upon the resource and the type of information in question. Most tended not to cite technical information, or anything acquired from Wikipedia, the TWiki or email exchanges with colleagues. Pre-print articles are cited, since the gap between finding and publishing information is so large that the most relevant and up-to-date information is found in pre-prints rather than published articles. One participant gave more detail:

For brief write-ups or for conference proceedings and internal documentation we cite the preprints with the arXiv reference number (if the preprint is not yet published). For articles that are submitted to journals the preprint will generally only be cited if it’s accepted by a journal (and has a DOI) or has been published.

Finally, in the case of preprints which introduce computer programmes, are a programme manual, or describe a technique and are not likely to be published, then I have cited them simply by the arXiv reference number (usually for programmes also with a link to a web page).

Our interviewees had high confidence levels in materials disseminated by their experiment collaboration. As one said: “You usually start working within a subgroup of people, focused on similar topics, and then you have to go to the large group, and then to the larger group, and then you send your draft to the whole collaboration and so at each step you’ve got editorial boards that are reviewing and questioning every aspect of your analysis. So things are looked with a lot of care so that you don’t go and publish something that is not done properly.” Similarly, another interviewee remarked: “High energy physics really – they kind of vet their own papers before they go out and so it’s very different than other areas of science in which people just write a paper and publish and then it’s just the journal reviewers that have the final say.” All participants felt that the community reviewing process is much stricter than a journal and in order for a paper to be approved – and to be uploaded on arXiv – it has to be accepted by the whole collaboration. This gives them the confidence in the content of such papers and means that they are happy to cite them.

Approval of the collaboration is more important than the number of citations received. One participant said that “the most cited papers are the ones that are wrong ... citation on its own is not a measure of anything.” Most said that “the bar is set very high for the standards of authorship within the experimental collaborations,” and individual recognition is a clearer indicator of good work than the number of citations.

The most common tool for managing and creating citations among participants is BibTex, a referencing database tool for LaTeX.

Dissemination practices

Particle physics collaborations are managed and coordinated through a complex network of channels, involving individuals and groups from different layers in the collaboration. Dissemination therefore occurs through members’ collaboration with colleagues via frequent face-to-face or virtual meetings through EVO, a video-conferencing tool, via extensive email exchanges and telephone conversations, or through the collaboration’s TWiki pages.

Although participants used traditional means such as publications, internal notes, and conferences to disseminate their findings, they all emphasised that frequent meetings and emails are the most important ways of disseminating information and knowledge. One participant stated: “The

real communication goes on by email, by group meetings where, you know, people get up and give progress reports about their work. We have lots and lots of meetings, some are daily, some weekly or biweekly, where we usually present talks. So in these working groups we get to show progress and even say, like, we are stuck in this part, and having this problem, what should I do now or what are your recommendations?" One participant had exchanged three thousand emails with a single colleague in the past year.

Participants also stressed the importance of informal face-to-face conversations over coffee breaks and meals, discussions in corridors, or by socialising in the pub. Expertise and knowledge is seldom lost because information is disseminated through these avenues, including dynamic documentation which grows with time. One participant said that "if someone is a real expert, then people will follow him around in order to learn from him and acquire his skills." Another way that individuals learn is by volunteering to perform tasks not relevant to their job descriptions, a common practice among all physicists.

The particle physicists we interviewed used a number of technical tools for disseminating information. One, for example, described frequently using the CDS to give feedback when participating in a note. He remarked that CDS is well-suited to this task because it has a built-in

system that allows members to publish notes, but it also has a mechanism for submitting comments and for the authors to then respond to those comments in a structured way. Another tool all participants frequently use for disseminating information about the meetings taking place is the CERN INDICO server.

While most did not use Twitter or blogging to disseminate their research, one respondent maintained a blog to discuss research topics and a few others used Facebook for work-related issues such as arranging meetings or telephone conversations. Senior scientists were more likely to use EVO and the telephone for conversations with colleagues, but doctoral students and postdoctoral researchers used instant messaging tools (such as iChat or Skype) for constant interaction with peers (e.g. for short and brief communication amongst fewer than 5 people). However, for 'public meetings', where anyone is welcome to attend, they too preferred EVO.

Collaboration

Particle physicists have a long tradition of being a collaborative community. Their work demands expertise in different fields of enquiry, and therefore collaboration between various globally distributed academics is key. These collaborations are established and maintained

through relatively simple means, including email, mailing lists, putting relevant material on websites and the TWiki pages, uploading shared programming codes to repositories, phone conversations and video conferences and face-to-face meetings (formal and informal). Researchers needed to become involved in the community in order to have a clear overview of what is going on, to learn how to 'be physicists' and to acquire a set of skills.

Collaborations have very informal organisational structures, with no clear division of labour. The leader uses charisma and soft leadership techniques in order to drive their community forward. The decision-making process is based on discussions, compromising and convincing; a decision is approved when they reach consensus.

Particle physicists rely on trust, autonomy and volunteerism. While this is true of any academic field or discipline to some extent, for particle physicists these are not just background traits. One interviewee stated: "We have trained ourselves ... that collaboration is one of the most powerful tools." Most interviewees were generally given freedom to carry out their work, usually without clear instructions or strict supervision because, in their view, their community involves people with commitment, intelligence, and self-motivation. All participants explained that their work demands faith and trust in what other people have done. Members valued

reputation, and recognition of expertise is important. Maintaining their reputation as good collaborators motivated them to complete tasks on time and keep their projects on track.

Those we interviewed often mentioned the shared goal of 'doing new physics' as being one that drives them. One participant stressed the crucial role of keeping all members engaged and making them feel ownership of the project by providing a set of structures which give individuals recognition, by building community bonds, by making information available and by inspiring them to work towards the common goal: "The shared goal is very important. If you don't have that, there isn't this common view of where we're aiming for. So it's very important that we have, you know, high-level aspirations to real significant physics discoveries, and that binds people together...It's a way to make them belong." All participants mentioned the importance of spending unstructured time together and establishing personal relationships with their collaborators, something which again helps build a sense of belonging and ensures efforts are directed towards the shared goal.

Most agreed that, although competition exists within the wider community and different experiments have their own personal goals, competition is minimised in order to achieve the higher goal. One senior participant stated: "So, we have

slightly different goals, we all have our own physics analysis we want to do, but the means of achieving that higher goal is collaboration; that's why collaboration becomes then the natural tool, much more so in our sort of type of community than, in, say, the corporate structure where I think the shared goal peters out after you get down to the first few layers of management. And I think that we have this history and we have learnt that collaboration works." Being open helps them collaborate, and minimises internal competition since, as one participant said, "there is no need to compete about the things you've previously done and resolved."

All the people involved in an experimental collaboration are included as authors in publications. Most participants were pleased with this, as it acknowledges the important effort of those who did the physics but did not author the paper. One participant explained: "They contributed to the building of the detector, running the detector, making sure the software worked. So without all these people you couldn't have done the physics, and there's never one person who writes the paper, anyway."

Finally, identifying collaborators through traditional means, particularly via word of mouth, is important. Apart from the technical skills required, being familiar with the particle physics culture and mentality is important for membership in this community.

Transformations in practice

Most transformations in practice were related to new possibilities offered by the World Wide Web. In the past, physics was performed by small groups of people in physically-isolated locations and information was disseminated via journals. Today, particle physicists work in "virtual research communities or virtual organisations, where scientists work together and there is no isolation in terms of physical location." One participant said:

The Internet is the way which binds us into this common laboratory. For example, if you're working in this building and one of your colleagues is in the basement doing some interesting experiments, I wouldn't wait for him to publish it, I'd go down and talk to him, right? And that's what the Internet does for us.

The Web, therefore, has facilitated the growth of this virtual laboratory. Particle physics experimental collaborations now consist of 2000-3000 widely-dispersed scientists. More senior participants reported that their collaborative practices had become more democratic. One explained:

A lot of the earlier experiments...were better described as benign dictatorship. There was the leader of the experiment who would surround himself or herself with a small group of people who represented the power

places – so perhaps the big institutes – but, basically, it was much more dictatorial. I think as it becomes bigger there has to be much more obvious democracy. So lots of boards and committees, and processes...It does mean that you sometimes pull your hair out because you have to get permission for everything. But I don't see there's any alternative.

Similarly, another senior participant stressed that “the bigger the collaboration, the more communication channels needed to make it work.” All agreed that their collaborations require more effort to keep all people engaged and make them feel ownership. Information must flow continuously and researchers need regular discussions and social activity to stay focused on the shared goal.

The web, and particularly communication technologies such as Skype and EVO, have made such communications easier. As one participant said:

The video-conference call is now very easy to use and I think it enables things that couldn't happen before. So the fact that it's very, very easy now just to set up a conversation with somebody in a completely ad-hoc way, and very quickly without hooking anything or going to a special room – that probably had more effect than we thought. I don't think five years ago we would have seen that we were all going to be sitting in our office basically plugged into headsets all day using EVO and so on.

Developments in web 2.0 technologies have also affected communications. While static websites and emails predated the web within particle physics, wiki pages provide a much more interactive element to their collaboration and make them more interconnected. All participants indicated that the wiki has become a mainstream way of communication within the community.

Real-time communication is and will become more important than email over time. As one interviewee stated: “Mailing-list style communications are going to die away. I guess in two or three years' time, you know – already actually, looking at my young post-docs, those guys don't make telephone calls anymore. They're simply on real-time sort of chat the whole time. They don't even email each other anymore. In fact, if you conduct your business by email you're regarded as being a bit of an old-timer. So I guess this is sort of creeping up.”

Most felt that major changes in their research practice were due to innovations in technology and availability of information online. All visited the library very rarely, and rarely consulted books or peer-reviewed journals, as they are 'out-dated'. One reported:

I haven't actually used a library for about twenty or thirty years, and, so, in our field, you don't use libraries because any information that gets into a library in

printed form is almost certainly out-of-date. Even published papers, because of the long refereeing and publishing process, they're already a year out-of-date. So, anything I want to know in terms of my research is going to be in electronic form these days, because it's going to be pre-publication, it's going to be pre-library, it's going to be even pre-most journals, so, most of the information I know exists electronically.

New questions

Most respondents felt that although access to information is faster and easier, this does not mean that they conceive or answer new questions per se. Rather, it enables them to address the questions already conceived.

Some believed that the work they currently pursue would not have been possible without networked technologies. One participant said:

What has changed now is the questions that we're trying to answer in the LHC experiments, which require a huge amount of organisation and collaboration to actually get a handle on. So some of these questions involve analyses and operations of the detectors and data handling and processing that is so complex that we would not be able to address some of these questions in a group of just a few people, we wouldn't be able to do it even in

one country. The scale is just too large. So, if we had attempted to do this without the sort of information resources and the communication resources we've got now, I'm not sure it would be possible.

The doctoral and postdoctoral participants in our study had never experienced particle physics without the web, and so did not observe any major changes in their way of working. But more senior participants clearly indicated the difficulties of doing physics 20 years ago, where more research was done by mostly co-located small groups of people compared to today.

New technologies

Most participants were satisfied with the way they acquire information. Nevertheless, they identified some limitations among current field-specific technologies. For example, most did not use SPIRES or arXiv because of their complex interfaces and inflexible, non-intuitive search tools. Senior participants, in particular, were more comfortable with the simplicity of Google. A number of interviewees wanted to enhance SPIRES and arXiv with a user-interface and a search tool that is as easy and as intuitive as Google's.

The most common complaint among particle physicists was information overload – particularly non-relevant information – and the need for tools to overcome this.

One suggested enhancing Google and their TWiki pages with “some sort of scoring system providing a hierarchy of quality,” something similar to Amazon's star ratings.

When asked what other communities could learn from particle physics, all participants suggested their culture of collaboration. As they argued, other communities have issues of strong competition, mostly because the problems they have to tackle are not so large-scale. Most also highlighted the value of spontaneous, unrestricted communications within a research group. Expertise should be shared in order to avoid information loss, and this can only be achieved with lots of communication and socialising. As one said: “I'm just thinking about how things used to be and if we were still very compartmentalised, we were still working within university groups, we were still not able to collaborate quite freely around the world, then we would not be able to do these experiments, and I think other people will soon find themselves in the same situation.” However they acknowledge that collaboration does come with a price as “every paper has 2000 people in it.”

Participants described themselves as early adopters of new tools. Their mentality is to always be prepared to invent a technology if it is needed in order to explore new physics. One stated: “If you need a tool, be very prepared to think about your communication, your information resource requirements, and if you need a technology which is not

there yet, then be prepared to invent it. But be prepared, when you're thinking about your project and your organisation, and how much funding you need, to build that in, right. So this sort of meta-project that supports the big project, yeah? Allocate resources to communications and to information sharing, because otherwise you get a problem with a big project, it's just going to fall apart.” They believe this approach provides an important lesson for other communities.

Participants also mentioned open access journals, highlighting the importance of freely-available information. One interviewee explained that CERN has taken the “brave decision to publish half of the work in open access journals where there is no subscription fee.” He hopes that this decision will change the publishing industry in general and that other communities will follow CERN's example.

Astrophysics gamma ray burst

Gamma ray bursts are rare cosmic events that produce high energy light, and can only be detected from space-based instruments. Studying the emissions of gamma ray bursts enables scientists to understand better how galaxies form and evolve. Gamma ray bursts also provide information about radiation emitted by matter accelerated to close to the speed of light. They are transient and usually last for a few seconds, thus necessitating a rapid response. The position of an event is radioed to the ground by an observing spacecraft within a matter of seconds and scientists receive notification within minutes via text message or email. The scientists can then quickly request observations of the afterglow effects using space satellites and ground-based telescopes. In some cases, the alerts are sent directly to telescopes for automatic observations seconds after the burst.

The transient nature of gamma ray burst events, coupled with the need for rapid response, places unique demands on researchers, particularly in the area of data collection and collaboration. We explored these practices by interviewing six senior academic astrophysicists studying gamma ray bursts, plus an additional senior astrophysicist studying high energy gamma ray astronomy. We held a focus group

discussion for six astrophysicists at the University of Leicester where two interview participants, two graduate students, and two postdoctoral researchers participated.

Information retrieval

Typically, a peer-reviewed journal article takes between six months and a year to go from submission to publication. The nature of work in this field demands rapid dissemination, and thus researchers frequently post papers to the astro-ph area of arXiv. From our discussions, most gamma ray burst astrophysicists read arXiv every day to stay abreast of new work in their field. Other uses of arXiv included literature searches, to re-locate articles previously read, and to check citation details. One participant described arXiv as “pretty close to essential.” Most also used SPIRES to determine which papers are citing other papers.

The Astrophysical Data System (ADS), a searchable database of astrophysics journals run by NASA, was similarly described as an indispensable means of accessing relevant articles. Most described ADS as capable of more precise searches than arXiv, since its database only contains astronomy journals and allows for searching a range of

dates, whereas arXiv can only limit searches to a specific year. The ADS also maintains links to citations for each paper. The consensus among participants was summarised by one respondent: “In the vast majority of cases it’s so straightforward with ADS and arXiv, the combination of the two, you get 99 percent of everything you’re after.” In fact, they reported that Google does not serve a primary search role in their research because they have such good systems in place. Speaking about ADS, one senior academic said that, “unlike Google, which searches everything, this searches trusted sources.”

The search strategy of the participants depended on the nature of their task. For example, a researcher might use telescope images to conduct research and analysis, but would use Google Images to get a general sense of an object or to find a photo to use for teaching. Senior academics reported training their undergraduates and masters students to use ADS for journal searches and arXiv for daily reading of new publications. They also host weekly paper discussions over coffee to discuss recent findings because “the pace is so rapid, no one is an expert, no one is up-to-date.” During the focus group, senior academics spoke about trust and the challenge in approaching the vast body of literature¹ during undergraduate and early graduate studies. They described the way in which trust is built by

familiarity with the literature, for example, identifying well-known astronomers and following their work, or avoiding the work of individuals seen to publish lower-quality findings.

Gamma ray burst astrophysics also boasts a wealth of object and image databases. One senior academic said of these resources: “Information seeking takes little time because the tools are strong.” Most of these tools provide advanced search facilities, including object names and co-ordinates, and some provide additional services such as data download, bibliographic references and citation counts relating to specific objects. Another astronomer used archived data provided by the Hubble Space Telescope Science Data Archive, Gemini Science Archive, and the European Southern Observatory Science Archive Facility, all of which provide keyword searches on object name or coordinates. For some tools, such as the Extrasolar Planets Encyclopedia and Exoplanet Orbit Database an astrophysicist reported that “a lot of people are using that as their source of data rather than the published literature.” These exoplanet websites post current observations and format data for download and analysis.

Information management

Most participants in our study retrieved papers from arXiv and ADS each time they used them, rather than storing them on a computer or printing them and storing in a folder or filing cabinet. One researcher said: “It’s so quick to find things that often I just open them online and view them directly.” Another said: “As long as you have subscriptions through your university you can get whatever paper you like and download it and you can cross reference everything and so on. So it’s a very convenient system – so half the time I can’t remember where the paper was and so I have to go and find it again.”

This practice of retrieval may be due to the importance of very up-to-date information for this group of researchers. For example, a user of the Exoplanet Orbit Database described challenges with storing the data: “As soon as we download and keep it, it instantly becomes out-of-date, because one can look at these things and see that they’re constantly being updated.” But when using data tables, most either print or download the database files, with others maintaining a plain text read-me file alongside their notes. Most preferred to keep notes in a digital file because they believed they would lose handwritten annotations. Those who printed papers seemed to do so for a particular purpose. In addition to data tables, astrophysicists reported

printing papers that seemed important, were complicated, or contributed to a collection of similar work. One senior academic described printing around 4-5 papers per day and organising them in box files by topic. When the files are large enough (10-12 articles), he concludes that he has enough background literature for a paper.

Data collection

The study of gamma ray bursts requires a number of satellites and ground-based telescopes to record observations. Most data used by astronomers in the gamma ray burst community are collected by instruments on satellites maintained by NASA, the European Space Agency (ESA), or Japan. These data are combined with observations from ground-based optical and radio telescopes. Data from the satellites are stored in databases using a standardised format so that they are accessible internationally, across projects and facilities. Thus, databases around the world store archives of the observations, all using the same standardised format.

Some projects, such as NASA's SWIFT, make image data available immediately to the public. Participants in our study believed that making information public improves and serves the scientific process. Additionally, scientists

believed that access to these events increases public engagement with their work. During the focus group, a postdoctoral researcher described feedback received from amateur scientists expressing their excitement at getting observation time alongside the astrophysicists and having the opportunity to see what the scientists see.

Other projects identify data as proprietary for a specified timeframe—usually one year—so that project members who contributed to the development and implementation of the spacecraft or telescope can conduct analyses and publish their work. Typically, scientists request an observation block in which they can focus on pre-determined coordinates.

Data analysis

Most participants use customised software for their data analysis that has either been developed at their institution, been developed by project members, or been developed for earlier projects. In the 1980s and 1990s several large, general purpose suites of astronomical software were developed, but funding for such initiatives has declined in recent years. This makes it difficult to continue to support the existing software. Because the field is moving so quickly, much software for astrophysics is developed in response to problems, or to support analysis of data from a particular instrument, rather than in anticipation of needs.

Analysis methods depend upon whether an astrophysicist is theoretical or observational. For theoretical astrophysics, synthesis has value, so researchers read arXiv daily to stay abreast of broad topics and understand how they apply to other aspects of astronomy. Theoretical astrophysicists also undertake modelling and simulation, which face limitations of time and computation: “Sometimes the data sets are so huge that it takes a long time to crunch through them. We run into that problem not as much with data analysis, but when we want to do a simulation of some object in the sky... there can be huge simulation codes that run for weeks on supercomputers to get a result.”

Citation practices

Despite heavy use of arXiv and ADS to learn of new research and to carry out literature reviews, researchers tend to cite the journal source rather than the database. Exceptions occur when an article is not yet published and is only available on arXiv. ADS will link to these citations, but will update links from arXiv to journal publications as they become available. For those who use BibTeX to organise their source material, citations from ADS include a note that appears for the writer, but is optional for the final text: “Provided by SAO, Smithsonian Astrophysical Observatory/NASA Astrophysics Data System.” However, most do not feel it is necessary to reference the database. One participant summarised this practice: “I think that they (ADS) request that people who make use of these services mention them in the acknowledgements of their papers, but I think now that they’re so universally used, in our field at least, people often don’t bother, because they think ‘Well, it’s obvious that I used ADS.’” Therefore, bibliographical reference counts may not be a reliable measure of the popularity of these resources.

Astrophysics is known for long author lists that can number more than 500.² There is an expectation within the field that funding sources and tools will be acknowledged. An acknowledgement may appear as follows:

Based in part on observations obtained with the European Southern Observatory’s Very Large Telescope under proposals 077.D-0661 (PI: Vreeswijk) and 177.A-0591 (PI: Hjorth), as well as observations obtained with the NASA/ESA Hubble Space Telescope under proposal 11734 (PI: Levan).

This example includes the tools used (i.e., European Southern Observatory’s Very Large Telescope), the grant identifiers showing the sources of funding, and the Principal Investigators for each project (e.g., PI: Vreeswijk). In addition, resources, such as satellites or telescopes, are thoroughly acknowledged in the data analysis section with the aim of providing enough information for others to reproduce the analysis. This information can include versions of software tools used and databases used.

Dissemination practices

Traditional forms of dissemination, such as conference papers and articles in peer-reviewed journals, remain important for recognition and promotion within the field. But because the field moves quickly, other techniques are employed for rapid dissemination, such as shorter notes. Much of the early scientific communication on gamma ray bursts is through ‘circulars’ on the Gamma-ray burst Coordinates Network (GCN) or, more rarely, the

Astronomer’s Telegram (ATELS). These are typically posted within a few hours of a new observation. This same system is used to rapidly send out the GRB alerts (positions and times) as ‘notices’.

Circulars are brief notes describing the object observed and are sent to email lists that have members numbering in the thousands. These circulars are available online, archived by ADS and mark observations, with details such as instrument used and location. Circulars serve as the basis for later papers.

One participant described the field as “quite a talky community...It’s not one talk a year, it’s more like ten I would say for some people.” In addition, astrophysicists attend several face-to-face meetings with their collaborators to discuss their research.

The majority of gamma ray burst astrophysicists we spoke to do not post to blogs or Twitter, but do make use of online communication tools within a project. Such tools may include wikis, large email lists, or, in rare instances, Facebook groups. In the case of the exoplanet community, its main websites are acknowledged as credible information sources and observations posted to the site are favoured over published journals as an information resource because of their currency. While the participants did not report frequently posting to blogs, they did report reading them.

In fact, though, most of the blogs described by participants were not research-oriented, but job rumour websites or discussions about research politics or funding.

Most established researchers maintain personal or departmental websites with links to their publications, but younger scholars use arXiv as a way to direct others to their work. In fact, a study showed that scientists post to arXiv at specific times to ensure their article appears at the beginning of the list (Gentil-Beccot, et al., 2009), since this portion seemed to receive a higher number of citations (Haque & Ginsparg, 2009). One senior academic feels that arXiv and its daily digest serves to democratise exposure of new research: “This morning, I looked at arXiv, and I went through today’s lot and I printed out what I’ve got in front of me now, four or five articles. They could be written by anyone, so in that sense, they get far more democratic access to my time than if I went to the library and looked at the preprints.”

Collaboration

Gamma ray burst research requires access to facilities across the world for satellite and ground-based images, meaning that most telescopes and satellites are shared. Observational astrophysicists therefore engage in relatively large, geographically dispersed collaborations. External collaborations are still small compared to nuclear and particle physics, with up to 100-150 members, and researchers also collaborate internally with research groups co-located at universities or observation facilities.

The intensity of contact varies, depending upon the collaboration’s purpose. Some collaborations set out to build an instrument or satellite and therefore involve many meetings and email conversations to discuss logistics and to coordinate efforts. Other collaborations are focused on astronomical events that will result in a paper, which also involve, to a lesser extent, emails and phone discussions. When a gamma ray burst event occurs there will be an explosion in communication between collaborators, again usually by phone or email.

External collaborators primarily use email for communication. Collaborators often meet at conferences or schedule week-long meetings or infrequent teleconferences. Some groups use open or password-protected wikis to

communicate recent activity; but a project wiki: “tends to die once its sort of immediate reason for existence has passed.” The usefulness of a wiki depended heavily on whether the scope of the project warranted the extra effort needed to visit another information source. Others preferred wikis as a way to reduce emails and to better organise correspondence.

The rapid response required in this area of astrophysics results in fluid methods. As one gamma ray burst astrophysicist said: “You don’t tend to set up video conferences, you tend to just—when one of these gamma ray bursts goes off ... you get the email, you get the text message on your phone telling you that the burst’s gone off, and then everyone sort of scatters—races to do things.” A collaborator on one project may be a competitor for another, with groups coming together around a shared interest or funding stream, and re-configuring for the next project. In addition, researchers help keep spacecraft functioning well and able to do the science. Thus, different teams will share responsibility for being on-call.

Transformations in practice

Many of the participants we interviewed divided their time between teaching and research. When considering transformations in their teaching, a few described the ease with which they can now find illustrative images on Google to share with their undergraduates. Others described the value of podcasts – the ability to refer students to expert lectures and discussions about cutting-edge research, citing in particular, the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara. In fact, a few believed that podcasts represent a strong resource which is sometimes underused due to current limitations in organisation and categorisation.

Most participants in our study described the same transformations in their information use, such as preprints replacing journal subscriptions. Similarly, most do not visit the library. However, one senior academic reports that he still reads Nature in its printed form over lunch, citing it as an exception to his online reading practices. The cohort reported that the Web of Science Index, once the primary source of publication information in their field, has been “rendered redundant” by ADS. Senior members of the field recalled the anticipation with which their department would receive printed pre-prints, which have now been replaced by daily digital digests. Additionally, access to unpublished

articles has increased significantly in the past 10 years. Indeed, the amount of material, both data and journal articles, has been steadily increasing.

The Internet has enabled more data to be shared with the worldwide scientific community and the general public. This sharing has also resulted in larger collaborations. The change in the volume of data and its online accessibility also means that research itself has expanded: for example, instead of studying one object for a thesis, current students can potentially study a thousand. As one senior academic reports: “You can tackle a computer-intensive problem more easily today simply because the information is online and you can get it into your computer easily. You don’t have to type it all in from a journal as you may have had to do in the past, and all the data from the spacecraft and the telescopes are online, so you can just grab those data straight into your computer and process them.” Some of those interviewed argued that this easier access to data has increased the signal to noise ratio, meaning that more papers are published, but that quality has not necessarily increased.

When discussing resources and interdependence of facilities, many of our astrophysicists also mentioned the fragility of the current system. One senior academic highlighted that much of the general-purpose software built for astrophysics and in use worldwide is over ten years old: “They don’t

continue to develop, and the best you can hope is that they at least don’t die completely because we’re all still using the same programme.”

Budget reductions around the world affect astrophysicists’ capacity for collaboration, data collection, and data analysis. For example, some space missions and ground-based telescopes are the result of international collaborations. When one country reduces its funding for astronomical research, it therefore affects these collaborations, putting additional pressure on other teams to support the effort and potentially causing scientists in the country which has cut its budget to be denied access to future data.

In a field where rapid response is essential, these types of challenges place researchers in a potentially vulnerable position. In discussing cuts to personnel, one scientist described concerns that too many cuts may impinge on the smooth running of space operations: “There’s no sense in people starting to tell us, oh, well, three months ago there was a very interesting event but I’m afraid it’s gone now, you’ve missed it. It’s an operation that either succeeds in real-time or it fails altogether, really.” Another astrophysicist at NASA said: “The future is challenging. It’ll be hard to sustain this level of observation in the future because there won’t be as many observatories.” When describing collaborations with NASA and ESA, one researcher said,

“You are dependent on those agencies carrying on funding them, and if they didn’t then we’d have a major problem.” Indeed, given the international interdependence upon the resources provided by agencies such as NASA, a reduction in resources would potentially have ripple effects to other programmes.

Fragility was also evident in the field’s dependence on NASA’s ADS system for journal access: “it had become, after just a few years, so completely indispensable, there was a suggestion that they were going to pull the plug financially and there was a mass outcry from the world’s astronomers.” In addition, arXiv recently requested donations from its users, so astrophysicists are aware that the systems upon which they depend are potentially vulnerable to funding cuts.

New questions

Technological developments in the past decade have significantly advanced gamma ray burst research. Improved communications capabilities have enabled more rapid sharing of information and opened access to international datasets. Data is available for analysis through publicly available databases, allowing researchers without direct access to a specific satellite or telescope to conduct analysis. They also enable researchers to compare events across datasets: “It was very hard in the past to assemble data from different observatories and different satellites and put it all together, and that’s become much easier because of these archive centres where all the data is collected in standard formats.” Astronomers said that in the recent past databases required unique knowledge to use, making the process of extracting information and comparing it across datasets challenging.

Additionally, advancements in computational capabilities have allowed for larger simulations and 3-D modelling. Thus it is now possible to simulate how a star explodes in three dimensions, with more precision and detail. Indeed, many astrophysicists said the improvement in precision and sensitivity of the instruments had advanced their work. Currently, satellites cover a broader range of the electromagnetic spectrum, enabling more sensitive observations of the spectrum. A researcher at NASA described the difference: “It wouldn’t have been possible in the past to decipher what it all means nearly at the speed we’re doing, or maybe even at all.”

New technologies

When asked about a wish list for future technologies, most participants instead spoke of the need for sustainable, long-term funding and job security. Researchers with permanent jobs expressed concerns related to research, in particular access to telescopes and the sustainability of current initiatives, but researchers employed by grants worried that project funding would be cut and their jobs discontinued. Recent graduates were concerned that secure positions will be minimised. One postdoctoral researcher said that he “can’t see past the next bid” because his employment hinges on grant awards.

Due to budget cuts, the UK recently pulled out of an international collaboration with the Gemini observatory, a partnership of seven countries that originally included Australia, Brazil, and the US. Formally, UK scientists will be locked out of the partnership. More broadly, scientists are concerned that this decision affects the image of the UK as an international partner and collaborator.

While most participants believed that barriers to future advancements relate to lack of funding, rather than the limitations of technology, a few improvements were suggested. Many expressed a need to link the different databases and archives of data so that searching for an

object would show all the information available for it. One senior academic suggested a system in which he could input coordinates for a specific part of the sky and it would pull all images that exist, together with all information published. A preliminary system has been attempted by the European Virtual Observatory; however, it encountered challenges in coordinating different archives. Further, they hoped the system would allow for online real time analysis. A few mentioned Enabling Virtual Organisations (EVO), a collaboration network hosted by Caltech, as a possible way to meet these needs. Its distributed architecture allows for large file sharing of high resolution images.

Another senior academic described a way to access the collective knowledge of experts – a system that includes podcasts of lectures and discussions, but extends further to a repository or database of what experts are listening to and reading, so that one could follow important sources based on who is reading them rather than waiting for citations to appear in the literature. Since a majority of the cohort wished for filtering tools that could sift through the vast amount of information currently available, this suggestion offers a means of identifying key works and could serve as a virtual supplement to citation chaining and peer recommendation.

Nuclear physics

In the United Kingdom, nuclear physics is an important area of scientific research, but has faced a series of funding challenges over the past two decades. The last major nuclear physics research facility in the UK was closed in 1993 as part of an economy drive. Since then, nuclear physicists largely rely on international collaborations and must travel to international laboratories to carry on their research. While there are still some small facilities, according to respondents “local work is mainly for student training” and “nuclear physicists have to go around the world to find suitable laboratories” to do research. The UK nuclear physics community is relatively small, with fewer than 100 researchers and similar numbers of doctoral students (Ion, 2009).

The nuclear physicists who were part of this case study ranged from pure nuclear physicists, who study topics such as the structure of nuclei and how energy generation occurs in stars, to very practical applied nuclear physicists who work with energy generation facilities to understand the processes of nuclear energy production and often have a strong engineering aspect to their work. The issue of building new equipment, however, can apply both to those working with the nuclear power industry and those contributing to international research facilities.

Information retrieval

One of the most important information sources identified by respondents is the National Nuclear Data Center (NNDC), often referred to by respondents simply as the 'Brookhaven Database' since it is housed at the Brookhaven National Laboratory in the United States. Participants noted several advantages of the Brookhaven Database over tools such as Web of Science:

- *“It searches journals for specific information, and you can look up authors as well [as topics], which is quite useful. Because it’s a restricted range of journals to do with nuclear science, it tends to be very quick and relevant.”*
- *“The other nice thing about Brookhaven is that it doesn’t require any special login. It’s open access, whereas [for] Web of Science, I have to make sure I’m connecting through the right route.”*
- *“The National Nuclear Data Center...has a huge amount of experimental data on different nuclei as well as a...reference database”*

One added-value aspect of the NNDC is that it includes databases which have been carefully constructed to combine information from multiple data sources to find the best overall evidence:

There may be several papers that have been published on the same isotope measuring the same information... [but] the individual peer review can only peer-review the paper and the effect it has on past history if the referee knows about the past history. [There is a] data network of people who do evaluations, so periodically, they will look at all the isotopes with [a particular] mass...For each isotope, they will sit down and look at all the papers published on that particular isotope and actually evaluate the information. They will look very carefully at the different papers and try to make a judgment as to whether the information all agrees and is consistent or whether there’s a rogue paper where the data is off, and they will try to understand why, and they’ll try to come to some conclusion about what is the best data set to use, and that’s the evaluated data set that they then put on the webpage.

Because the field of nuclear physics is relatively small globally, most respondents felt that they had a comprehensive sense of which journals they needed to consult in order to stay on top of developments in their field. As a result, participants tend to go straight to the journals they know first, and only turn to search engines when looking at a topic more broadly:

You go directly to the journals if you know what you want to look at. And deciding what you want to look at involves either your own knowledge, based on what’s been written before, what you’ve read, references in previous papers, and so on. So if you know where a paper is that you want to read, you would go directly for it. If you’re sort of doing a more general survey of what has been done, then you’d probably start with some of these other tools, like Web of Science and the SPIRES database, and start searching there for specific topics, usually, or specific people who you know have been working in that area.

This strategy is consistent across the interviewees, which overall reflected a mature field that is not suffering from an information deluge. Because of well-established publication norms in the field, there is a relatively constrained set of information sources which need to be tracked:

I go through the list of contents online and download the papers that I want to read, and I keep them on my computer, and then if something comes up that I want more depth in, I go to the reference list of the paper that gave me some information. I might do a category chain through references, and seeing what that reference quotes... [By] starting off in very recent journals, then I believe I’m capturing [information] fairly well. And

then if I get very serious, I will also do a database search, too. But usually, they're way behind the recent publications in keeping up to date with the data.

This practice of citation chaining seems to be particularly useful in small fields such as nuclear physics, where many of the researchers know each other's work, and often know each other personally via conferences and experimental collaborations.

General tools such as Google were also mentioned as important for search, but this was generally in the context of starting to research a new idea and trying to discover whether anyone was already working on the topic. Some participants did not see themselves as particularly proficient information searchers, but they did not believe that this had hampered their careers in any significant way since the information in nuclear physics is tightly bounded and discoverable without sophisticated strategies. One of the participants, for instance, remarked in the interview that prior to having filled out the user survey for this project, he had never even heard of Google Scholar, but felt he "should take a look at it" to see if it might be useful.

Information management

In the interviews, most participants report that they rarely visit libraries in person: "I think in this day and age, the days of trotting up to the library, getting a book or a journal out, and reading it or putting it through the photocopy machine are, thankfully, in the past," and "I'll tell you, I used to love and go looking up historical things in the library. It all got put in the archives, now, and it's very hard to get them out... there's no browsing of old articles any longer." Nevertheless, the participants in this case were one of the few groups to report awareness that their libraries are key players in maintaining subscriptions to the online journals they require—that is, they are aware of who holds and pays for the subscriptions:

We get access to them because our library has a subscription to various journals. And I'm not quite sure what the mechanism is, but when I log in from my computer here, it automatically recognises that I'm on the [university network], and it gives me access to these things. Now, there are various ways of getting them. You can go through the university library site and get completely lost, because they seem to have some other idea of what they're about, and finding electronic journals is—you know, about the tenth page in, you find the link. So, I mean, that's what we use the library for, but finding it that way is not straightforward.

At least one respondent mentioned the library can still be an important resource:

Books and things like the good old conventional textbook. There are some very, very good textbooks, surprisingly enough. And going back and really trying to understand things from those is not a bad place to start. I would use a textbook for trying to understand the fundamental principles. So often you'll come across something in a piece of research that somebody's done, and you don't quite understand what it is that they've done and why it is that they can do it, and it sometimes requires some background reading to really understand what the context is. Textbooks are very good for that. I'm thinking reasonably high level textbooks, of course, but not—so not introductory material, but with a good library, a good university library, you have those sorts of things.

These uses of library resources and materials show that libraries still play a role, although their ability to communicate this to researchers is uneven.

Information management strategies in nuclear physics reflect the simple ecosystem of resources described above. Many of the respondents indicated that they have done more reading on screen in recent years, but printing out an article and reading it offline is still relatively common: “I have colleagues who have a sort of database of relevant papers, but I’m a bit more chaotic; I tend to sort of just look for them when I need them ... if there’s an article which is particularly important or interesting, I usually print it out and carry it around with me for a bit, you know, then try and read them on the train or something.” For those who read on screen, the main disadvantage was the lack of annotation tools to mark up the on-screen copy. One, for instance, had recently acquired the professional version of Adobe Acrobat which allows annotation:

Well it’s a bit variable, but I’ve now got an Acrobat thing that I can stick post-it notes on with and make yellow blobs on the screen ... Yes, so highlighting, as with a marker pen. But I’ve only started using that quite recently. But yes, it seems good. Otherwise it’s rather laborious, writing notes in a separate file.

One disadvantage of digital files is the relative inconvenience of the reference list. Rather than being able to quickly flip to the last page of the paper article, the reader needs to scroll to the bottom of the document, and then try to find their place back in the text of the document. One participant got around this by making a separate file for the reference list when there were a lot of references they wanted to consult. But this navigation problem was balanced by the big advantage of digital documents: search. According to one participant, “especially if I’m researching something that I’m less familiar with, then I’ll look for keywords in a range of papers and find them, and I don’t have to read it all. So there are some big advantages in electronic versions.” This also is reflected in reading patterns that can be more cursory, which is consistent with previous research (Tenopir & King, 2008). One participant said: “When you just download, you think, ‘Oh, that looks interesting,’ and you’re more browsing and speed-reading it than if you had walked all the way down to the library to look at it.”

The convenience of electronic access also affected participants’ choices about what to read: “There are some journals which we don’t have access to online. Of course, what that means is that you tend not to use those journals anymore.” When faced with a journal to which they do not have a subscription, most participants simply looked for other easily-accessible information sources – unless there was an important reason to track down an article by requesting a PDF from the author. Several participants mentioned journals which they knew might contain interesting research, but that were only available to them when they were visiting institutions in the United States, so they did not bother to follow what was published in those outlets. Several also remarked that they would like to access articles on their personal laptops while travelling with the same ease as while working in their offices, but that they did not know whether this was possible at their institution.

Data analysis

While nuclear physics experiments generate large amounts of data, it is not at the same petabyte scale as that generated by particle physics experiments which require the Grid for storage and analysis; the volume of data is increasing, however, and some nuclear physicists are increasingly involved in Grid computing and using cloud-based storage services. Many interviewees used a scaled-up version of a method that dates back to the earliest method of transporting punch cards and reels of magnetic tape: carrying the data by hand. The low cost of portable hard drives, coupled with the fact that the scientists already travel from their home to a research facility to perform experiments, has resulted in widespread use of this very simple, cost-effective and reliable means of moving large amounts of data from experimental facilities for analysis:

We bring some of the data back [to our facility] for analysis; some of it is pre-processed on site, and then reduced amounts of data are brought back for analysis. Getting it back is an issue; it can be done over the Internet, but that's slow ... [so it is] often put on disks and just brought back in someone's suitcase.

Typically, now we would take a portable hard drive. I think you can get a terabyte portable hard drive these days, which can fit in your hand luggage. So we take that and bring it home after. If we go to some accelerator, they would have their own copy as well. It would be archived somewhere...It's good for travel, physics, these days. All you need is a laptop, really.

The analysis of these specialised data, like many of the other physical science cases, relies on bespoke software that is passed around between colleagues and among collaborations: “The detailed data analysis, I would say the best description is software developed by colleagues. Sometimes PhD students, but quite often somebody in the Americas has written something that's useful. Then it gets shared around the community.”

Citation practices

The nuclear physics participants, as with many of the other cases, do not have clearly articulated practices with regard to citing databases. According to one participant:

If people were going to refer to...the sum data as a whole, they would normally refer to the edition of Nuclear Data Sheets in which it was published, so the journal rather than the database. However, I think if you're referring to specific subsets of data, I would always refer to the primary source – in other words the journal paper it came from – because otherwise it's unfair on the people who've done the original work.

This question of how to best refer to data so as to acknowledge the contributions of others is one that many fields are still trying to resolve.

Another aspect of citation practices was raised in this case as well: the ascendancy of bibliometric measures such as citation counts and the h-index. One participant noted that “citations...in terms of management was a curiosity 10 years ago, but...now if I'm sitting on a promotions panel or looking at applicants for a job, I will use that information to help me evaluate.” The citation measures are not being used blindly, since participants noted you can have a high citation count for bad science as well as good science, but noted that the information available by measuring citations is much more important today than it was in the past.

Dissemination practices

The main dissemination strategy is traditional publication in peer-reviewed journals. There was little evidence of less traditional dissemination routes, although many interviewees would include their articles on their department's webpage. This seems to be related to the fairly well-bounded nature of the field: the relatively small number of journals relevant to nuclear physics means that as long as work is published in those journals, authors can be reasonably sure that it will be seen by the right people. According to one nuclear physicist, "99 percent of our scientific output is reported through conferences or published in scientific journals."

A nuclear physicist working on the applied end of the spectrum offered a different perspective. For him, a key output was helping to populate important databases with experimental data: "It's actually very detailed, very intense, and the major aim is not to get a publication in Nuclear Physics or Phys Rev Letters and so on; the major aim is to get that database right, because that's what people are going to use." This particular scientist had established a career path that relied on measures of success other than journal publications. However, he admitted that this focus on non-traditional outputs caused some problems when the work he and his colleagues were doing was compared to other projects, and he sensed that it might result in a worse performance in exercises such as the UK's research assessments (RAE and REF).

Collaboration

The nuclear physicists we interviewed were all engaging in collaborative science, but the size of the collaborations ranged from those working in smaller groups of 10 to 30 scientists and technicians, to larger collaborations of 100 or more sharing beam time. In general, respondents suggested that the size, cost, and rarity of equipment played an important role in dictating the size of the collaboration. Generally, respondents pointed out that nuclear physicists engaged in smaller collaborations than particle physicists, and this was at least partly because "the particle physics kit is more expensive, so there are fewer facilities" and scientists have to share with large groups if they want to do their scientific experiments.

The length of collaboration can vary widely. One pure nuclear physicist working on the structure of nuclei reported working with 10-15 collaborators at a time, but said that the membership of collaborations was fluid and changing:

[The group's interactions] wouldn't be anywhere close to daily. It would maybe be monthly. So the collaboration's ad hoc in the sense that the collaboration is formed around a particular measurement that we want to make, so they're not

long standing, typically. They will be formed, and they will achieve their scientific objectives and then move on. We would, close to the time around the measurements, talk to each other via email on a regular basis, but then that contact would be at a lower level.

But many of the other nuclear physicists we spoke to followed a model of long-term and long-standing collaborations. One suggested that his work was “not like some of the other sciences, where people might collaborate with one or two people who are interested in a specific topic...and then you move on to something else; our collaborations are large and they are long term.” Another argued, “we work in a field where the experiments take many years, and there’s a history, and a sort of incremental progression, and so we have established these collaborations a long time ago.” Part of the reason for this collaboration is the complexity of the equipment: “the experimental work is always collaborative, it has to be, because we need many people to get the equipment to work.” The UK nuclear physics community, in particular, must collaborate internationally, since there have been no major research facilities in the UK since 1993. Several participants said that this has shaped how UK scientists get involved in collaborations: they send people and equipment to collaborative facilities elsewhere that can extend the

facility’s capabilities to do things that otherwise would not have been done. One suggested that the decisions to collaborate internationally has nothing to do with a desire to be involved in large collaborations particularly, but is simply a matter of practicality: “we all want to do research and there’s limited access to facilities.” Some participants suggested that collaborations have been getting larger in the past decade as the research equipment has been getting more complex and expensive.

The tools of collaboration are the same as some of the other physical science cases, including email, telephone, EVO for large periodic teleconferences, Skype for smaller meetings held every few weeks, and face-to-face meetings once or twice a year. Face-to-face meetings remain important, even though the tools for online meetings reportedly work well, because the meetings are “the opportunity to build up the relationships and to talk about things in-depth, and to explore ideas for new work. I mean, if you have a one-hour conference and it’s fitted in to your schedule, and you’re discussing just some small aspect of things, this just doesn’t happen.”

Transformations in practice

There were few suggestions that computing had radically changed the kind of science that was being done or the ways in which researchers work. However, nuclear physics is beginning to involve larger collaborations and more complex research technologies. Several participants reported that since 2000, there has been an increased emphasis on two-stage accelerators which are intrinsically larger than the facilities they supersede.

The biggest transformation otherwise has been related to speed: “the Internet really speeds up the exchange of views and interpretation of the data. Otherwise it was by post, or fax was quite important for a while. I think that all this pressure of rapid communication is [both] good and bad ... There used to be time to think about things in between letters.” Several also mentioned that this increased pace is reflected in publication pressures, with more papers expected to be published than in the past.

New questions

As with transformations of the science, there was little evidence that computing technology itself has opened up fundamentally new research questions.

I'm not sure that it enables you to ask new questions within the field because normally nuclear physics is pushed forward experimentally, and so it's not a subject like history or social science, where I think the retrieval of pre-existing academic information is actually the main activity. So I don't think that it enables me to ask new questions, but it certainly makes things go a lot faster.

Participants also felt that technology enables them to be more thorough in their background research. By having everything online and searchable, participants felt that it was “less likely that you will miss something” when looking for previous research.

However, some of the advances in science are linked to computing capabilities: data analysis techniques such as lattice QCD, partial wave analysis, or the use of Markov Chain Monte Carlo simulations rely on computing power, and can now be applied to problems that were intractable a decade ago.

New technologies

Participants emphasised the desirability of better ways to read and annotate PDF files. For instance, one participant felt that PDFs were difficult to read because it was difficult to flip back and forth between the text and the references compared with his experience with paper. Related to this was another participant who wanted to be able to “chase the reference threads from papers, rather than just going through each paper one by one and clicking.” He wanted a tool that would allow a researcher to start from one paper or a collection of papers and be able to automatically see all the references and visualise whether there is an emerging research theme linking back to the same primary papers.

One interesting example reported during the study involves repurposing off-the-shelf technology: the use of graphics card processors (GPUs) to speed up analysis. Speed improvements of up to three orders of magnitude are possible using this technique, which is available largely because of advances in the gaming industry.

There was little evidence that the nuclear physicists participating in this case study spent much time or effort actively seeking out new information tools and strategies, although many were open to new approaches that would make their research easier and more productive. They appeared to rely more on word-of-mouth to learn about approaches that colleagues had adopted, or on serendipity such as discovering a new tool or website that allows something useful to be done and which can be adopted within their work practices relatively easily.

Chemistry

Chemistry includes a wide range of sub-fields and approaches to work. Some chemical fields have significant crossover with biology or physics, and therefore have borrowed some of their approaches to information use and dissemination. In contrast, the 'pure' chemists exhibit information use and management strategies which are quite different from the other groups in this study. Participants in this case come from the fields of inorganic and organic chemistry (synthesising chlorophyll, protein engineering with bacterial enzymes), researching MRI imaging agents, chemical biology (drug design for cancer targets), and physical and theoretical chemistry and computational physics (computer simulation of electron spin dynamics).

Chemists' information behaviours have been the subject of previous studies which noted their reliance, as a discipline, on journal literature (Davis, 2004). In Davis' study from nearly a decade ago that measured accesses to the American Chemical Society (ACS) servers, the large majority of traffic at the time (84%) was referred by SciFinder Scholar's database of chemistry abstracts. More recent work has found that chemists spent longer than other disciplines (except for physics) when viewing ScienceDirect articles, and that

chemists were more likely to view the articles rather than just the abstracts (Nicholas, Rowlands, Huntington, Jamali, & Salazar, 2010).

This case focuses on a single department of chemistry: we interviewed six chemists at the University of Oxford representing mostly younger scientists—one senior researcher, one postdoctoral fellow, and four current PhD students. Additionally, we conducted a focus group with seven chemistry students, three of whom had not already been interviewed. During the focus group and interviews with students, we explored the students' perspective on information use in their discipline, focusing on their acquisition of the disciplinary culture, and which aspects of it they accepted without question or were inclined to challenge.

While a single department cannot show the range of activities across a discipline as large as chemistry, it does provide a comparison to the other cases in this study which were not made up of co-located scientists. The University of Oxford chemistry department is one of the largest and most prestigious in the UK: it was one of the top research departments in the 2008 Research Assessment Exercise,

with an average ranking just behind Cambridge and Nottingham, and with the largest number of staff submitted for assessment of any department in the U.K.

Information retrieval

The main tools used by chemists for literature searches are Web of Science and SciFinder. Any variation from this is field-dependent—for example the biomedical and drug discovery chemists use PubMed extensively, while the magnetic resonance researchers, who have some crossover with physics, also use pre-print resources from arXiv.

Most of the chemistry case study participants work in laboratories, performing experiments to test their hypotheses. The computational chemists, as their name suggests, work entirely with computer models. As well as the search tools named above, several participants mentioned that they will seek background information from Google, Wikipedia and Google Scholar if, for instance, they are doing some cross-disciplinary work and need to go back to basic concepts.

Library use is decreasing. The more advanced doctoral students and postdoctoral researchers previously used the library for access to some foreign-language journals, but this now happens rarely, as most of the resources they need are digitised. A fourth-year undergraduate in chemistry had

never needed to access a journal within a library: “I never use the old-school journals on the shelf, or anything like that ... I’ve heard of people going to the library and looking stuff up in the old journals, but I’ve never had to do it myself.”

Information management

Most of the chemistry students we interviewed print out papers and store them in an office outside the laboratory, and annotate them with highlighter pens. They import bibliographic details into bibliographic software such as Endnote or BibTeX (for LaTeX users) when they download the paper. A minority of those interviewed read papers onscreen and annotate them using PDF readers. One reported using the online file management tool Dropbox to store papers: “because for some reason my computer doesn’t connect to the server, it allows me to transfer files between different servers.”

Data analysis

Experimental data is created in several forms. Machines in the laboratory produce files which can be read and manipulated in the proprietary software for that machine, or exported to Excel. Most researchers we spoke to have a series of Excel spreadsheets for data manipulation which are kept on a networked drive. A core part of the computational chemists’ work is producing software tools

such as the Spinach, an open-source function library written in MATLAB.

Production of graphs and figures for publication is also common. For this, researchers use a variety of off-the-shelf software, ranging from image manipulation tools like Adobe Illustrator, to more specialist tools that produce molecular models, such as HyperChem and Chimera.

Citation practices

All those interviewed reported following the citation chain of papers they find in searches, and several remarked that they found SciFinder particularly useful for this: “it allows you to ... check what’s been cited by or who it’s citing, and so that’s quite good, because you can follow the trail of papers once you’ve found that one key paper.” None of the chemists interviewed cited databases or other online resources they use, instead mainly citing published journal articles.

Teaching and learning

We explored how researchers in the study had acquired the practices outlined in this case. The students were unanimous in agreeing that most of their information practices were taught to them by slightly more senior researchers—that is, senior doctoral students or early career postdoctoral researchers.

The formal training in research information practices which they received as undergraduates was not perceived to be particularly useful, since it was quite short and targeted at tools they found they did not really use once they embarked on their research careers. The training “teaches you every other web-based search mechanism that you could think of, apart from SciFinder, which is the one that everyone uses on a day-to-day basis.” Some students marvelled at their supervisors’ ability to stay on top of the literature and to suggest new techniques or approaches based on their readings, but remarked that “when it comes to anything more specialist or analytical software, things like that, he’ll come and ask one of us to do it with him.” However, several students also remarked that they received most of their training in specialist software from their supervisors.

Dissemination practices

Chemistry students displayed a high degree of trust in peer-reviewed publications, and they and their groups largely did not use preprints, as: “ultimately the source will not be trusted or referenced unless it’s been published in an established peer-reviewed journal.” Conferences are less important than publication for dissemination and, where there is biomedical crossover, there are also fears of being scooped. Researchers always attempted to have a paper submitted or accepted before even talking about it

at a conference: “There’s too much risk for people to be able to copy the work—and that has happened. Not to me, but to others, where they’ve presented at a conference and someone else has repeated it and published it.” The perception of risk, even if there is little evidence to suggest academic theft in actual practice, shapes the behaviour of scientists in this case.

In practice, multiple teams can easily be working on parallel research projects aimed at similar ends without any suggestion of theft or dishonesty on the part of any of the participants, since the leading edge of research in science is cumulative, and thus all are relying on the same body of previous research that suggests particular directions for new investigations. As a result of this culture, the chemists’ publication cycle is short, usually weeks, not months, to avoid the scenario of even accidentally being scooped: “It could happen that, you know, you’re four weeks away from getting a review back, and someone’s just published what you’ve submitted.” Most laboratories will link to papers on journal sites, but will not host the PDFs on their own servers for download. Where there is crossover with physics, as with the magnetic resonance researchers, arXiv preprints are used. Publishing in gold open-access journals is not common, and one participant remarked that their group was unimpressed with the slow turnaround of reviews when they submitted to an open-access journal.

Collaboration

Most of the chemistry students’ research groups were involved in collaborations with other groups. However, these differed from the kinds of collaboration seen between physicists in several ways. First, the collaborations are much smaller. One participant cited twelve as a large number of collaborators. Second, the collaborations are normally undertaken to extend the capabilities of the laboratory—that is, they set up collaborations with groups who have equipment they do not have, or in the case of the theoretical chemists, with experimental groups who can test their theories. The collaborations are often set up through strong social ties, such as those between former supervisors and students now in different institutions. In the case of the magnetic resonance researchers, collaborative work is also done with other groups within the university, who are located close to the chemistry department.

Several technologies were mentioned in relation to data and knowledge sharing within collaborations. In the biomedical chemistry collaboration, data sharing is facilitated by online services like Huddle or SharePoint, or networked Oracle databases for raw data. Most communications are managed via email and face-to-face meetings.

Transformations in practice

Several transformations in practice were identified by participants. They felt that the digitisation of scholarly publications should eliminate unnecessary reproduction of research, because all extant research should be easily discoverable: “You should never, ever, make something that’s already been made. Which presumably in the old [pre-digital] days they did quite a lot, and then only realised when they got to the stage of some publication of it.”

A senior researcher remarked that for him, the transition of everyday computing into the research arena has had a marked impact on work practices:

When EndNote, the bibliographic database manager, started connecting directly to the various databases and pulling out references as opposed to you having to type them in, that was a big difference which made a change. Then when mind-mapping came around, I could suddenly offload large amounts of things I have to remember onto a trusted source somewhere on paper, or on disk if it’s properly backed up. That has relieved my memory of a considerable amount of what is largely irrelevant scheduling information. So there have been these significant paradigm shifts in the ways that I was doing research. My adoption of Microsoft Outlook ... [has] ordered up my life quite a lot, scheduling and so

on. And the switch from lesser drawing tools to Adobe Illustrator has made a sea-change in the quality of graphics that I have in the papers, so yes, there are these abrupt switches perhaps throughout, precipitated by new resources becoming available.

These changes support the notion that many of the transformations in practice have more to do with speed and ease of access to information, rather than evidence that disciplines are engaging in completely new research information behaviours.

New questions

All participants agreed that the questions they ask are mostly unchanged, but that the speed at which they can be answered is much faster: “Just because of the fact that we have a search engine we can put in the structure, and find that, and find that structure, wherever it comes in the literature. That probably saves, for every structure, an hour’s worth of work. And that’s—maybe you do that twenty times a day,” and “[in the past] we didn’t have SciFinder, so we had to go through the physical copy, which used to be volumes, and volumes, and we used to fill up a shelf of that size every year. And it used to be, you know, an afternoon worth of work to find out if you should be doing it. Which you can do in five minutes now.” Thus, participants assert that the scope of research has changed, but not necessarily the research questions.

New technologies

One participant felt that the work they do is in some ways held back by the limits of current technology, because they see where they want to go but are not yet able to reach it: “Our group is getting to the point where we’re needing new hardware, and therefore new software, to be able to advance our research because at the moment the stuff that we use is just not sufficient to kind of get the separation we need. So, that is a perfect example that our chemistry is basically pushing our requirements of software and hardware.”

The group’s wish lists almost all revolved around easier retrieval and storage of PDFs as well as easier retrieval of bibliographic data. They seem happy with search (and confident that they are finding everything they need) but find bibliographic management a chore. One felt that Zotero, a Firefox browser plugin, solves this problem: “If you use Zotero ... when you add the DOI to that database it saves when you last accessed it, and will keep an up-to-date database. It’s very, very good.”

During the focus group, one participant also expressed a wish, supported by the rest of the group, for centralised resource management of laboratory materials, to reduce waste and time spent ordering or looking for solutions in the laboratory.

Earth science

Our fifth case examines information practices in the field of earth sciences. Since the field broadly encompasses the study of geologic history, natural hazards, resource availability, and climate change, we were able to explore diverse research practices. While sharing a focus on geologic issues and phenomena, researchers varied dramatically in their collaboration practices, data collection methods, and means of dissemination. Many of their differences related to either the timeliness of their research or the need for shared facilities. Volcanologists, for example, tended toward more rapid dissemination routes than those modelling climate change. Likewise, researchers who shared facilities, such as hydrologists and seismologists, engaged in larger, normally international, collaborations compared with those working in smaller labs.

We interviewed six earth scientists and hosted a focus group discussion for four graduate students at the University of Bristol. Additionally, one geophysicist participated in our interdisciplinary focus group discussion held at the British Library.

Information retrieval

Unlike physicists and astrophysicists, earth scientists do not make extensive use of pre-print archives. As a result, they lack a single point for finding new information, so the majority rely upon their peers and citation chaining. Most interviewees emphasised the importance of personal contacts to learn of new research. As one scientist said, “many things are discovered by talking to people; somebody else will have discovered something, and will tell you, or someone will have heard something at a conference.” Another described the process of making connections: “In time, you just build up knowledge...you know people, and you know, in your field, who is working on what.” Participants learned of new research in a variety of ways, including participation in projects and collaborations, at conferences, and through conversations within departments.

Conferences were frequently mentioned as an important place to make connections and learn of new research. Indeed, conferences and journals were often mentioned together: “Some of the best ways to learn about new research, new techniques, and moving things forward is through journals and academic presentations.” Sub-fields

within earth sciences have few enough top-ranked journals that most scientists can easily keep abreast of new work. One scientist described having a starting point of about ten resources that he regularly checks. A few receive email alerts for new publications, while most browse journal databases or regularly check the websites of key journals.

A few scientists expressed dissatisfaction with their current practices of information sharing and retrieval. One described his method as “not terribly systematic” and another mentioned the limitations of depending upon colleagues to send emails when they publish new work. The latter felt that his methods of discovering new research were somewhat haphazard and described a recent experience of missing an email about important new work: “I think it’s mostly through speaking to people directly that I would find out—sort of sideways leaps, if that makes sense.”

All participants used Google Scholar or Web of Science, but for different purposes. Some browsed Web of Science to look at specific journals or to find an article they had heard about, while others used it as a search engine to “see what comes up.” Web of Science was described as easy to use and a complete resource: “At a click of a button I can download any paper that I’m interested in.” Some use Google Scholar and Web of Science interchangeably, listing both as their starting points when beginning research. These sites are

used daily when writing and less frequently when teaching. Many described Google Scholar as more useful than Google because results are limited to journal articles. One scientist described Google Scholar as enabling him to do keyword searches without the deeper context required when searching journals using platforms provided by publishers and institutions.

Earth scientists used Google to stay up-to-date with world events, for example, to track statements that governments released pertaining to natural hazards. Along with Wikipedia, Google was used to gain basic knowledge of new topics. Graduate students in particular used these resources to develop familiarity with new aspects of their field. Additionally, a few scientists mentioned using Google to search specific databases, even where the database offers its own search function. For example, a volcanologist described using Google to search the Smithsonian Institute’s Global Volcanism Program. Thus Google serves several purposes, from staying abreast of current events, to gaining familiarity with a new topic, to offering a stronger search tool for existing databases.

Information management

Although no clear file management strategy emerged from our interviews, participants seemed generally satisfied with their use or non-use of citation management tools. A few

used citation management tools such as Papers or Zotero to organise articles they downloaded, yet even those who use these tools retrieved the articles online when re-using, rather than searching their print or desktop files. An earth scientist said, “I don’t even go to the PDF any more. That’s what Google Scholar’s good for or Web of Science. You type in the article and it just goes straight to it.” Most participants only printed articles they found particularly important and annotated printed copies with notes and highlights.

Data collection

Methods of data collection are also dependent upon research area, and may include experimentation, fieldwork, using satellite imagery, or a mix of all these. For example, a scientist engaged in climate modelling gathers raw data, including satellite imagery, from sources such as the National Snow and Ice Data Center in the US, the European Centre for Medium Weather Forecasts (ECMWF), and the UK’s Met Office, all of which are online databases providing data to registered users. A seismologist reported using remote sensing data requested through NASA and ESA collected by satellites and ground-based facilities. A computational mineralogist used computer simulation and neutron scattering in a lab setting. Those using satellite imagery described using multiple sources for their research, explaining that by combining resources they can usually

achieve acceptable, if not complete, coverage of a given area. One seismologist described using different data from different sources as straightforward: “I’ve never had any difficulties with getting hold of data.”

Data analysis

Across research areas, earth scientists reported that skills in computer programming are essential to their work in order to prepare data for processing, process the data, develop graphs, and perform statistical analysis. The scientists and graduate students we spoke with were technically competent, with skills ranging from simple spreadsheets to multiple computer languages, and expertise with specialised tools such as ArcGIS, a mapping and visualisation tool. Most scientists learn new tools when they are needed for a project, citing this flexibility as important to collaborations: “When I’ve collaborated...they might have sent me some code, so I needed to be able to manipulate it and use it.” Commercially-available software is somewhat more commonly used than bespoke tools, however, since both are widely used, there is broad support in the field when it comes to trouble-shooting.

However, one climate scientist cautioned that, despite widespread availability of raw data and tools to manipulate them, people may not know what to do with the data: “If there are no people who are experts on how to interpret the data, you end up having some numbers, and you can get the wrong conclusions if you don’t know exactly what you’re looking at.” This reliance on other scientists is acute in earth science, because it would be very difficult for a single scientist to have the diverse range of skills needed to work with data from different sources which requires different approaches to analysis.

Citation practices

As with the astrophysics case, earth scientists in our study used the acknowledgements and methods sections of their papers to mention software and facilities used in their research. Satellite data that is available to registered users, for example, from the Japan Aerospace Exploration Agency (JAXA) or European Space Agency (ESA) are usually mentioned in the methods section, and data that is more challenging for the agencies to collect receives an acknowledgement; however, imagery that is easily discoverable through a Google search and freely available is generally not cited because scientists assume that anyone can find the images themselves.

Dissemination practices

Peer-reviewed journals and conference presentations are the main way for earth scientists to disseminate information. Conferences and project meetings were seen as crucial, and equivalent to journals in conveying new findings. Other dissemination practices differ depending upon the urgency of the data. For example, scientists specialising in volcanology mentioned posting information about eruptions to Twitter. During our focus group discussion, students recalled a lengthy Twitter debate about the temperature of lava flow. In comparison, students did not mention using Facebook groups for communicating research; they used it mostly to post photos of fieldwork.

Web 2.0 technologies were viewed as promotional tools rather than as a means to disseminate research findings. Two of our respondents were suspicious of blogs, with one saying “I’m more interested in actually having a career in science, rather than just getting publicity for myself.” Indeed, while some scientists reported getting information from blogs, none were active bloggers, expressing concerns that blogging would distract them from their research. Students and early career researchers felt they should focus their efforts on publishing in highly ranked peer-reviewed journals rather than on blogging.

Even though blogging was seen as a potential distraction, self-marketing and promotion are emerging as a new concern for all kinds of researchers, particularly as funding bodies are increasingly asking individuals and organisations to demonstrate their impact beyond their academic influence. Scientists are learning that they need to engage more effectively with the public. Graduate students are realising they need to establish an online presence. Heads of research groups describe spending a portion of their time on public engagement. Across all positions, scientists recognise that they need to raise awareness: “Social media networking, that’s definitely kind of revolutionising the way, especially we reach out to young people.” Some say that these efforts are not changing the way the actual science is done, and that the “nitty-gritty of how the science is transmitted is still going to meetings, meeting people.” Others feel there is growing awareness of a need to engage the public, describing that now “every research statement has its own webpage.”

Departmental websites act as digital business cards, providing contact information and links to recent publications. A few described their departmental websites as a crucial means of disseminating information because they are usually listed first in Google results, so keeping these pages up-to-date was a priority.

Collaboration

Collaboration sizes varied significantly depending upon the research. Those sharing large facilities engaged in large international collaboration: “Economies of scale lead to collaborations on field projects between universities, because it’s cheaper to have a big camp than a small camp.” Those using freely-available data and performing analyses in their home institution generally collaborate with others in their department, or form small external collaborations composed of 4-5 other researchers.

Most communication occurs face-to-face, by phone, or by email. Some participants preferred face-to-face communications, while others stressed the convenience of email for collaboration on large projects. A majority said their collaborations last for several years and tend to form naturally over time, with the same researchers moving within a few projects. Those engaging in large collaborations described email distribution lists as “absolutely critical” for rapid dissemination of natural phenomena. One earth scientist described discussion on the distribution list as “much more informed than anything you’d see online, because it’s such a specialist group.”

Document sharing occurs primarily via email, with a PDF, Microsoft Word document, or LaTeX file sent as an attachment. Graduate students used Dropbox and FTP servers for larger files. Many participants respond to drafts with a list of comments, or, if working in close proximity, will print the draft, write comments, and hand it to their collaborator.

Transformations in practice

Technological innovations over the past twenty years have had a remarkable impact on research practices in earth science. For seismologists and volcanologists, Global Positioning Systems (GPS) allow increased accuracy in reporting the timing and location of natural events. Centralised, connected data repositories allow researchers improved access to satellite and ground-based images; they can compare such images across datasets to develop and validate models for climate change or resource sustainability. Archived satellite data allows groups such as the European Centre for Medium-Range Weather Forecasts (ECMWF) to provide historical records of weather patterns. Improved data storage allows free access to large datasets and enables sharing of large datasets across research teams.

Improved communications options, such as email and telecommunication, support large collaborations and enable geographically-dispersed groups to work together in real time. Further, improved communications make it easier to move large datasets between research groups.

Most participants used journal collections online, and never visited the physical library. Electronic journal articles speed the process of citation chaining by providing links to bibliographic references, links to who else is citing, and in some cases, downloadable data.

Teaching, too, has changed dramatically. Whereas earth scientists used to teach from a single textbook, “now everything’s done in sort of the PowerPoint or Keynote and you sort of harvest. You know, you get on Google and you get exciting pictures of things and you look at other people’s lectures.” As lectures, slides and images are posted online, academics borrow and learn from the best of their colleagues’ work to create their own teaching materials.

Despite rapid technological progress over the past 20 years, most scientists complained about slow processing speed and memory limitations when managing large datasets. A graduate student said that he’s “at the limit of the memory” available at his university. One scientist believed that journals avoid storing large data files because they will increasingly be pressed to store larger and larger files.

A few scientists expressed dissatisfaction with the current journal system, which differed from the high levels of satisfaction reported by scientists in fields and disciplines that make heavy use of arXiv, such as astrophysics and particle physics. A postdoctoral researcher complained about the high cost of publishing articles in journals, questioning why the person submitting the article has to cover costs of printing colour images, when most readers access and prefer the digital copy. His perspective was that “there’s no difference in the cost of making a PDF in

colour versus in black and white.” He further described costs associated with publishing in top journals as prohibitive for researchers starting out, saying that sometimes projects aim for lower-level journals because of lower publication costs.

Participants expressed feelings of overload both in terms of information and overall workload. Students in particular experienced difficulty when first gathering information about a new field. Others said that they downloaded more PDFs than they can read. In terms of workload, one senior academic said, “I get involved in too many projects, and I don’t have time to actually work on everything.”

New questions

The focus group participants considered how the field of earth science would look without the Internet. One student remarked that new research would take longer to disseminate, saying that it would be impossible for a student to access an unpublished article. They initially focused on the practical difficulties of writing letters and posting data on disk to their collaborators, but then asked how they would locate people doing relevant research in the first place. A first-year graduate student said, “with no Internet I don’t think I’d have any data.” She is studying historic records that have been collaboratively collected and maintained, a practice enabled by the Internet.

One seismologist said, “the nature of what I’m doing hasn’t changed that much, but the quantity of data and the quantity of the results we’re producing has gone up by orders of magnitude.” Echoing findings from the Gamma Ray Burst and Zooniverse communities, this seismologist said that while 20 years ago he would publish a paper about 20 earthquakes, now he can publish one about 10,000. Comparing trends across a larger dataset strengthens scientists’ capacity to test theories and develop models. Seismologists can also now spot new phenomena that were not evident when viewing small numbers of seismograms:

“When you take all those data together and you sum the signals up – what we call stacking – now we can see these really subtle signals. And these subtle signals are very important because they tell us – they really allow us to much more accurately constrain the material properties of the earth’s deep interior.” Another seismologist said that new technologies enable researchers to view the earth’s deep interior at “a resolution that’s completely unprecedented. In fact, 20 years ago people have said it would never be possible, so it’s a very exciting time.”

New technologies

Recommendations for improvements to information resources focused around centrality and increased access. One geophysicist wanted a central resource similar to Wikipedia with basic geologic information for scientists. Another described his vision of a system that allowed for easy sharing of data and methods of manipulation and analysis that would capture their analytical actions as they occurred and could be centrally stored for easy access. A climate scientist wanted a facility to put out unfinished work for someone else to pick up and continue. Building upon the idea of improved transparency and access, a seismologist argued that journals should be completely free because “science tends to be funded by public money, and it’s right that anyone can read the results of that, and it’s sort of holding back—I think it holds back science, if anything, by charging for access to journals.” Of course, this can exacerbate the tension mentioned above, since journals that charge for publication may charge even higher author fees to provide open-access to articles.

Nanoscience

Nano-scale science and technology have seen rapid growth in recent years. Nanotechnology is defined as science, engineering and technology related to the understanding, control and use of matter at dimensions of roughly 1-100 nanometres where unique characteristics enable novel applications (Shiri, 2011). Scientists from many domains, including biology, chemistry, electrical engineering and electronics, material science, medicine and physics, are engaged in research in this emerging field. Nanotechnology is already employed in commercial products and promises significant breakthroughs in areas such as medicine.

Nanoscience was selected for this study because it is such a new and multi-disciplinary field. Nanoscience scholars were invited via email to participate in our case study. We interviewed researchers from a variety of backgrounds such as condensed matter physics and electrical engineering. Five physicists and two electrical engineers – two senior academics, three PhD students and two postdoctoral researchers – participated in the study.

Information retrieval

Participants generally began their data collection with a web search on Google or Google Scholar. As one stated, Google is the “first port of call because it still gives you what you are looking for, but it is easier and faster” than the alternatives. Google is used to find academic papers, but also to find broad information from sources such as Wikipedia, newsletters, and other research groups’ web pages. More rarely they might access Web of Science for scientific paper searches but, as one interviewee joked, “if it’s not on Google, then it does not exist.” A physicist argued: “I think it seems to be the people of a slightly older generation grew up using Web of Science, and people slightly younger grew up using Google Scholar.” One participant with a background in electrical engineering also reported occasionally using ScienceDirect, and/or IEEE Xplore for scientific searches.

While the nanoscientists are aware of other information resources provided by their universities, for example the online library catalogue, these are not used because they are considered inflexible and constrained. Journals’ own websites are rarely accessed directly for search purposes, with only one participant indicating daily access to

particular journals. ArXiv is not widely used as participants do not believe it provides publications relevant to their field.

Traditional methods such as citation chaining remain an important way to find information, with a few participants using Web of Science for tracking such citations. Most also agreed that there are a number of books that explain the fundamental principles of the field, such as the Handbook of Chemistry and Physics, which they have to revisit occasionally. Most participants also mentioned talking to people at conferences as an important information resource.

Our nanoscientists also subscribe to mailing lists and check other research groups' websites to see their latest publications or work. Most used social tools, such as Wikipedia, blogs, Twitter, Google Books and online lectures, to gather information when they started working on a new area. Patent searches are considered important in nanoscience, since research in this field often has commercial value which needs to be protected.

Information management

Organisational strategies vary among nanoscientists, reflecting their different disciplinary backgrounds. The physicists used LaTeX for word processing, while one electrical engineer preferred to use Microsoft Word. Participants often read the abstract or skim-read the paper

online before deciding whether to save it to their personal computers. One respondent used Zotero to manage his information sources.

When a document is a key resource or difficult to read on screen (e.g. it is lengthy or it is in a hard-to-read font), most interviewees print it out, read it on paper and make annotations, such as highlighting key pieces of text or writing notes in the margins on the hard-copy. Some also annotated texts by copying and pasting relevant passages, or writing useful notes in a digital text file that they later incorporate into their article drafts.

Data analysis

Most participants felt that, when faced with a specific problem, about 30% of their time went towards finding information, while about 70% went towards analysis and problem-solving. But they also stressed that in their field, most of the time, these tasks are performed simultaneously.

Analysis in nanoscience is cumulative, involving a number of steps and sometimes requiring collaboration between people from different fields and disciplinary backgrounds. For the physicists, analysis is mostly about programming as they must write code to produce data, to analyse data, to create plots, and to undertake other related tasks. The information that they have gathered is used to situate their

work within current practice and to inform their coding and the techniques applied. A physics PhD student said: "I normally have MATLAB open and just test if certain things work, I try and test as I go along, as I read it and make sure I understand what's happening." MATLAB was the most common programming language used, but others that were mentioned include LabView, Perl, Fortran, C and C++.

In contrast, the analysis done by electrical engineers involved developing experiments, analysing experimental data and comparing and contrasting these with different theoretical models, rather than writing code. Information that they have acquired is mostly used to help them address challenges when setting up the experiment or when something goes wrong. The participant working in an industry setting had a slightly different focus, with analysis involving finding the latest papers or patents on the technology they aim to develop and on writing software to control the specific equipment they develop.

Citation practices

For most participants, citation practices depended upon the nature of the resource. Most did not cite online resources, such as Wikipedia or websites, when they used them. They also tended not to cite general concepts from major papers, as these are assumed knowledge. Most did not trust

information which is not peer-reviewed, such as pre-print publications, and thus citations tend to come from refereed, well-known journals or books. As one stated: “The vast majority of the stuff that we cite – you know, 99%, will be papers in the primary literature, which have a journal, have a volume number, have a page number, have a year, etc.” BibTeX was the most popular citation tool for physicists, with one student saying, “I’m a very hands-on person, I like computers, I like to do everything myself. The thing I like about BibTeX is the fact that all the source code is there, all of the ways it formats it are in files [and thus] I can go in and hack the formatting file until it does it exactly as I want. So, I can’t get on with programmes like EndNote where it’s all in a black box, and their code, and if it doesn’t export how you want, well, tough. So, I refuse to use Microsoft Word, or EndNote because it just doesn’t do exactly what I want it to do.” An electrical engineer interviewee, however, preferred EndNote.

Dissemination practices

Traditional dissemination channels such as peer-reviewed journals and conference presentations are the primary dissemination route for all participants. Inter-disciplinary conferences are particularly valued as opportunities to meet people in the field, attract investors, and create collaboration networks. One interviewee also published

in general magazines, such as New Scientist, for the wider public. Most argued that they would not publish their work in electronic journals that are not peer-reviewed. As one stated: “I don’t use non-peer-reviewed journals and I would never publish in a non-peer-reviewed journal. There’s no quality control and therefore there’s to some extent an issue of credibility.” Publications are not submitted to online repositories (such as the condensed matter archive) before acceptance in a peer-reviewed journal because, as one interviewee said: “I have particular qualms about uploading papers to the archive before peer review because in many cases peer review acts to improve a paper – sometimes considerably. Uploading a paper prior to peer review then becomes difficult because two (or more) different versions of the paper are publicly available.”

In line with other research findings, academic publications are not a major concern for industry. The industrial partner that we interviewed said: “We have to protect our intellectual property so publishing is not a good thing anyway.” He further argued that “to justify our research grants as a company we have to be seen to be active, so, conferences, giving presentations to schools, universities – less specific tasks, more sort of general information so that the public is able to perceive the benefit – is sort of more important to us.”

Dissemination also usually occurs through more informal means. Conversations with colleagues – face-to-face, online or via email – allow researchers to discuss technologies and techniques, often establishing or challenging standards in the field. Most participants do not use social media to disseminate their findings, but one senior physicist maintained a blog to discuss general research findings, while another contributed to other groups’ blogs by posting comments. For those we interviewed, not using Twitter or Facebook or blogging is a well-considered decision. As one nanoscientist described:

The problem with things like that is they’re quite high maintenance. I mean, the point of a blog is that you want people to read it because you think you’re important enough that people should listen to what you have to say. And the point is that you have to want to publish, essentially, to want to write, and I think there’s a certain kind of personality type that, you know, feel that what they have to say is important enough that they’re going to put in a blog for people to read.

Most indicated that even within larger collaborations, people usually ignore the wiki-type interactions and prefer communicating via email. However, nanoscientists who do blog see a significant value in the practice: “The blog type of thing and getting involved with posting comments where you

get some interaction there, that can lead you in directions you wouldn't have had before. And it's that interactive element that is pretty important as well."

Most respondents also post versions of their papers on their personal websites, or their university's website. Interviewees from the University of Nottingham used YouTube in an innovative project which informs the wider public about physics and nanoscience research. One said: "We are pretty keen on outreach and public engagement, so the school actually participates in something called Sixty Symbols, which is a project we set up together which is a set of YouTube videos. Not all of which are linked directly to research, but many of them are. And they've picked up something like 3½ million hits altogether, now, so that's a good way of disseminating, but not only disseminating to the research community, but also disseminating to the wider community. And they're targeted largely at people who don't have a large background in science. So what we try and do is take bits of physics and try to put them in terms that are comprehensible to somebody who has never studied science."

Collaboration

All participants engaged in collaborations, although the nature of those collaborations depended on the type of the

work to be performed. For example, one interviewee only collaborated with colleagues at his own university, as they need "to claim total control of the technology we build, of our findings."

Participants described nanoscience as a field which requires collaborative efforts, between different scientific fields and between industry and academia. One participant said: "I can't think of ever doing any individual work. We only collaborate, this is the nature of nanoscience." A senior physicist said:

When you're doing nanoscience you're sort of working at the convergence of the traditional physics, chemistry, biology and computer science – I think you'll find that most people in nanoscience you talk to will have quite a wide range of interdisciplinary collaborations ... I particularly have collaborated with material scientists, with chemists, with computer scientists, with life scientists, people in biotechnology and bionanotechnology, etc.

Collaborators communicate via frequent email exchanges, subscriptions to shared mailing lists, uploading relevant material to websites, phone conversations, video and audio conferences using Skype and face-to-face meetings at least twice a year. In some cases, there are frequent visits to collaboration sites, where co-located collaborative work is

performed for a couple of weeks. One participant reported using Dropbox to share files with his collaborators as well as a co-ordinated version control system for shared code contributions.

Typically, collaborations are not very large. The intensity of the collaborative work changes depending on the kind of work and the problems faced and there is usually a clear division of labour with specific roles and a strong hierarchy. One interviewee described "astrophysicists, particle physicists, and condensed matter physicists as very different species," something which he thought is reflected in their culture and in the way they collaborate. When co-authoring publications in nanoscience, the first author is usually the student or the postdoctoral researcher that has done the majority of the work and the last author is always the group leader. The paper always states who contributed what, and all interviewees felt that this is a fair way of reflecting the input of each collaborator.

Participants identify collaborators through traditional means. Most indicated that personal contact is important, as is reputation. One participant described this as "a very sort of organic, multi-faceted approach. You might see them speaking at a conference or they might publish a paper which you think is particularly good."

Transformations in practice

Perceptions of transformation in practice varied depending upon whether those we interviewed began their careers in the age of the World Wide Web. However, all participants very rarely visited the library and consulted books even less frequently. One interviewee argued that nowadays it is “more about rapid access, being able to keep on top of things, being able to keep up-to-date” and this is where the Internet helps. Another described his current practice as “using the Internet like walking to the library. In the old days, I would do a literature search in the library and go to the shelves and pick up the books. Now I do the same thing but on the Web. So it is not that my work has changed, it has just been made easier.”

Increased access to online content - journal publications, blogs, Wikipedia, websites, and so forth - has helped scholars begin to answer their research questions more quickly. Some researchers had previously used scientific paper searches, but their “struggle to get results out of that” led them to Google and/or Google Scholar, which they consider the perfect tools for finding information. One interviewee used Google images as a way to access the “thousands of images that are inside individual papers and texts.” A couple of interviewees used Google Books

to preview texts before deciding whether to seek a printed copy, either from the library or by personal purchase. The industrial partner that we interviewed pointed to the significance of sites such as eBay – in their field – in providing the most up-to-date information on different experimental equipment and guaranteeing the best available prices.

New technologies have increased the ease and speed of access to information, and enabled geographically-dispersed research collaborations. Information sharing was more challenging before the Web and web-based tools such as Skype allowed digital file sharing and video conferencing. Many interviewees stated that their work had always been possible, but would have involved significantly more time and travel. One interviewee said: “Nanotechnology is very popular, but if there was no rapid communication on the Internet then it would grow very slowly, because nobody would know what anybody else is doing, or what is interesting. So, I think nanotechnology just happened to be around at the time when the Internet was really coming into its own. And you know, the Internet makes the world a much smaller place.”

New questions

Most participants agreed that the breadth of information available ultimately makes their research questions more open-ended rather than prompting new questions. One participant described: “Back then, it was more targeted research. I’ve already had the background knowledge for a lot of the stuff, so I was really focusing in and trying to find the latest publications, the very latest research that someone has done. Whereas now, maybe 50% of the time I’m looking at things from a broader scope.”

But some participants believed that by enabling more complex analysis, new technologies allow for new questions. The scale of analysis has expanded, although most agree that there is still an urgent need for new technologies and tools for analysis and experimentation which will enable them to ask different questions.

New technologies

Participants felt that current technologies had a number of limitations and looked forward to progress on practical matters such as the development of tools to manage the bibliographic information in one’s own computer. A physicist nanoscientist wanted a better tool for organising literature and biographies: “If there was something that was kind of more sophisticated than Zotero, something almost like Web

of Science, but something that did that within your own bibliographic library. Something that could help you store and consolidate and visualise bibliographic information and that exported it properly to BibTeX.”

Participants also recognised gaps in online content, and wanted digitised versions of books and journal back issues. A PhD student said:

Especially, one thing I don't like about the library is that you can't – well, books don't have a CTRL+F, you can't just find something randomly in a book without trying to sit there and look through an index. If absolutely everything was on the computer, and so I didn't actually have to leave my desk, that would save me a lot of time. It's easier to search, you don't have to get up, it's easier to find the same thing, you can't lose it. A lot of journals haven't scanned all the old copies in; they're only available on book. If every single journal article in the library, and every single book in the library was scanned in and made a PDF, that would make me very happy.

A postdoctoral nanoscientist similarly argued: “Most of the journals we can get online, whereas it's frustrating – you know, sometimes there's a book that, you don't really want to get the whole book, you just want to read a few pages. Google Books for example is really good, but it's quite

limited because a large amount of books aren't on there or are unavailable for different reasons. Basically, having them as PDFs would allow you to browse books at the same speed that you're able to browse the journal archives, essentially.”

Most participants were satisfied with the tools they use to access information. One stated: “I'm pretty satisfied with what I've got at the moment ... if you ask me what in your career would make a big difference to how you do your research, access to information would fall pretty far down. Another said: “Different pieces of toolkit, more qualified students, etc., these are the type of things I need. But access to information is not something I fret about.” However, many had problems with information overload and needed better tools with “more clever searches” to overcome this (although no-one could describe what these ‘clever searches’ should be) and which could make everything more efficient. For example, one suggested Google Scholar should provide direct links to journal articles, rather than external links to a publisher's web page that can then direct you to the appropriate journal. While most felt that Google Scholar is a ‘perfect tool’, a few suggested enhancing it with a better facility for tracking citations. While Web of Science tracks citations, something important for nanoscientists, most participants complained about its keyword search tool, which is less flexible and intuitive than Google Scholar.

Moreover, many participants said that Web of Science does not include some important papers. Most participants suggested enhancing Google Scholar with the good features of Web of Science. A faithful user of Web of Science, on the other hand, recommended improving the speed at which new publications are uploaded or added to the database.

Most participants indicated that not having subscriptions to specific collections, archives or journal databases presents an often insurmountable barrier. One said: “One major concern is open access. What I find absolutely absurd is that we write the paper, we sign the copyright over to Nature, or Science, or PRL, or whatever, and then if the university doesn't have access to that journal, we have to pay for the bloody paper.” The industrial partner also said: “It annoys me when journal articles aren't free, or I don't have access to them. It's very frustrating.”

Zooniverse and citizen science

“Each of our projects always seems to start from someone saying, ‘I’ve got way too much data, and I can’t process it myself.’” Galaxy Zoo and the Zooniverse group of projects (which currently comprises eight astronomy-based projects and two based on transcribing historic documents) are created within a software framework that allows non-experts to identify features within photographs. In Galaxy Zoo and the other astronomy projects, these identifications are used to create catalogues of astronomical objects. The citizen science approach allows classification of many more objects than was previously possible, and has led to the discovery of several new astronomical objects.⁴

This case is unusual because the data-gathering stage of research involves engagement and interaction with the general public and media, including use of online communications such as blogs and Twitter. Thus, in this case we focus not only on the information practices of the scientists involved, but also the implications for public engagement with science and the emergence of new forms of data analysis in astrophysics.

Six members of the Zooniverse project – five scientists and one software developer – were interviewed for this study. To

gain some insight into the citizen scientists’ view of the project, the forum moderator of the Zooniverse was also interviewed.

How the project works

The goal of the Galaxy Zoo and similar projects is to classify galaxies from images provided by the Sloan Digital Sky Survey (SDSS). One of the main tasks for scientists working on the project is to encourage citizen scientists to create data, which they can then analyse in order to produce scientific data. The scientists’ roles are necessarily extended from the routine astrophysical procedures, as they must interact with the general public via the Zooniverse. In addition, the project has several full-time software engineers who work on developing and improving the Zooniverse cyberinfrastructure. Scientists working with the public on Zooniverse projects are known as *Zookeepers*, while the public call themselves *Zooites* (Raddick, et al., 2010).

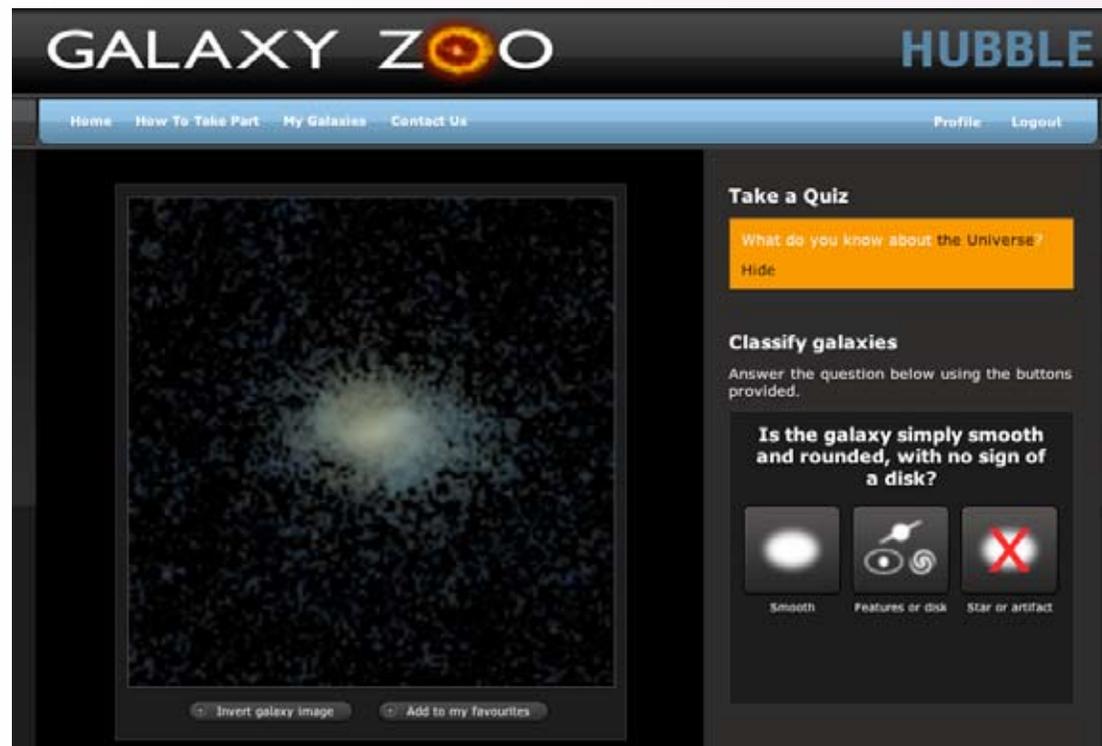
The first Galaxy Zoo project was a response to a data deluge problem experienced by astronomers. The Sloan Digital Sky Survey (SDSS) produced far more

photographs than previous projects: the initial Galaxy Zoo sample had almost 900,000 objects, but previous work with SDSS data ranged from 2,500 to about 50,000 manually classified objects (Lintott, et al., 2008). Therefore, astronomers could not inspect the entire catalogue, especially since multiple independent classifications of each galaxy are needed if researchers are to have confidence in the results (Lintott et al., 2008). Part of the survey had been professionally categorised before the Galaxy Zoo project began; this provided a baseline against which to measure the citizen science contributions.

Citizen scientist contributors register with Zooniverse and then choose which project they want to contribute to. For Galaxy Zoo, an image of a galaxy is shown in the browser, and the user clicks one of six buttons on the right of the image to classify the type of galaxy (Raddick, et al., 2010). For more complex projects, the user may be asked to identify more features or types of object, or to measure objects by selecting them.

The Zooniverse team were surprised by the number of contributions to the first Galaxy Zoo: the strong public response was aided by mainstream media publicity. The original Galaxy Zoo project launched on July 8, 2007, and was covered by the BBC on their website and a morning radio show on July 11th, followed by coverage by other news

Figure 1: Galaxy Zoo Hubble galaxy classification interface



outlets. Within one day of this coverage, nearly 1.5 million classifications had been completed by more than 35,000 volunteer classifiers (Raddick, et al., 2010, p. 3). Over the two years of the first Galaxy Zoo project, around 70 million

classifications were made by over 180,000 volunteers.⁵ The scientists working on the Galaxy Zoo project estimate that the data provided by the citizen scientists is equivalent to maintaining approximately 150 full-time classifiers.

In Galaxy Zoo, each astronomical object is classified approximately 40 times by different citizen scientists, to provide confidence in the classifications. Such confidence is reinforced by comparing the individual citizen scientists' community classifications with professionally classified objects. This comparison also allows the scientists to rate the classification quality of individual citizen scientists. These data comparisons mean that obviously-wrong answers can be quickly discarded and, as the technical lead of the Zooniverse remarked, "internet trolls get bored pretty quickly with us" because they are not able to reliably disrupt the classifications.

The scientists consistently refer to the Zooniverse projects as a collaboration between scientists and citizen scientists. Thus there are two facets to the collaborative work in the Zooniverse: first, the collaboration between the scientific institutions involved in the collaboration (Oxford University, University of Nottingham, University of Portsmouth, Yale University and Johns Hopkins University), and second, interaction and communication with the citizen scientists.

Interaction with the citizen scientists is mainly via the Zooniverse forums and the scientists' blogs on the Zooniverse website. Although the majority of citizen scientists do not use the forums, those who do are

generally very active, and the community that has grown up around the forums has been central to making several new astronomical discoveries. The Zookeepers are careful to choose new projects that allow the citizen scientists to actually contribute scientific data: "I think it's really about giving them credit, and it's not us against them, or science team versus them doing work for us. It's really sort of a very collaborative effort."

Questions of credit and acknowledgement in authored papers are still being developed and clarified as the project progresses. Generally, the scientists are the authors of the papers, and the citizen scientists who have made significant contributions to the papers are recognised in the acknowledgements. However, the scientists also consider some citizen scientists' observations worthy of an authorial credit, such as Dutch schoolteacher Hanny van Arkel's discovery of a new astronomical object, Hanny's Voorwerp (Lintott, et al., 2009).

Information retrieval

In terms of information retrieval of already-published resources, most of the scientists interviewed predominantly use the SOA/NASA Astrophysics Data System (ADS) to search for papers, as it indexes most astrophysics publications. One participant remarked that it is not too difficult to keep track of new publications, as "I think astronomy has a smaller number of journals than other sciences, so there are really only sort of half a dozen journals that people publish in... But they're all on the ADS, so I don't really worry about the journal too much ... and astronomers are very good about posting their papers on the arXiv, so that also tends to be kind of an aggregate place for stuff." One participant mentioned that they can check for new resources by subscribing to the arXiv RSS feeds on their smartphone, and that they are also alerted to relevant new publications by colleagues posting links on Twitter.

Data analysis

The astrophysicists within the Zooniverse projects acknowledge that they deal with large data sets, similar to the Gamma Ray Burst astrophysicists described elsewhere in this study. However, unlike the Gamma Ray group, they are not usually relying on existing software to analyse this data. Instead, they usually write their own code to analyse the classification data sets, occasionally borrowing useful snippets of code from colleagues or other projects. They generally use Interactive Data Language (IDL), a programming language designed to manipulate visual data as well as to create plots, and draw data from the Zooniverse projects using MySQL to query the database.

Most papers are written collaboratively using LaTeX, then sent via email as PDFs to collaborators for revision; sometimes papers are discussed over Skype. Most use BibTeX for citation management, although a couple of participants also mentioned Mendeley software, which allows users to share their PDF libraries.

Citation practices

Within astrophysics itself, citation and attribution is generally expected and freely given. One participant remarked: “If I see papers on the arXiv that I feel should cite me, and don’t, I will actually email them and tell them. And it’s helpful that people often post on the arXiv before the paper is finalised, so that can often be changed.”

Dissemination practices

Knowledge dissemination from the Zooniverse projects occurs in a number of ways, partly because it needs to communicate with different audiences. Formal publishing (journal publications, pre-prints and conference papers) disseminate findings to the astronomy community, and blogs for each Zooniverse project are used to communicate with the public in a more accessible format. Several participants also mentioned using Twitter, both to disseminate their own publications and to receive news about new papers: “We’ll tweet about papers, but other people also tweet about their papers and such. So sometimes, when I have a specific question, I will go and ask people on these networks about it, and I’ve found that has been quite useful.”

Approaches to journal paper publication in astronomy and astrophysics vary widely even within sub-fields—some scientists upload preprints of their papers, but others would not consider it:

Astro-ph [on arXiv] is normally a preprint server, but depending on where you are in the field, people have very different attitudes. [Name] next door is a theoretical cosmologist, and it’s considered rude not to put your paper on Astro-ph before you submit to the journal. You put it on Astro-ph and ask for comments, and then you submit it.

If you go through to, I don’t know, Milky Way star formation people, it’s considered incredibly forward to put your paper on Astro-ph until it’s been accepted.

Each project can take time to produce enough data for a paper—often a year from project idea to useable data (that is, a sufficiently-complete catalogue). This issue of scientific pace occurs throughout our investigations: attitudes towards speed are shaped by disciplinary practices and expectations, and these attitudes in turn shape behaviours around openness, sharing and collaboration.

Zooniverse policy means that data from the projects is generally openly available: “Some of our projects come with a limited proprietary period of six months or a year that is, I guess, to reward the efforts of the initial people. But there’s a very clear statement in the kind of teaming agreement that we have between us all, that the data belongs to the community and therefore is public.” Visitors to <http://data.galaxyzoo.org/> can download the processed Galaxy Zoo data in various formats. The project asks researchers to attribute any relevant papers, but does not provide a template for citing the raw data.

Collaboration

The scientists within the Zooniverse projects frequently use email and project email lists to communicate, but have also begun to use Skype extensively. One participant remarked that Skype was now a central part of her research communications—as a graduate student she would walk down the hallway to talk to someone, but now all of the project team are always on Skype, so she can instant-message them or call them at any time. They have had a few face-to-face meetings, but these have been infrequent (less than once a year). The group is thinking of introducing more frequent teleconference meetings to supplement the extensive email communication.

Transformations in practice

The amount of data that can be processed via the citizen science approach is changing the scale and speed at which astronomical observations can be made, much like whole-genome sequencing transformed research in genetics. Astronomy is changing from “make an observing proposal, go and look at 20 objects, spend a year working on those 20 objects” to big surveys and data, like the SDSS and beginning soon, the Square Kilometre Array.

New questions

Citizen science data has the potential to transform visual analysis of astronomical image data. The citizen science data from the Galaxy Zoo and other projects can be used to train computer classifiers—and if they can become as accurate as humans it will transform the scale at which astronomical data is produced, and enable new questions and further new discoveries. The Zookeepers are already investigating this approach—in 2010 a conference paper titled *Data Mining the Galaxy Zoo Mergers* was published, which examines the feasibility of several approaches to identifying “correlations between human-identified patterns and existing database attributes” (Baehr, Vedachalam, Borne, & Sponseller, 2010). Baehr et al. found small information gains, but also identified promising directions for further studies in this area. Success in this approach would help to resolve a possible new data deluge in astronomy: “When the next-generation telescope comes along ... and it produces 100 billion galaxy images instead of 1 million,” data mining techniques would be essential, as even the four hundred thousand Galaxy Zoo volunteers would take many years to process that amount of data.

Several new discoveries have been identified from this project, for example, the new object Hanny's Voorwerp mentioned above, and the public aiding in the discovery of a new type of galaxy called Green Peas. New Zooniverse projects are also stimulating interdisciplinary work. The recently launched Ancient Lives papyri analysis project employs a postdoctoral researcher "jointly appointed between Physics and Classics" at the University of Oxford to use similar crowd classification methods to understand ancient papyri.

Most of the other groups interviewed for this study asserted that although technology made their work much faster in terms of the amount of information they could find and process, the tools did not allow them to explore completely new areas of their fields; rather they were doing "more of the same." The Zooniverse project is similar in this regard, except that "more of the same" has exceeded their expectations, with the contributions by citizen scientists increasing their data exponentially. The project has also prompted a rediscovery of visual morphologies (ie. classification based on visually-assessed, and thus labour-intensive, criteria) in astronomical research; as one participant noted: "Because the sample sizes of galaxies had got so large, there had definitely been a shift towards ignoring visual morphologies" and instead using measurements such as colour and concentration which can be measured computationally.

New technologies

Astronomers have a long history of using computers to store and analyse their observations, but have been slower to formalise computational methods as a way of extending the field. An astronomer who is also the technical lead on the Zooniverse projects hoped to train researchers in computational methods in astronomy: "I think they [biologists] realised that there was a whole area of specialism—you know, data-intensive biology. They named it, they called it bioinformatics, and it—you know, these rare computational methods applied to their research area. And what surprises me about astronomy is that astronomers haven't done the same yet. There is a term—astroinformatics—but it's not very widely used." It will be interesting to see in future years whether, as with bioinformatics, a specialisation will develop in astroinformatics, or whether computational methods will become part of the standard methods of all astronomers, or if these techniques will remain relatively less developed in astronomy when compared with fields such as biology.

In terms of a wish list for the future, the Galaxy Zoo team are working on refining the user interfaces for the projects. They realised that they need to make better connections between the citizen scientists and the scientists, as "one of the main problems with having more projects is that we need people to filter the important questions that only we can answer." Further ideas included a desire for general journal discussion, similar to journal clubs in the biological sciences, via Twitter and/or Skype, "where you can sit down with a paper and go, 'I don't understand Figure 3.' Or, 'I think this paper's completely crazy,' and someone else explains why they don't think it is."

Tools and practices of information

Each participant in the study was asked to respond to a short survey designed to gather information on their strategies for finding new research, the software they use in their work, and their dissemination strategies. 76 participants completed the survey. The tables and charts below, which report the results of the survey, should be seen as illustrative rather than definitive, since they are not based on a statistically significant sample. Nevertheless, a number of clear patterns emerge.

Demographic data was also collected via the survey. For this study, the average age of respondents was approximately 39 years old, ranging from 22-73, and respondents had finished their highest degree about 14 years earlier, on average (range <1-47 years). Across the seven cases, there was relatively little variation on this with the exception of the chemistry case study, which relied heavily on postgraduate student responses and thus had a lower average age of about 28 years (n=10), and the Galaxy Zoo case study, which had an average age for respondents of about 33 years old (n=8). The oldest average age were the nuclear physicists, averaging 46 years old (n=14). These differences reflect the snowball sampling strategy of this study and should not be used to conclude anything about the nature of the fields studied.

In terms of academic degrees, the vast majority of respondents have a doctoral degree (80% of respondents), while 14% hold a master's degree and 6% a bachelor's degree. A range of academic ranks is also represented, ranging from senior academics, professors, heads, readers, and emeritus professors (44% of respondents in total), to academics, lecturers, researchers and postdoctoral researchers (34% of respondents), to students (23% of survey respondents).

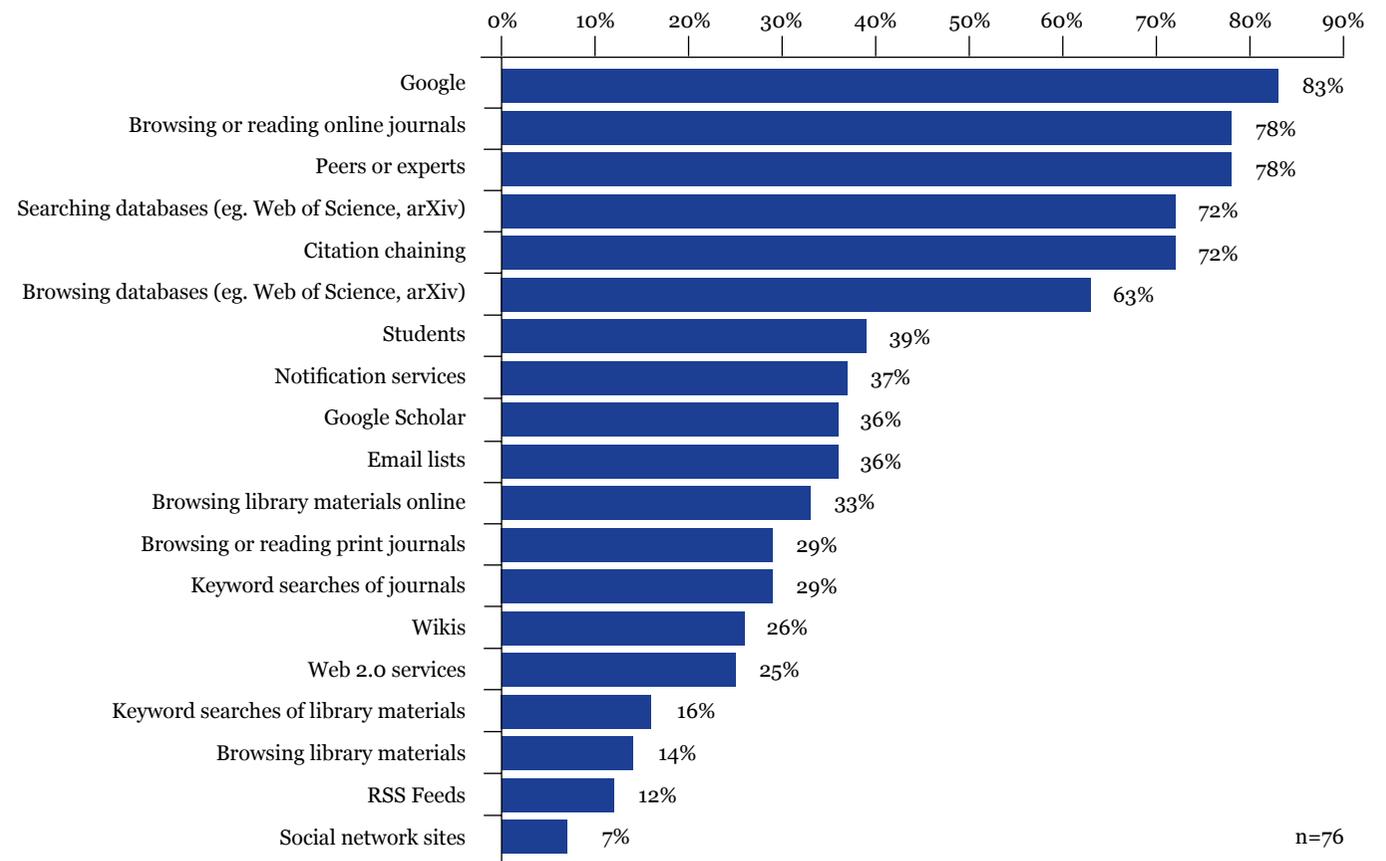
In terms of collaboration, very few respondents (16%, n=12) feel that online modes of interaction are more productive than traditional means of interacting with colleagues such as scholarly conferences and symposia, whereas 40% (n=45) of respondents disagreed or strongly disagreed that online modes were more productive, with the remainder expressing neutral opinions. However, even though these suggest that participants do not hold an uncritically positive view of technology, we will see below that all rely heavily on information technologies for their research, although the exact mix of tools varies widely.

Information sources

In Figure 2, we see the most common strategies employed for finding new research. For this question, respondents were asked to identify all the strategies they use to discover new, relevant research. Consistent with all our other research, Google has become one of the most frequently used sources for information, with 83% of respondents reporting that they use it to find new research. However, nearly as common is browsing or reading online journals (78%) and consulting peers or experts (78%). Searching (72%) and browsing (63%) databases such as Web of Science or arXiv is also very common, as is citation chaining, or following citations from one article to the next (72%). All other strategies are much less common, although as we will show below there is some notable variation by case.

More interesting differences become apparent when looking at these data separated by case study. Figures 3-10 identify those strategies for finding new research which are reported by the participants in each. In these figures, the totals are greater than 100% overall because respondents were asked to report on all the strategies they used when finding new research. Later, we will discuss how looking only at the top two most important strategies changes the picture slightly, both overall and for some of the individual cases.

Figure 2: Strategies for finding new research



Google, which is the most commonly-mentioned information source overall, also shows relatively little variation across cases: across all cases, more than half of respondents (and in some cases, all respondents) reported that Google search is an important way of finding new research information. Use of Google Scholar, on the other hand, is much more uneven across the cases. While more than 70% of the respondents in the earth science and nanoscience cases used Google Scholar, only a small proportion of respondents from the particle physics and gamma ray burst astrophysics cases use it.

Figure 3: Google

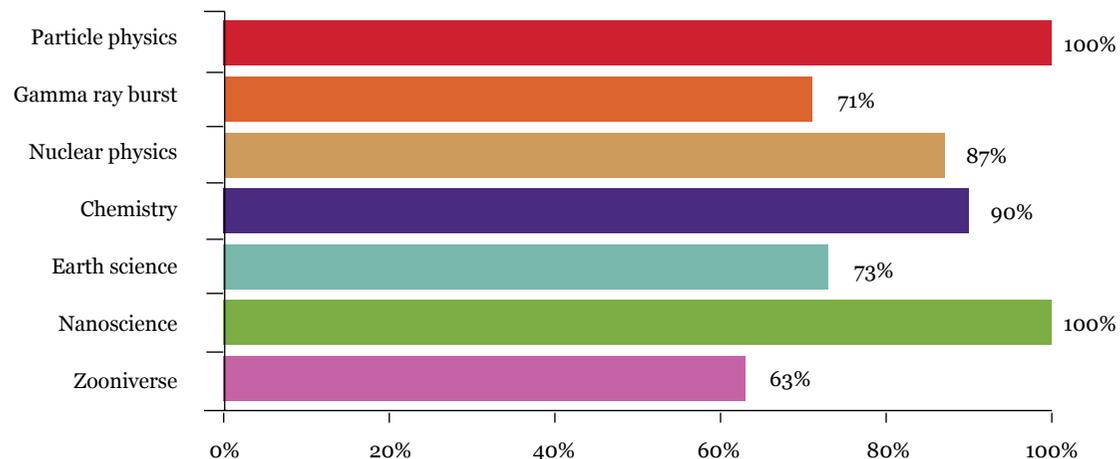
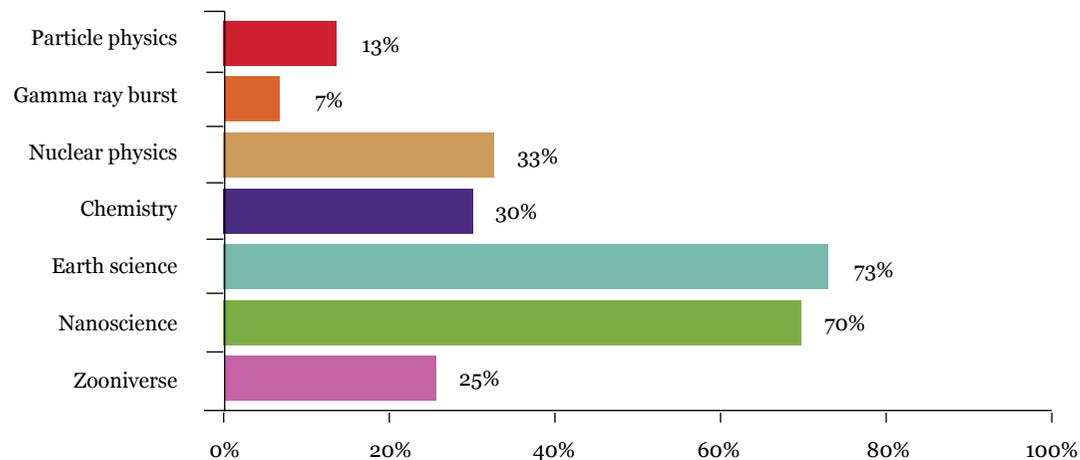


Figure 4: Google Scholar



Chemistry (100% of respondents), earth science (100%) and nuclear physics (93%) reported the highest reliance on browsing or reading online journals to find new research information. In the case studies, the nuclear physics respondents in particular noted that the literature they need to follow is well-defined, which is one of the reasons they gave for relying on journals for new information. The particle physicists, on the other hand, relied more on pre-print archives, and felt that waiting to read information in a journal would mean falling behind the latest developments in their fast-moving field.

Few of the respondents across the cases browsed library materials online. In some ways, these two tables highlight one of the challenges for libraries: even though many of the online journals being accessed by scientists are made available to them via their institutional library subscriptions, few perceive these to be library materials.

Figure 5: Browsing or reading online journals

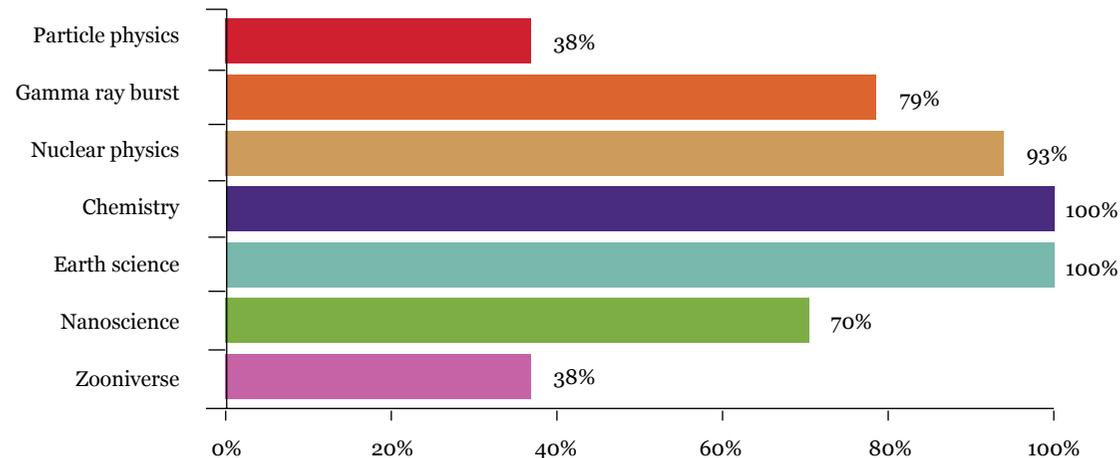
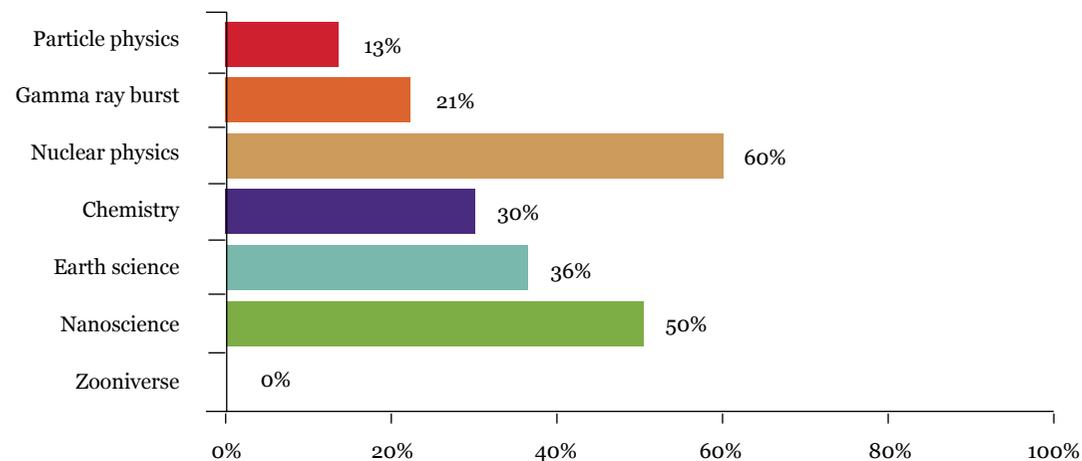


Figure 6: Browsing library materials online



Peers and experts are important across the cases as sources of new research information, highlighting the continued importance of supporting opportunities for face-to-face interchanges even in the digital environment. However, email lists played relatively little importance except for particle physicists and the Zooniverse scientists.

Figure 7: Peers or experts

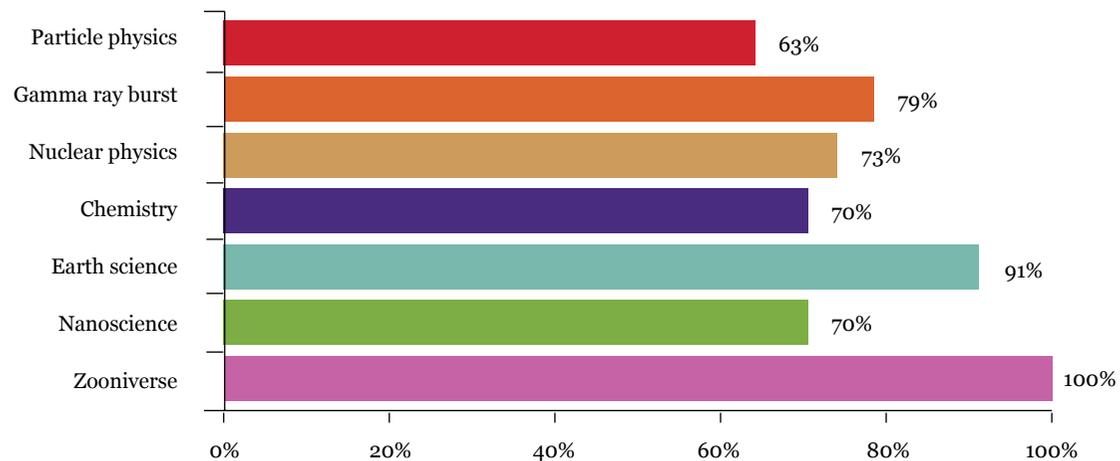
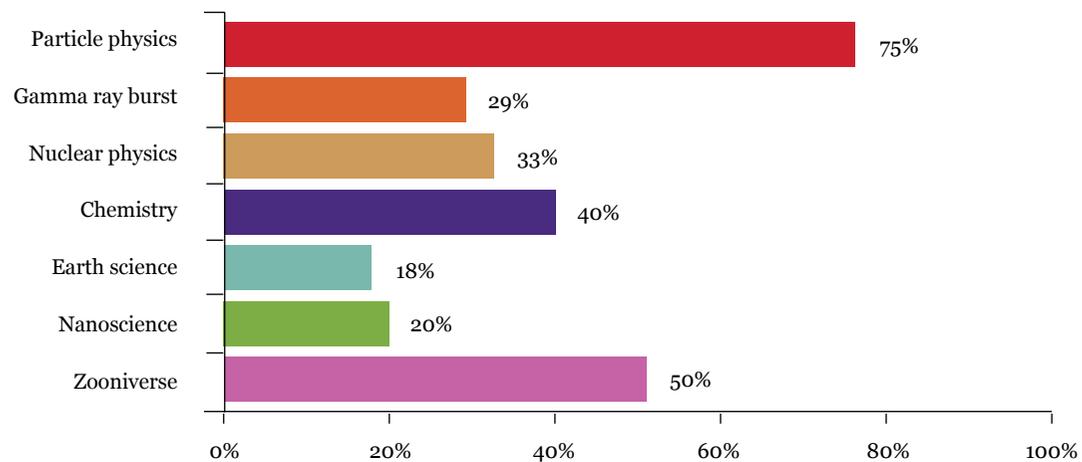


Figure 8: Email lists



Scientists in this study seem relatively slow to adopt Web 2.0 services. With the notable exception of the Zooniverse scientists (who are more likely than others to communicate with the general public because of the nature of the Zooniverse project), few participants reported that Web 2.0 services played a role in finding out about new research. Participants did not seem to view wikis as Web 2.0 tools, and they were only important to particle physicists (who are using CERN's TWikis) and chemists.

Figure 9: Web 2.0 services

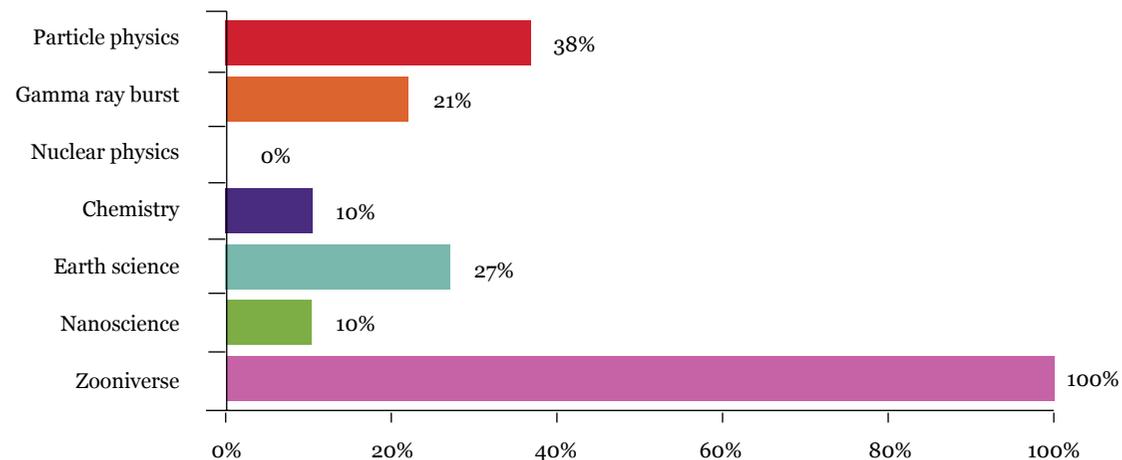
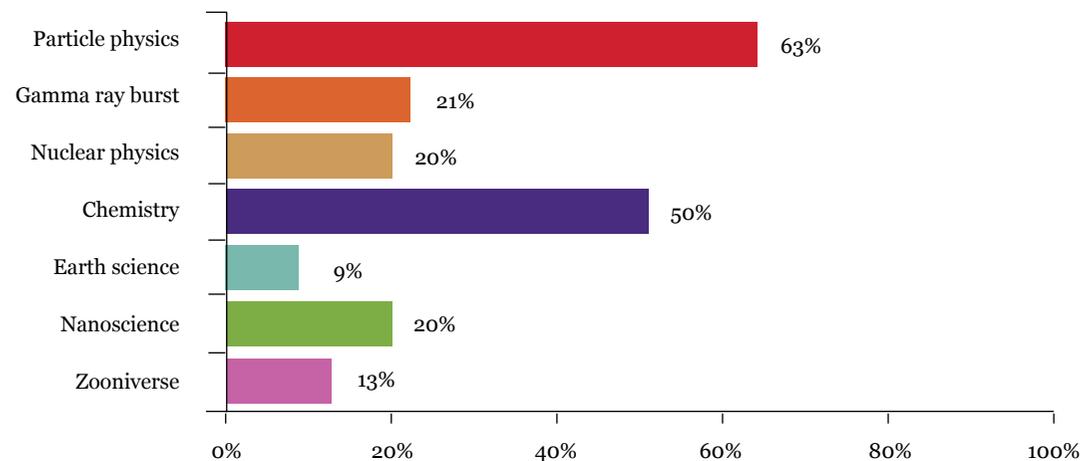


Figure 10: Wikis



Patterns by case

The particle physicists in our study were more likely than other participants to use Google, email lists, wikis, and RSS feeds, but less likely to use online journals, Google Scholar, or to go to the library to browse materials. The respondents in this case were the only group in the study to report that they rely heavily on email lists, with three-quarters of respondents finding information about new research this way.

All of the astrophysics gamma ray burst participants reported browsing databases for new research; in interviews they said the most frequently used resource is the NASA Astrophysics Data System (ADS). Respondents for this case used were less likely to use Google Scholar than other cases, and none reported browsing library materials in person or using RSS feeds. This is interesting because it is possible to subscribe to the ADS via RSS feed, and their astrophysics colleagues in the Zooniverse case study used this function regularly.

Most of the nuclear physicists in the study reported using online journals, and they were also likely to browse library materials online or in person. None reported using Web 2.0 services to learn about new research.

The respondents in the chemistry case study reported using online journals, Google, notification services, keyword journal searches and wikis. They were likely to get new

research information from students, a finding which probably reflects the high proportion of students participating in this particular case.

The respondents from the earth science case were the most likely (with the chemists) to get new information from online journals, and from Google Scholar. They also relied heavily on peers and experts. They are less likely to browse databases, and none reported browsing library materials in person.

The nanoscience participants were second only to earth scientists in the study in their use of Google Scholar. Searching databases was an important strategy for most respondents, but they were also the group most reliant of any of the cases on print journals.

The Zooniverse scientists are particularly interesting. Their use of Web 2.0 services to discover new information sources is much higher than any other group. Likewise, the participants in this case all relied on peers or experts as a source of information about new research, and were the only case other than particle physics to report much reliance on RSS feeds, which they later explained was used to subscribe to the ADS feed. On the other hand, Zooniverse respondents were less likely than others to rely on online journals for new research information, and none used print journals or library materials, either online or in person.

The charts shown here tell an interesting story about disciplinary differences, and how these differences in disciplinary culture, norms of behaviour, and the demands of the science all strongly shape the ways in which scientists access and use information. Zooniverse participants, who are engaged in a collaborative effort to classify and analyse data about the universe, are the most likely to rely on their peers and to use Web 2.0 social media tools to communicate with each other. Tools are used very differently in the various fields: even a generic and readily available tool such as Google Scholar was used by only 7% of the gamma ray burst astronomers but 70% of the nanoscience participants. These findings all suggest that the desire for general purpose tools to serve the needs of scientists is, at some level, a pipe dream. Even the ubiquitous search engine Google, which is certainly highly used, shows some variation.

To better understand some of these differences, respondents were also asked to rank the top five resources from among all those they are marked as the most important (with 1 being the most important of all). If we use these data to look just at the information strategies that respondents ranked 1st and 2nd most important, we see a slight re-ordering of some of the data, and some additional information emerges.

In Figure 11, we see a stacked chart of the items which were the first and second most important resources for finding new research. Whereas Google was the most commonly reported resource in the data above, with 83% of respondents reporting relying on it, it is the first or second most important source for a more modest 29% of respondents. Searching databases (such as arXiv, ADS, or Web of Science) is most important for the highest proportion of respondents (32%), and browsing databases is also important (19%) among the participants in this study.

We can also divide these data out by case, which we have done in Figure 12. In this chart, we have combined the data about searching and browsing databases together (with a positive response to either counting as a single positive response for the combined variable). The differences between cases becomes immediately clear, and we can see that for several of the cases (astronomy, Zooniverse, nanoscience) over 60% of respondents ranked browsing or searching databases as one of the top two strategies for finding information about new research topics. At the other end of the scale, fewer than 15% of particle physics and earth science respondents reported this as one of their top two sources of information.

Figure 11: Information strategies identified as first or second most important for finding new research

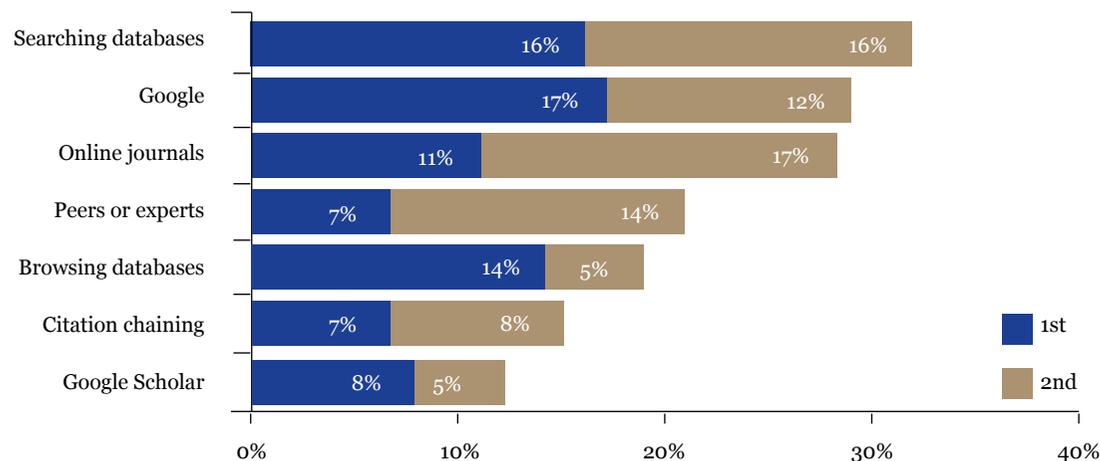
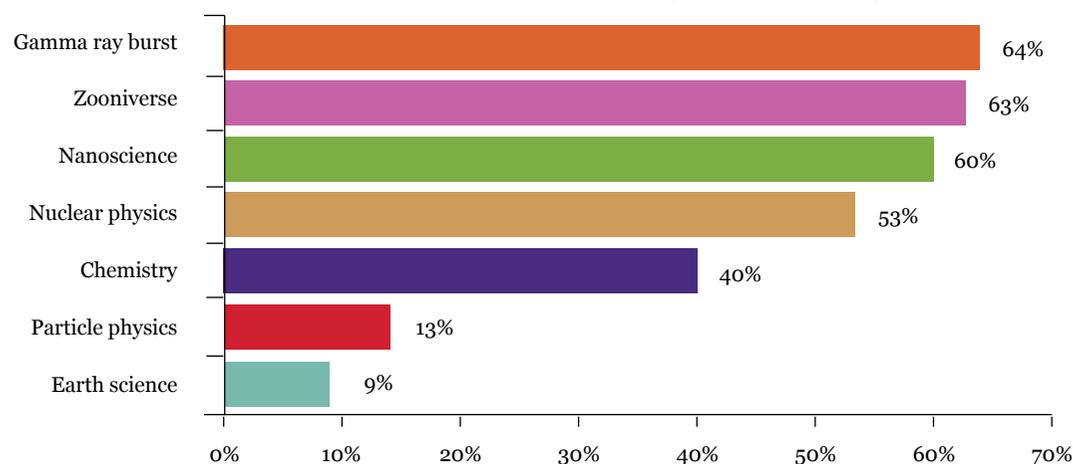
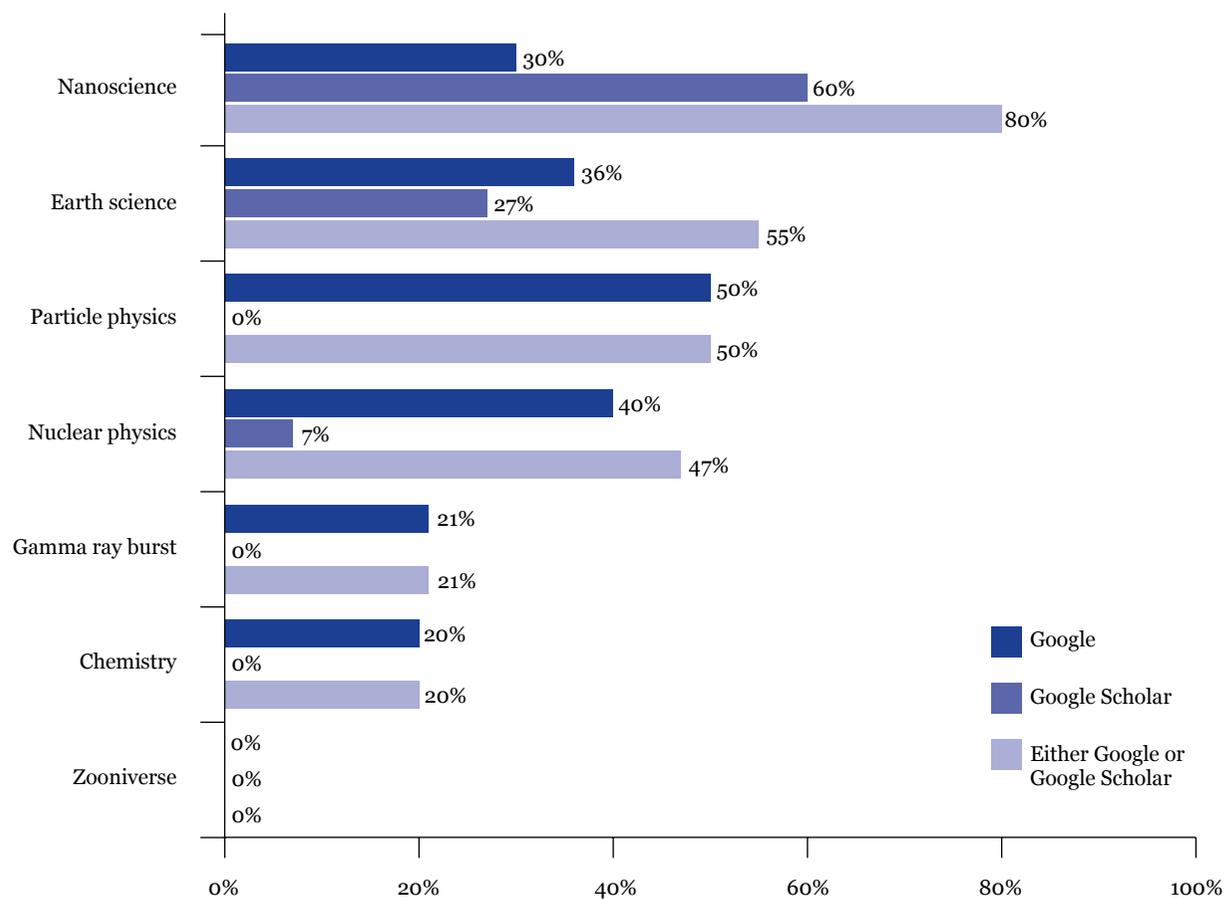


Figure 12: Searching or browsing databases as first or second most important strategy, by case



In Figure 13, we look at another set of important information sources in more detail: Google and Google Scholar. In Figure 2, we saw that overall, Google was used by 83% of respondents, and Google Scholar was used by 36%. We can see here that the use of either of these tools from Google not only vary widely by case, but also vary widely in terms of which of the two Google products is preferred for finding information. So, in the nanoscience case, 80% of respondents ranked either Google or Google Scholar as one of their first two most important strategies, and of these, Google Scholar was the more important (60% of respondents). None of the other cases reported that Google Scholar was more important than Google, and four of the cases did not have any respondents identify Google Scholar as one of the two most important resources (particle physics, astronomy, chemistry, and Zooniverse).

Figure 13: Google or Google Scholar as first or second most important strategy



Research software

In Table 2, we see the proportion of respondents in each case who report using different software tools in their research. The most commonly used software are applications developed in-house, a finding consistent with research we have reported elsewhere (Meyer & Dutton, 2009). Zooniverse scientists were most likely to use in-house software, while the chemists in the study were least likely to do so.

Databases and spreadsheets were used by about half to two-thirds of respondents fairly consistently across the cases, as were programming packages such as MATLAB (although fewer chemistry participants reported using software in this category).

Image processing, simulation, and modelling/3D imaging software were much more variable across the cases. For instance, while chemists appear to rely heavily in modelling/3D imaging software and very little on simulation software, the reverse is true of particle physicists.

There are several other points to note in the data. First is that the particle physicists are reliant on the Grid. This is unsurprising as the respondents were recruited via their participation in the GridPP project, but none of the other cases report any significant use of the Grid. Also, while the use of citation management software is relatively uncommon, the chemistry respondents – who were predominantly students – had a much higher reliance on citation management tools.

Table 2. Software used for research (percentage of respondents reporting any use for each type)

	Particle Physics (n=8)	Gamma Ray Burst (n=14)	Nuclear Physics (n=15)	Chemistry (n=10)	Earth Science (n=11)	Nano-science (n=10)	Zooniverse (n=8)
In-house software	75	79	73	40	75	75	75
Databases and spreadsheets	63	57	67	70	63	63	63
Programming, including MATLAB	63	57	53	30	63	63	63
Image processing	13	57	13	40	13	13	13
Simulation	63	50	53	10	63	63	63
Visualisation	25	29	40	40	25	25	25
Citation management software	13	7	7	70	13	13	13
Modelling or 3D imaging	-	21	7	70	-	-	-
Scanners or cameras	13	-	27	40	13	13	13
Statistics	25	21	13	20	25	25	25
Grid	100	14	-	-	9	-	13

Dissemination

We also asked respondents several questions about their opinions about online dissemination. Overall, nearly half of respondents disseminated their research online, either always or regularly. Some posted research online sometimes or infrequently, but only a few never disseminated their results online.

Online journals,⁶ public repositories, personal websites and departmental websites are all common outlets for disseminating research online among our respondents. Respondents were also asked about the advantages and disadvantages of online publication, and the responses are shown in Figure 15 and Figure 16.

Figure 14: All outlets reported being used to disseminate results online (n=76)

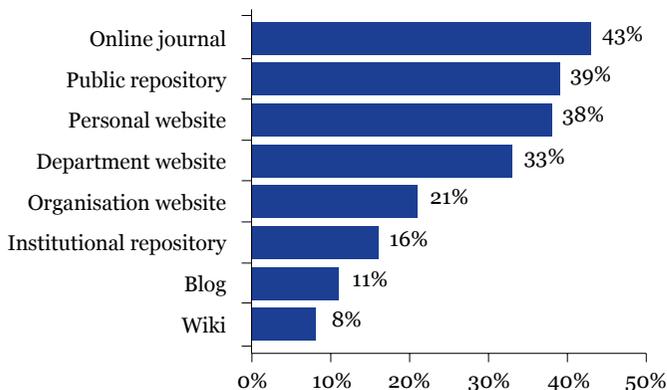


Figure 15: Advantages of online publication, top three in order of importance (n=76)

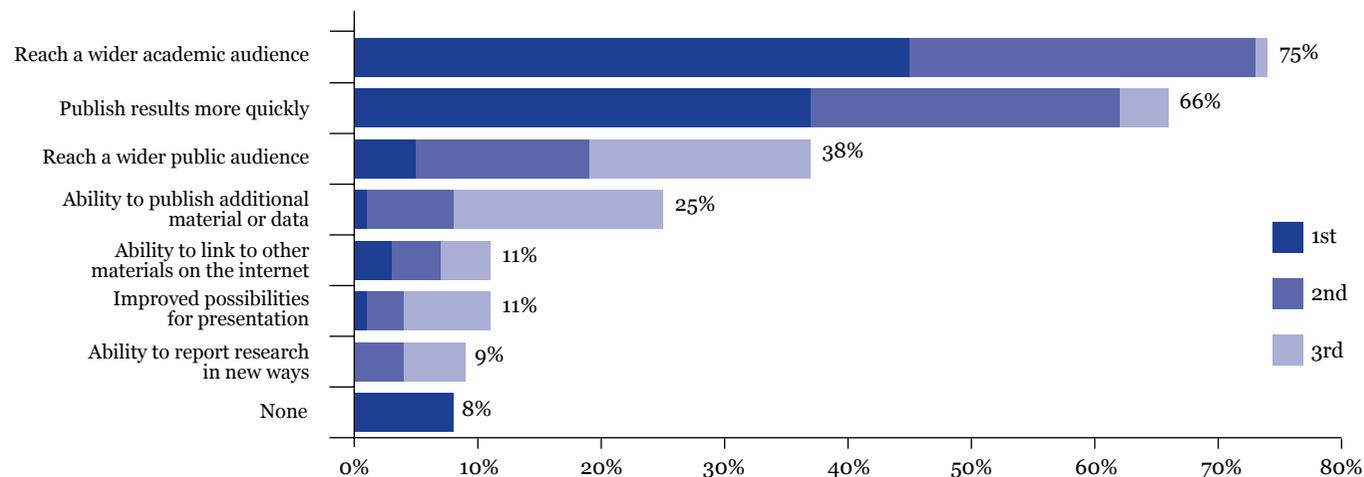
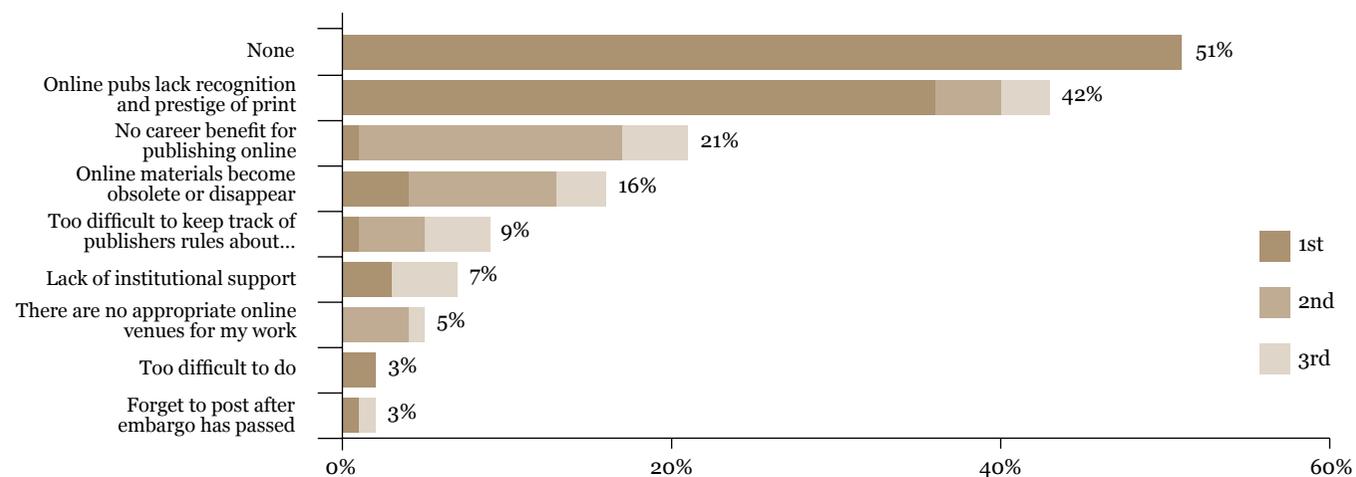


Figure 16: Disadvantages of online publication, top three in order of importance (n=76)



Participants were asked to rank up to three advantages and up to three disadvantages of online publication. The top perceived advantages include reaching a wider academic audience and publishing results more quickly. Relatively few felt that linking to other materials, presenting findings in new ways, or reporting the research in new ways are key advantages of disseminating material online.

Most respondents felt that online publication had no disadvantages, the most common response. Among those who identified up to three disadvantages, a perceived lack of prestige of online publications was the most commonly identified. Few found publishing online difficult to do, or lacking in appropriate venues.

Complexity

One concept that drove the design of this study is what we call the complexity continuum. Information in today's world is created and stored in increasingly complex ways. We can conceptualise the range of information seeking, access, use, and dissemination behaviours in the physical sciences as a two-dimensional complexity continuum, shown in the diagrams below. Any individual researcher will use and create information materials from across this continuum, as virtually all researchers today employ a mix of digital and analogue resources exhibiting a range of complexity. However, certain patterns have emerged from the case studies.

The horizontal axis ranges from lower to higher complexity in terms of the user perception of the computational resources required to create, maintain, access, and use research information resources. This axis of computational complexity is labeled from lower to higher rather than from low to high because the perception of computational complexity changes over time and at different speeds in different fields and disciplines. Twenty years ago, electronic full-text journals were considered to be complex computer-based information resources compared to other information resources of the day. Today, however, these same full-text versions of journals would be considered fairly mundane in comparison to other more complex resources. The scale does not imply

any value judgement: information resources with lower computational complexity may be just as important and valuable, even more so, than their more complex cousins. However, from a socio-technical perspective, the relative complexity of one set of resources in comparison to another has important implications in terms of types of users, uses, and the likelihood of uptake.

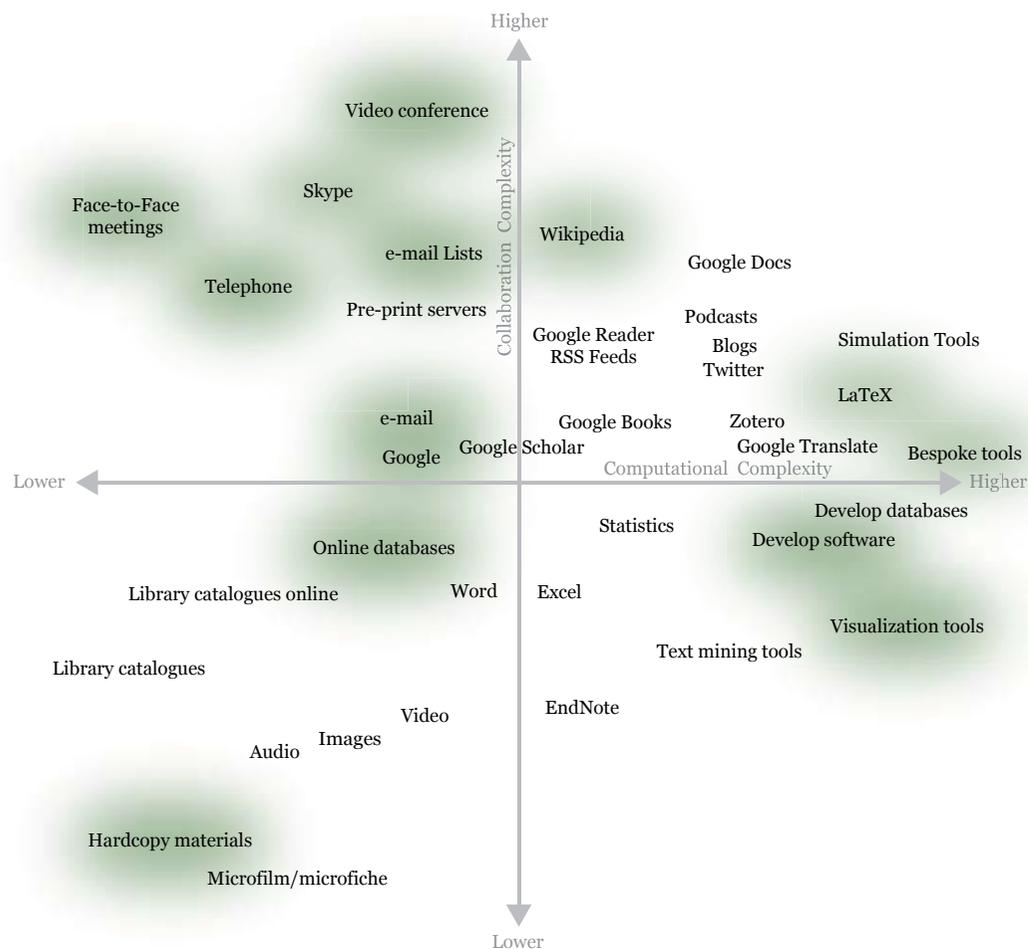
Similarly, the vertical axis indicates the collaboration complexity of various types of information behaviours. Considerable research has shown that the social complexity of collaborative relationships is often one of the biggest challenges faced by those implementing complex information systems and resources (e.g., Cummings & Kiesler, 2005; Fry, Schroeder, & den Besten, 2009; Meyer, 2009). The lower portion of the complexity continuum represents those information behaviours that take place in isolation, or by way of relatively small or simple social arrangements. The upper portion includes those information behaviours that are linked to more complex social coordination and collaboration.

The scientists in the case studies use tools from across the spectrum, as shown on the complexity continuum above. This shows their tools in a conceptual space, with some of the most commonly used resources highlighted. The shaded

tools were identified by participants in at least six of the seven cases as important resources or types of tools for their field. As seen in this conceptual diagram, physical scientists across disciplines appear to share a need to collaborate, but often do so using relatively simple technological tools, such as the telephone, Skype, videoconferences, email, and email lists; face-to-face meetings also remain important. But the choice to collaborate using less complex means does not appear to be due to any innate problem with engaging with technology: the other area of broadly shared uses are at the far right end of the computational complexity axis: developing and using bespoke software and tools, visualising results, and using specialised word processing software such as LaTeX instead of more general purpose tools such as Microsoft Word.

Some general purpose tools are used widely, including Google, Wikipedia, and published online databases (which can include Web of Science, ADS, and a range of other discipline- and field-specific databases). Finally, hardcopy materials continue to be used, but not necessarily in the same ways that they have been used in the past. So, for instance, while the nanoscience and nuclear physics participants identified several standard physics texts which are consulted regularly, few of the other cases specifically mentioned printed books as important resources.

Figure 17: Physical sciences complexity continuum



Some things are mentioned exclusively in the humanities cases, such as reliance on audio and video, consulting materials via microfiche, using text-mining tools, and using public tools such as Google Translate and Google Docs. Likewise, the physical sciences in some cases use tools such as the Grid and code repositories when appropriate. Overall, this data tells an important story: while general purpose computing has pervaded the academic fields in both studies, researchers are critical users of technology and have been shown to be adept at discovering, adopting, adapting, and sharing tools, resources, and social arrangements appropriate to their needs. This reinforces the notion that researchers are not uncritical ‘users’ of technology, but are critical social actors (Lamb & Kling, 2003), who operate within important professional domains that help them to engage with new technology options that are available to them. This is important because it reinforces the notion that the failure of technology to be adopted is often not because ‘users’ have failed to understand how to use the technology (which is what developers often choose to emphasise), but because technologies have been built which have been judged by researchers not to be appropriate for their needs. When appropriate and useful technologies are made available, their uptake is often widespread, as can be seen from the success of Skype and Google.

Table 3: Resource use by case

Resource	Particle Physics	Gamma Ray Burst	Nuclear Physics	Chemistry	Earth Science	Nano-science	Zooniverse
Higher computation and Higher collaboration (top-right quadrant)							
The Grid	●	●					
Code repositories	●		●	●	●		●
Simulation tools	●	●	●		●	●	
LaTeX	●	●	●		●	●	●
BibTeX	●	●	●		●	●	●
Bespoke tools	●	●	●	●	●	●	●
Wikis	●	●	●	●			
Wikipedia	●		●	●	●	●	
Podcasts		●					●
Google Reader					●		●
RSS Feeds	●						●
Blogs		●				●	
Twitter					●		●
Zotero				●	●	●	
Mendeley							●
Google Books					●	●	
Google Scholar			●		●	●	●

Collaborative yet independent:
Tools and practices of information

Resource	Particle Physics	Gamma Ray Burst	Nuclear Physics	Chemistry	Earth Science	Nano-science	Zooniverse
Lower computation and Lower collaboration (top-left quadrant)							
Google	●	●	●	●	●	●	●
Pre-print servers (such as arXiv)	●	●	●				●
email	●	●	●	●	●	●	●
email lists	●	●	●		●	●	
Video conference	●		●	●	●	●	●
Skype	●	●	●		●	●	●
Telephone	●	●	●		●	●	●
Face-to-face meetings	●	●	●	●	●	●	●
Higher computation and Higher collaboration (bottom-right quadrant)							
Develop databases	●				●		●
Develop software	●	●	●	●	●	●	●
Visualisation tools	●	●	●	●	●		
Statistics	●	●	●	●	●		
Excel				●	●		
EndNote				●	●		
Lower computation and Lower collaboration (bottom-left quadrant)							
Hardcopy materials	●	●	●	●	●	●	●
Library catalogues			●				
Library catalogues online		●	●	●	●		●
Audio							
Images		●		●	●		●
Video						●	
Online databases (such as <i>Web of Science</i> , ADS, and other published databases)	●	●	●	●	●	●	●
Word		●		●	●		●

Table 3 shows resources that were identified in the interviews and surveys by more than one respondent per case as important sources of research information. We can see that, as visualised in the continuum, the lower computation but higher collaboration resources are broadly shared and widely used. In the cells corresponding to the top-right quadrant of high computational and high collaborative complexity, we see that there is considerable evidence that physical scientists engage with these complex resources, but there is much less consistency across cases as to the *particular* resources used. Thus, while only particle physicists and astrophysicists identified the Grid as an important resource for their science, five of the seven cases identified shared code repositories and/or simulation tools as important resources. There is also uneven use of social media, such as blogs, Twitter, shared bibliographic software (Zotero and Mendeley), podcasts, or RSS feeds.

Taken together, this complexity continuum and resource table paint a complex picture of information practices across these seven cases, and reiterate the idea that there is rarely a one-size-fits-all solution that is appropriate for science. This is underscored by the importance of bespoke software, developed for specific scientific purpose, across all the cases. Google, of course, is one of the stand-out examples of the opposite: a general-purpose tool developed for a very wide audience which is indeed useful across the cases.

Conclusions and recommendations

Several clear patterns emerged from this study about information practices in the physical sciences. The first is reflected in the title of this report *Collaborative yet independent*: while the physical sciences rely heavily on collaboration, the individual scientist and individual fields remain very important. Within collaborations and within research fields, there is often broad agreement about the important questions to be pursued, and researchers make considerable shared efforts to pursue them. However, individual choices and efforts are still important in terms of information use and career progression. Even within collaborations there is considerable variation in individual choices, and most scientists are still judged independently of one another.

Information retrieval

While Google, and to a much lesser extent Google Scholar, are important information seeking and retrieval tools, the cases in this study show that a much broader range of tools are in general use, and that these vary from case to case. Specialised tools such as arXiv, SPIRES, the Astrophysical Data System, the National Nuclear Data Centre, SciFinder, and individual journal websites are all important. The cases show both convergence and continued diversity: it is undeniable that many fields are converging upon Google as a general purpose tool but it is only one of many information search and retrieval strategies. Beyond Google, there is a clear diversity of specialised tools suited to individual fields and disciplines.

Peers have always been, and seem destined to remain, important. Talking to peers and experts seems likely to remain one of the most important ways to learn about new research across all fields and disciplines. This reflects the importance of trusted sources: just as peer-reviewed high quality journals help to inspire trust in the information they present, people grow to trust their colleagues and rely on this trust to weigh the information coming from their peers. Peers who prove trustworthy will have their ideas and opinions trusted more in the future.

Information and data management

Information overload was not present in every case. For instance, while particle physicists complained about information overload, nuclear physics participants generally felt able to keep up with important developments in their field.

However, information overload was a reasonably consistent problem when it came to handling emails. Reading, replying and storing to hundreds of emails each week takes considerable time. This growth in email seems to be from existing colleagues working together on projects, papers and proposals, plus administrative information, rather than people highlighting new information sources. It is exacerbated by an increasing expectation of round-the-clock work at home, while travelling, and at other times when researchers used to be unreachable.

Individual habits of storing and re-accessing research information varied widely. Some still printed out papers to read, but many more either saved PDFs to their computer, or relied on the continued availability of a known copy online that they can refer to if necessary. Those relying on online copies assume that the copy will stay online, and that their institution will continue to subscribe to the service that provides access.

Data analysis

For our participants, the most complex computing they undertook related to data analysis, and many relied on powerful tools and large datasets. Unlike searching for background literature or finding supporting data and information, which can be seen as a necessary task that supports scientific progress, data analysis is seen as central to the scientific endeavour. For instance, while respondents in several of the cases reported spending 20-30% of their time searching for information about a new problem, they could easily spend 70% of their time analysing and understanding their research data for a new problem.

Across the cases, there is a wide variety of commercial, open-source, and bespoke software used for data analysis. The generally high reliance on bespoke tools, built for specific research needs, suggests the need for flexibility in research infrastructures. While general purpose tools fulfil certain needs, specific scientific questions seem to need specialised tools.

The complexity of data analysis in each case is, to a certain extent, dictated by the available research technologies, and the nature of the data they produce. The data being generated by CERN's Large Hadron Collider is at the petabyte scale, and requires a huge amount of computation,

storage, and processing power. In nuclear physics, datasets are too big to be transported by networks, so scientists use cheap and portable hard drives in the terabyte range to carry research data by hand from distant laboratories to their home facilities, where they are analysed. Many of the chemistry students, on the other hand, were working with much smaller datasets, which could be stored and analysed in off-the-shelf tools, particularly Excel spreadsheets.

Programming skills are increasingly important in the physical sciences, whether programming functions in MATLAB, or writing programmes in C++, or writing code that will run in any of a wide variety of specialised programmes. Across the cases, scientists identified the need to either have programming skills or to have access to programmers. Much of the data generated by many of the scientists in all the cases needs to be cleaned, transformed, and analysed, so automated programmes, routines, and tools must be created to assist in this work. Pieces of code may also be shared across collaborations and with other research teams, who often re-use or modify the code to suit their own needs.

Citation practices

Citing the work of others, and being cited by others, is one of the ways in which science progresses. Science is a progressive endeavour, and new work inevitably builds upon previous work. At a more pragmatic level, however, citation measures are increasingly important in judging individuals and departments, with measures such as the h-index becoming standard ways to evaluate a scientist's productivity and impact.

Even though most work is accessed electronically, researchers generally cite the printed version of journal articles. Few felt it necessary to reference databases, particularly in fields where the sources used across the field are so standard that other scientists would already know which databases they probably used. The Astrophysics Data System, for instance, has a suggested text to acknowledge use of the database, but gamma ray burst astrophysics participants felt it was unnecessary to use this text because it is obvious that ADS was used.

There is little agreement on how to cite databases, or otherwise assign credit to the scientists and technicians responsible for the creation and maintenance of databases. This is important at several levels. First, without a means to be assigned credit for their work, those responsible for creating data have fewer career incentives to engage in such efforts. Second, the ability to replicate scientific work relies on being able to identify not only the data used for analysis, but the version of the data used, in cases where databases continue to grow and change. Being able to cite the version of the data used will help those interested in verifying or re-analysing data.

Dissemination practices

There is a perception in many of the fastest moving fields that the print process and, in some fields, even the peer review process, has made formal publication too slow. In many of the big collaborations or highly collaborative fields, articles submitted for publication have gone through extensive internal review, and are therefore considered to be of citeable quality, even before they are published. In other cases, such as the chemistry case study, scientists prefer to wait until an article has been formally peer-reviewed before citing the work. In general, however, the expected times from submission to publication appear to be shorter in the physical sciences (within weeks or months) than in the social sciences (where submission, peer-review, and publication can take many months or even years) and the humanities (where book-length publications can take many years of preparation and editing before appearing in print). These fast publication schedules raise important questions for the future of peer review in the sciences, especially in view of changes currently taking place across the publishing industry.

In some fields, much initial dissemination takes place outside the formal publication process. The first results are often shared in meetings (both formal and informal) and in email communications. The gamma ray burst

participants, for instance, suggested their field was quite a 'talky' community, with some productive scientists giving ten or more talks a year to share results. Nevertheless, formal publication remains the gold standard, and even researchers in those fields that share results more quickly still expect that those early results will eventually be written up, peer-reviewed, and published.

With the exception of the Zooniverse scientists, few of the participants were using tools such as blogs, Twitter, open notebooks, social networks, public wikis, or other public-facing technologies to share research information (although some such as particle physicists and astrophysicists use internal, private wikis). For most, these social media were viewed as distractions from the communications they had with their most important colleagues within their community of practice. The Zooniverse scientists, on the other hand, rely on the public to contribute to their scientific work, and thus have an incentive to keep the public informed of, and interested in, their work.

The most common means of disseminating results online (reported in the cross-case survey data) are via online journals, public repositories, personal websites, and departmental websites. Participants self-defined 'online publications' for the purposes of this question, so it is difficult

to know exactly what they mean, but their answers suggest that they are not talking about electronic versions of print publications. Most respondents felt that online publication allows work to reach a wider academic audience and permits faster publication. Relatively few, on the other hand, thought that online publication would allow them to present research in new ways enabled by technology or to link to other work. The biggest concern was that some online publications may lack the prestige of print publications, but around half of participants felt there was no disadvantage to publishing results online.

Collaboration

New and emerging technologies are changing the way scientists gather and analyse data, and have been doing so for many years. The research facilities on which some (but by no means all) physical scientists rely have been getting larger and more technologically complex, and generate more data than ever before. This has resulted in larger collaborations in some fields, as scientists coalesce around these rare, expensive, shared facilities. The process of collaboration has inspired new communications technologies, which in turn have changed the way the scientists collaborate. They work in less isolation, engage in more frequent and more rapidly-developing conversations, and report a more democratic approach to decision making among collaborators. In the case of nanoscience, for instance, one participant argued that the field itself would have emerged much more slowly without the Internet because there would be no way to learn of new and interesting developments.

While collaboration in general has increased, not all science is done with large teams. In the chemistry case, many collaborations are small, and the equipment rarely demands large collaborative effort to build and maintain. Also, the Zooniverse case demonstrates that large scale collaboration does not always require large infrastructure investments: the effort of thousands of citizen scientists has been harnessed for the relatively minor cost of building the Zooniverse web-based platform.

While ‘invisible colleges’ (which have existed since the earliest development of science and scientific disciplines) are greatly enhanced by modern communications, people still have only finite time available. Time spent dealing with email is time lost to local, personal communities. This raises an important question: are departments becoming less meaningful entities for research? If so, what is the implication of using departments as the focus for research assessment exercises, if scientists’ most important collaborations cross departments, universities or research facilities, or even countries?

Transformations in practice

Complex approaches to computation

Computational capabilities have increased dramatically, which has had a significant effect in certain cases, such as the particle physicists working with data from the Large Hadron Collider. Distributed computation ranges from the supercomputing power harnessed to the Grid to the power of human brains classifying galaxies via the Zooniverse. Data is available more widely, is generated and released more rapidly, and is increasingly available in standardised formats which support sharing and reuse.

These technological advancements are part of a positive feedback loop: as collaboration-enhancing technologies advance, scientists engage in more cross-institution sharing and international collaboration, which in turn creates demands for newer, more efficient, and larger scale technologies to support collaborative research. Rather than spending a career becoming an expert in the quirks and anomalies of particular datasets, scientists are able to access more data and more easily compare it to other datasets to advance their scientific research. It is not yet clear what this means for career trajectories and the evolving roles of scientific team members, but new opportunities are likely to become available for scientists skilled at large-scale data analysis.

Simple approaches to information

While physical scientists use complex and powerful technology for their research, there is some evidence that many are less sophisticated users of information sources than the researchers in the previous life sciences and humanities case studies. Few were using innovative information search and retrieval strategies, most relied on relatively simple systems for organising the information they discovered, and many did not understand the full capabilities of the tools they were using. For instance, some participants were dissatisfied with their ability to annotate PDFs, apparently unaware that the technology exists (in tools such as Adobe Acrobat Pro) to meet some of their expressed needs. Likewise, Google was used widely, but more specialised (but still generally available) tools geared towards academic work such as Google Scholar were used much less frequently. Across the cases, participants relied on standard off-the-shelf tools for information search, which in turn leads to a somewhat generic experience and set of results. Again, this is in contrast to the widespread practice of building their own highly-specialised tools for science, with a very high level of use of in-house or other bespoke software solutions for science. Such creativity was not evident in the tools used for information search.

One explanation for this is that many of the cases here are very well-bounded, and exhibit what Whitley (2000) calls high mutual dependence and low task uncertainty. In other words, any individual scientist in the collaborative physical sciences relies strongly on his or her colleagues and on their shared facilities to contribute to scientific progress (high mutual dependence) and has a well-defined understanding of what constitutes important research and where the results of that research can be found (low task uncertainty). When research relevant to one's area of science is only likely to appear in a handful of journals which can be easily monitored, there is little incentive to build elaborate strategies for information discovery that is unlikely to yield much additional important information.

Disciplinary and field differences

The 'disciplinary difference' literature has often focused on the broad differences between, for instance, physics, engineering, chemistry, and so forth, but here we have seen that even within disciplines there can be considerable field differences. For instance, nuclear physicists and particle physicists are very different in the ways they find new research information. And within fields or collaborations, there can be differences, such as the participant in the Zooniverse case study who indicated that a theoretical cosmologist would routinely be expected to post a paper on

arXiv for review and comment prior to journal submission, whereas a star formation specialist would only post a paper after it had been accepted by a journal.

New multidisciplinary approaches will create new challenges in terms of negotiating how information is discovered, shared, cited, and disseminated. In the Zooniverse case, the astrophysicists who started the project are now collaborating with humanities scholars who use the Zooniverse platform and community to classify ancient papyri. However, even when different scientific fields are collaborating, as in earth science or nanoscience, they retain their existing habits and practices (gained when learning their field and discipline), presenting a challenge that must be negotiated in multidisciplinary work. Disciplinary practices and expectations shape attitudes toward issues such as how quickly scientific results should be shared, and via which channels, and these attitudes in turn shape behaviours around openness, sharing and collaboration.

It is also worth bearing in mind that once a scientist or other scholar joins a community (typically at the age of 18/19), they are likely to stay within that community for most of their working life if they remain within academia. They also join a specific 'club' within that community when embarking on doctoral work focusing on a narrow question within a

field or sub-field, and many people stay in that club until they retire, and in some cases longer than that. This results in strong cultural norms and shared views within fields.

Two other key findings relate to disciplinary differences. First, just as there is variation between individual scientists, there is marked variation between the cases in this study in terms of information practices and priorities. Participants consider different information sources to be particularly important and these preferences seem to be clustered quite convincingly by case study, strengthening the argument that disciplinary norms are communicated effectively among communities of practice. Google Scholar is one example: the use of this easily available tool for academics ranged from a low of 7% of astrophysics gamma ray burst participants to a high of 73% of earth science participants. Based on our interviews, this cannot be attributed to a single factor such as whether certain journals are indexed within the tool. Rather, it appears to be a combination of the capabilities of the technology (actual or perceived), the attitudes expressed by peers, the existing work practices of participants, and whether other, more specialised, resources are seen as adequate. Thus, the word 'independent' in the title of this report can refer to individual scientists, but can also refer to the independent disciplinary choices of individual fields and specialties within the physical sciences.

The second finding is that when we look across all five cases, there is less overall difference between the physical sciences and the humanities than we expected. We hypothesised that physical scientists would use far more complex technologies, and would be engaged in far more complex working arrangements, than humanities scholars. Physicists in particular are often thought to be at the forefront of developing new technologies for research, and new methods of sharing research (such as arXiv), and this study did find evidence of this behaviour. However, looking across the cases from a broader sociological view, it is striking how much consistency there is across the fields and disciplines. Beyond the obvious case of Google, there has also been wide convergence on resources such as email, Skype, online journals via library gateways, and public resources such as Wikipedia. Furthermore, because humanities scholars often need to discover and track information from a wider range of sources than physical scientists, many of the humanities participants had developed more sophisticated strategies for dealing with information sources, even if they were generally using much less complex systems for working with their research data and materials.

Thus, it is inaccurate to stereotype either physical scientists or humanities scholars as more sophisticated users of research technologies. For instance, hybrid print-electronic practices are common for physical scientists just as they were for humanities scholars. It is far more accurate to say that researchers across the disciplines are adept at identifying tasks in their personal and disciplinary workflows which require computational tools and collaborative approaches, and then developing appropriate tools, skills, and strategies to address those tasks.

New questions

Many respondents did not feel that new technologies have resulted in their asking new scientific questions, instead choosing to focus on the speed and ease of access and the increased quantity of information available. This perception may be because the scientists themselves are part of a changing system, and each month-to-month or year-to-year change seems relatively small and evolutionary. But when comparing the kinds of scientific questions that could be answered in the past with those that can be answered today, it seems clear that many new questions are emerging. Advances in science and information technology have happened in concert with one another, and each has influenced the other.

Even in cases such as Zooniverse, where new technology (a website which supported the process of citizen science) enabled the discovery of completely new types of galaxies, the perception was this was actually 'more of the same' work they had been doing, but was 'more of the same' in ways that exceeded the scientists' expectations. Likewise, nanoscientists felt that new information technologies had allowed their research to have a broader scope, but that many of the fundamental questions remained the same.

This reluctance to credit new technology as an inspiration for new questions is widespread across many disciplines. This may be because researchers are reluctant to be branded technological determinists, or because scientific change and technological change are alternatively pushing each other forward, or because of some other reason altogether. But it seems evident, on balance, that new information technologies have opened up new avenues of exploration.

New technologies

In general, it seems that younger (doctoral and postdoc) students seem to be more comfortable with technology than their older colleagues. But it remains to be seen whether this will result in new technologies becoming deeply embedded into their routines as they age, or whether today's younger researchers will themselves fall behind their younger colleagues in the future.

As young technology-savvy researchers age it will also be important to monitor the extent to which the accelerating pace and volume of digital communication crowds out time for other important activities, such as deep, engaged reading and extended periods of writing. Extensive writing is an activity that some respondents feel is being 'squeezed out' by new forms of communication, although many recognise the value of first-class monographs written by experts. This leads to another interesting, if tentative, conclusion: should there not be mechanisms to encourage and reward this sort of activity?

The Zooniverse case hints at new possibilities for scientific research, and it will be interesting to see how this kind of citizen science develops. The technology that underwrites the Zooniverse website is fairly simple, certainly far simpler than the telescopes which gathered the data. However, the simple web-based technology harnesses a much more

complex system – many human brains – that still has no parallel in computer-based tools. How these and other methods are used to increase the power of science could be an exciting area in coming years.

New technologies for sharing data and for combining data from disparate sources are particularly valuable in multidisciplinary fields such as earth science and nanoscience. Unlike large datasets that are generated by a single machine in some of the other cases, datasets in these fields can originate from a wide variety of sources. The challenge of federating, mining, analysing and interpreting these data will be a key focus in coming years.

More mundane information tools are also of interest to researchers in many of the cases, including better tools for PDF annotation, better systems for managing and storing information, and better tools for citation management. A number of respondents wanted better tools for annotating PDFs, which has become the ubiquitous format for distributing final research papers. Researchers feel that existing tools do not allow them to work with PDFs in the way that they want to – possibly because they cannot access, or are unaware of, more sophisticated functionalities such as annotation which are available in tools such as Adobe Acrobat Pro. Researchers also noted their inability to easily flip back and forth between the text and the list of references in digital papers.

Recommendations

Removing barriers

The cases within this study suggest that there are several important barriers to better information use and management.

- Funding, as always, is an important barrier to developing new strategies, resources, and shared tools, but can also serve as an important driving force for collaboration and information sharing by setting out the parameters for how information should be shared. This is particularly true in areas such as data sharing, where standards and practices are still emerging. A perennial need in this area is identifying sources of sustainable funding for information resources developed as part of funded projects.
- Lack of access is a key barrier to finding and using information. Participants reported that they tended not to track down articles or data to which they had no subscription unless they were certain it was central to their work. The more research and research data that is made available via methods that are (or appear) free to the user, the less of a factor this will be. Whether this is via open access or via other business models is beyond the scope of this study, but it is clear, that lack of access

is often an issue, even for scientists at research-intensive universities.

- For resources available only via institutional subscriptions, remote access arrangements need to be either improved or better communicated to researchers.
- For some cases, information overload is a problem, and methods and tools to filter information more effectively must be developed. Some of these tools may already exist but have not been widely adopted, and others need to be refined to fit into the workflow of scientists. Others will need to be built. However, the volume and flow of information will almost certainly continue to increase, and tools, or changed practices, to manage this are crucial.
- Annotation tools are inadequate, and need to be better developed and more widely distributed. Many researchers believe that the tools for marking up documents such as PDFs are inadequate, and that this presents an important barrier to paper-free reading and use of information.
- The most pressing need for many physical scientists is new technologies and tools for experimentation and data analysis, rather than more information resources, which are mostly viewed as reasonably satisfactory. However,

new information technologies that fulfil unmet (and often unperceived) needs will certainly emerge, and are most likely to achieve uptake if they can fit into existing workflows. The example of Google is clear: 15 years ago no scientists felt they needed it, but now few could imagine working without it. The challenge for those designing new tools is to identify bottlenecks and gaps in current practices, and to build tools that can widen those bottlenecks and bridge those gaps rather than to design tools that require completely new ways of working.

Research councils and funders

There are clearly funding pressures on the physical sciences, although the question of funding for research falls outside the remit of this report. However, in terms of funding and support for information practices, two main areas of potential investment are **increasing existing efforts to link and share data, and providing extra support for training new researchers in how to use and manage information.**

Linking and sharing data was also identified as a potential area for investment in the previous reports in this series, on the life sciences and the humanities. The infrastructures that support shared and linked data in the physical sciences are

different from those needed in life sciences or humanities, which reinforces the idea that one-size-fits-all approaches may not be appropriate or successful. In particular, the physical scientists expressed a desire for new tools to access and analyse data that is generated in shared research facilities. Research funders can invest in the infrastructure and tools needed to enable this.

Funders can also target postgraduate students and postdoctoral researchers for training opportunities in how to manage their information more effectively. While physical science students are well-trained in research tools, there is little evidence that they are being systematically taught the best practices for finding, managing, and disseminating information. These training opportunities should be as targeted as possible: we have reported elsewhere that generic training delivered by computing or library staff is less effective and less engaging than training tailored specifically to demonstrate to a student how their peers are managing their information sources (Eccles, 2010).

Publishers

For the last 15 years, publishers have been facing the challenge of how best to meet demands for easy and free-to-the-user access to research materials, while still maintaining sustainable business models. Scientists

recognise that the advancement of science depends upon rapid availability of high-quality content which can be read by the widest possible audience, and can therefore be enthusiastic supporters of open access. But it is important not to underestimate the value of gatekeepers in science; these roles have been built up over the centuries to ensure that good science is propagated while bad science is not. Publishers have built the cost of maintaining the system of peer review and of disseminating results into their business models, and new models must take these issues into account. In addition, as datasets get larger and there are increasing calls to publish the data that underwrites scientific papers, publishers may need to consider how far they should engage with maintaining data archives and handling quality assurance, version control and access for such services.

Several of the cases suggested specific issues of relevance to publishers. Several earth science participants, for example, complained about the high cost of publishing articles in journals, particularly when colour images were required to convey their scientific results. One researcher suggested that publishers could offer to provide colour images only in the digital version of a publication, thereby reducing page costs. Participants in several cases wanted to associate supplementary material with publications: astrophysicists would like to be able to link research results to object and

image databases, and earth scientists felt that publishing the large datasets that underwrite many scientific papers in their field would increase transparency and move the field forward. Some fields, such as particle physics and some areas of astrophysics, have very long lists of authors and need to develop systems that can identify the specific contributions made by each author.

In short, this study suggests that **publishers must understand their customers, not just at a disciplinary level (such as physics or chemistry), but also at a sub-disciplinary level, which recognises the differences between fields when it comes to information needs and practices.** By focusing on the information landscape of their target audience(s), publishers can build tools that meet the specific needs of scientists. In some cases, this may be as simple as portals that allow access to shared back-end content via the channels used within a field. In other cases, it may mean that new tools are built or new methods of access such as APIs are opened to allow integration with other key resources for a field. In the most extreme case, users would be allowed complete flexibility in finding, accessing, and linking to information and data via tools and platforms developed by publishers, users, and third-parties.

Libraries

Libraries are not perceived as vital resources in the physical sciences. Few participants have visited their bricks-and-mortar libraries in recent years, as most of the important resources have been made available digitally. But many of these online resources are being delivered to users via library subscriptions. Ironically, libraries in many cases have been so successful at making this process seamless to on-campus users that few even realise that the library is responsible for their access until they try to use the resources while away from their campus and discover that they are unable to do so. Thus, the challenge for libraries is to find ways to be perceived as important and relevant brokers of information, while continuing to make this brokering function almost completely invisible on campus. They also need to make the process of accessing paid-for content easier for off-campus users.

The need for librarians to reinvent their roles as partners in the scientific and research process is acute. The future of librarianship, and how librarians' roles should evolve, has been a central concern of many professional library associations, particularly over the past decade. This study suggests that librarians need to be flexible when it comes to engaging scientists and researchers, so that they can tap into field-specific needs

rather than asking researchers to conform to librarian expectations. There are some examples of success: for example, fields within chemistry are engaging with librarians who have expertise in metadata to help them build specialised chemical databases. These opportunities, where library professionals become scientific consultants that can advise on information practices and policies in scientific collaborations, are one way for libraries to remain central to the research process.

Learned societies and professional bodies

One of the important roles of learned societies and professional bodies is to support the professional communities of practice through which disciplinary norms are learned and perpetuated. Conferences, newsletters, journals, training opportunities, websites, and other forms of communication all support this process. Learned societies and professional bodies can identify cutting-edge information discovery and management strategies in use by a minority of their members, and communicate those techniques to their communities. Ample evidence shows that professionals learn most effectively from the example of their peers, and so **opportunities for training should focus on linking experts using new approaches with their peers** in the same domain to demonstrate how these approaches can support their work.

Stakeholder cooperation

While each of the stakeholders listed above is important, there is a **pressing need for these stakeholders to work together**, with each other and with scientific communities, to build better, more effective, and more useful information practices in the physical sciences. As we have seen in this report, publishers and libraries can better serve science not just by talking to one another about subscription models and dissemination tools, but by engaging with funding bodies and professional bodies to help deliver the training needed by students and working scientists to improve their information practices. Scientists need to work with the other stakeholders to ensure that tools (and the training to use them) are suited to the needs, practices, and cultures of different scientific fields. Funders, publishers, librarians and learned societies must think radically and creatively, and work together to deliver the best information products and services to the practising scientist. Each of these stakeholders will have to consider how their current roles should be reworked and redefined to meet the needs of an emerging information ecosystem. But failure to work together will almost certainly result in some actors being excluded as their existing roles become irrelevant.

Next steps

This study demonstrated that information practices in the physical sciences vary not just by discipline, but also by sub-discipline and field. We must not view the information practices of researchers using broad-brush caricatures of the white-coated laboratory life-scientist, the lone humanities scholar labouring in the dusty archive, or the physical scientist seated at his or her computer crunching numbers. Not only are these caricatures often inaccurate, more importantly they mask the huge variety of activities taking place which contribute to the world's storehouse of knowledge.

There are several next steps that we could take. First, some major areas of research information practices such as the social sciences and the arts have not yet been studied using this lens; doing so would add to our understanding of the kinds of sub-disciplinary differences across all areas of research.

Second, this study and its predecessors have been small-scale studies, with an inherent bias towards the United Kingdom. Increasing the scale and scope would help us to understand differences at the national and regional level. For instance, by comparing similar fields in countries or regions with very different funding mechanisms, access to published information, and training regimes, we could begin to understand the extent to which scientific cultures transcend

national boundaries, and how much they are influenced by local policies and infrastructures. Larger-scale studies could also be used to test whether the patterns identified in this research hold true among larger samples of the scientific population.

Third, research that focuses on the processes by which funders, publishers, libraries, and professional bodies engage with each other and with domain scientists and researchers will help us to understand the steps that have already been taken to enhance the information landscape, and may suggest new ways to build more effective information ecosystems. Cross-stakeholder studies of this sort will be invaluable for shaping the strategies of all the stakeholders moving forward.

Finally, novel methods for understanding information uses by researchers such as measuring readership via publishers' access files, links via webometrics, and emerging areas of research via text analysis are being developed by a number of research groups. These potentially-promising areas of research all use a variety of large-scale metrics and measurements to understand how information is created and used and can, along with additional qualitative research, help us to understand the big picture of information practices in a digital world.



Endnotes

- 1 The perception that the literature is large is borne out in previous studies, which reported that “astronomers in the gamma ray burst community have been shown to have higher h-indices than their colleagues in other areas, partly because of the importance of the field, but also because of the size of the collaborations and the rate of publications” (Grothkopf, Melo, Erdmann, Kaufer, & Leibundgut, 2007, p. 66).
- 2 See Newman (2000) for a discussion of this. Several examples can be seen at <http://arxiv.org/abs/1111.2472> (with 498 authors), <http://arxiv.org/abs/1011.4079> (with 770 authors), and <http://arxiv.org/abs/1001.0165> (with 666 authors).
- 3 The Daresbury Nuclear Structure Facility was closed in 1993 following funding cuts. See R.C. Johnson’s 2008 public talk discussing the closure at www.ph.surrey.ac.uk/UserFiles/File/RCJSurrey50thTalk.doc
- 4 Cardamone, et al. (2009) describe the discovery of “green pea” galaxies; Lintott, et al. (2009) describe “Hanny’s Voorwerp,” discovered by Galaxy Zoo citizen scientist Hanny van Arkel.
- 5 Note that this is total number of classifications, not valid classifications after data processing. Cleaning the data removes approximately 4% of the total. For details of this processing see Lintott et al. (2008).
- 6 Online journals may or may not be peer-reviewed.
- 7 Time is a third underlying dimension to the diagram, although not pictured here for the sake of clarity of illustration.

References cited

- Attwood, T. K., Kell, D. B., McDermott, P., Marsh, J., Pettifer, S. R., & Thorne, D. (2009). Calling International Rescue: knowledge lost in literature and data landslide! *Biochemical Journal*, 424, 317-333. doi: Doi 10.1042/Bj20091474
- Baehr, S., Vedachalam, A., Borne, K. D., & Sponseller, D. (2010). *Data Mining the Galaxy Zoo Mergers*. Paper presented at the 2010 Conference on Intelligent Data Understanding.
- Berners-Lee, T. (1996). WWW: past, present, and future. *Computer*, 29(10), 69-77.
- Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, 57(13), 1758-1770.
- Bulger, M., Meyer, E. T., de la Flor, G., Terras, M., Wyatt, S., Jirotko, M., et al. (2011). *Reinventing Research? Information practices in the humanities. A Research Information Network Report*. London: Research Information Network. Retrieved from <http://www.rin.ac.uk/humanities-case-studies>.
- Cardamone, C., Schawinski, K., Sarzi, M., Bamford, S. P., Bennert, N., Urry, C., et al. (2009). Galaxy Zoo Green Peas: discovery of a class of compact extremely star forming galaxies. *Monthly Notices of the Royal Astronomical Society*, 399(3), 1191-1205.
- Chompalov, I., Genuth, J., & Shrum, W. (2002). The organization of scientific collaborations. *Research Policy*, 31(5), 749-767. doi: 10.1016/S0048-7333(01)00145-7
- Cummings, J. N. & Kiesler, S. (2005). Collaborative Research Across Disciplinary and Organizational Boundaries. *Social Studies of Science*, 35(5), 703-722.
- Davis, P. M. (2004). Information-seeking behavior of chemists: A transaction log analysis of referral URLs. *Journal of the American Society for Information Science and Technology*, 55(4), 326.
- Eccles, K. (2010). *Digital History Workshop: Connecting researchers to digital collections (JISC Final Report: Workshops & Seminars: Achievements & Challenges in Digitisation & e-Content)*. London: JISC. Retrieved from <http://www.jisc.ac.uk/media/documents/programmes/digitisation/oifinal.doc>.
- Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science*, 321(5887), 395-399. doi: 10.1126/science.1150473
- Fry, J., Schroeder, R., & den Besten, M. (2009). Open Science in e-Science: Contingency or Policy? *Journal of Documentation*, 65(1), 6-32. doi: 10.1108/00220410910926103
- Gentil-Beccot, A., Mele, S., Holtkamp, A., O'Connell, H. B., & Brooks, T. C. (2009). Information Resources in High-Energy Physics: Surveying the Present Landscape and Charting the Future Course. *Journal of the American Society for Information Science and Technology*, 60(1), 150-160. doi: 10.1002/asi.20944
- Ginsparg, P. (1994). First steps towards electronic research communication. *Computers in Physics*, 8(4), 390-396.
- Goldschmidt-Clermont, L. (1965/2002). Communication Patterns in High-Energy Physics *HEP Libraries Webzine* (6). Retrieved from <http://library.web.cern.ch/library/Webzine/6/papers/1/>
- Grothkopf, U., Melo, C., Erdmann, C., Kaufer, A., & Leibundgut, B. (2007). Using the h-index to Explore the Scientific Impact of the VLT. *The Messenger*, 128, 62.
- Haines, L. L., Light, J., O'Malley, D., & Delwiche, F. A. (2010). Information-seeking behavior of basic science researchers: implications for library services. *Journal of the Medical Library Association*, 98(1), 73-81. doi: Doi 10.3163/1536-5050.98.1.019
- Haque, A. U. & Ginsparg, P. (2009). Positional Effects on Citation and Readership in arXiv. *Journal of the American Society for Information Science and Technology*, 60(11), 2203-2218. doi: Doi 10.1002/Asi.21166
- Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S., & King, C. J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Berkeley, CA: Center for Studies in Higher Education. Retrieved from <http://www.escholarship.org/uc/item/15x7385g/>.
- Hemminger, B. M., Lu, D., Vaughan, K. T. L., & Adams, S. J. (2007). Information Seeking Behavior of Academic Scientists. *Journal of the American Society for Information Science and Technology*, 58(14), 2205-2225.
- Ion, S. (2009). *EPSRC/STFC Review of Nuclear Physics and Nuclear Engineering*. EPSRC and Science and Technology Facilities Council. Retrieved from www.stfc.ac.uk/resources/pdf/reviewnpne.pdf.
- Knorr Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
- Kyriakidou-Zacharoudiou, A. (2011). *Distributed development of large-scale distributed systems: The case of the particle physics Grid*. PhD thesis, London School of Economics and Political Science, London.
- Lamb, R. & Kling, R. (2003). Reconceptualizing Users and Social Actors in Information Systems Research. *MIS Quarterly*, 27(2), 197-235.
- Lintott, C. J., Schawinski, K., Keel, W., Van Arkel, H., Bennert, N., Edmondson, E., et al. (2009). Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society*, 399(1), 129-140.

- Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., et al. (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 389(3), 1179-1189. doi: 10.1111/j.1365-2966.2008.13689.x
- MacGregor, I. J. D. (2011). Ernest Rutherford his genius shaped our modern world. *Europhysics News*, 42(5), 18-21.
- Meyer, E. T. (2009). Moving from small science to big science: Social and organizational impediments to large scale data sharing. In N. Jankowski (Ed.), *E-Research: Transformation in Scholarly Practice*. New York: Routledge.
- Meyer, E. T. & Dutton, W. H. (2009). Top-Down e-Infrastructure Meets Bottom-Up Research Innovation: The Social Shaping of e-Research. *Prometheus*, 27(3), 239-250.
- Meyer, E. T., Eccles, K., Thelwall, M., & Madsen, C. (2009). Final Report to JISC on the Usage and Impact Study of JISC-funded Phase 1 Digitisation Projects & the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR). Retrieved from http://microsites.oii.ox.ac.uk/tidsr/system/files/TIDSR_FinalReport_20July2009.pdf
- Moed, H. F. (2007). The effect of "Open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047-2054. doi: Doi 10.1002/Asi.20663
- Nentwich, M. (2003). *Cyberscience: Research in the Age of the Internet*. Vienna: Austrian Academy of Sciences Press.
- Nentwich, M. (2008). *Cyberscience: The Age of Digitized Collaboration?* In G. M. Olson, A. Zimmerman & N. Bos (Eds.), *Scientific Collaboration on the Internet* Cambridge, MA: The MIT Press.
- Newman, M. E. J. (2000). Scientific collaboration networks. *Physical Review E*, 64, 016132.
- Nicholas, D., Huntington, P., Jamali, H. R., & Dobrowolski, T. (2007). Characterising and evaluating information seeking behaviour in a digital environment: Spotlight on the 'bouncer'. *Information Processing & Management*, 43(4), 1085-1102.
- Nicholas, D., Rowlands, I., Huntington, P., Jamali, H. R., & Salazar, P. H. (2010). Diversity in the e-journal use and information-seeking behaviour of UK researchers. *Journal of Documentation*, 66(3), 409-433. doi: 10.1108/00220411011038476
- Palmer, C. L. (2001). *Work at the boundaries of science: Information and the interdisciplinary research process*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Palmer, C. L. & Neumann, L. J. (2002). The information work of interdisciplinary humanities scholars: exploration and translation. *The Library Quarterly*, 72(1), 85-117.
- Pearce, S. E. & Venters, W. (2012). *How Particle Physicists Constructed the World's Largest Grid: A case study in participative cultures*. In A. Delwiche & J. Henderson (Eds.), *The Routledge Handbook of Participatory Cultures*. New York: Routledge.
- Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., et al. (2010). Galaxy Zoo: exploring the motivations of citizen science volunteers. *Astronomy Education Review*, 9.
- RIN. (2006). *Researchers and discovery services: Behaviour, perceptions and needs*. London: Research Information Network. Retrieved from <http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/researchers-and-discovery-services-behaviour-perc>.
- RIN. (2010). *E-journals: Their use, value and impact (Final report)*. London: Research Information Network. Retrieved from <http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/e-journals-their-use-value-and-impact>.
- RIN & British Library. (2009). *Patterns of information use and exchange: Case studies of researchers in the life sciences*. London: Research Information Network. Retrieved from <http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie>.
- RIN & NESTA. (2009). *Open to All? Case Studies of Openness in Research*. London: Research Information Network & National Endowment for Science, Technology and the Arts.
- Rogers, E. (2003). *Diffusion of Innovations (5th ed.)*. New York: Free Press.
- Shiri, A. (2011). Revealing Interdisciplinarity in Nanoscience and Technology Queries: A Transaction Log Analysis Approach. *Knowledge Organization*, 38(2), 135-153.
- Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, 41, 643-681.
- Tenopir, C. & King, D. W. (2008). Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns. *DLib Magazine*, 14(11/12).
- Walker, J. S. (2005). Recent Literature on Truman's Atomic Bomb Decision: A Search for Middle Ground*. *Diplomatic History*, 29(2), 311-334. doi: 10.1111/j.1467-7709.2005.00476.x
- Whitley, R. (2000). *The Intellectual and Social Organization of the Sciences (2nd ed.)*. Oxford: Oxford University Press.
- Zheng, Y., Venters, W., & Cornford, T. (2011). Agility, Paradox and Organizational Improvisation: The Development of a Particle Physics Grid. *Information Systems Journal*.

Full acknowledgements

Thanks to the Research Information Network (RIN), the Institute of Physics (IoP), and the Royal Astronomical Society (RAS) for funding this work.

Other project members provided editorial comments and support during the research process. The full list of project members is as follows:

Oxford Internet Institute, University of Oxford:

Eric T. Meyer (PI), Monica Bulger, Lucy Power

Information Systems and Innovation Group,
Department of Management at the London School
of Economics and LSE Enterprise:

Will Venters (Co-PI), Avgousta Kyriakidou

Oxford e-Research Centre, University of Oxford:

Marina Jirotko (Co-PI), Annamaria Carusi

Centre for Digital Humanities and the Department
of Information Studies, University College London:

Melissa Terras (Co-PI), Peter Williams

Virtual Knowledge Studio for the Humanities and Social
Sciences, Royal Netherlands Academy of Arts and Sciences
and Maastricht University:

Sally Wyatt (Co-PI), Andrea Scharnhorst

The project's advisory board was also key to helping source
appropriate disciplines and projects to investigate, and
offered important comments on the project's findings:

Kim Clube, Tajinder Panesor, Joe McEntee, Sean McWhinnie

Also thanks to Ellen Collins and Michael Jubb from the
Research Information Network, for funding this project
and their expert help and advice.

We are also deeply grateful to all of the participants in the
project, who generously agreed to help us understand their
information practices:

Atea Akraiam

Department of Physics and Astronomy, University of Leicester

Sitki Aktas

Department of Physics and Astronomy, University of Leicester

Melanie Auker

Department of Earth Sciences, University of Bristol

Steve Baker

Department of Physics and Astronomy, University of Leicester

Michael Barlow

Department of Physics and Astronomy, University College London

Cristina Oropeza Barrera

Department of Physics and Astronomy, University of Glasgow

Chris Binns

Department of Physics and Astronomy, University of Leicester

Sarah Bridle

Department of Physics and Astronomy, University College London

Philip Brohan

The Met Office, Hadley Centre

Jim Brooke

School of Physics, University of Bristol

Alison Bruce

Nuclear Physics Research Group, University of Brighton

Matilde Castaneira

Department of Physics, University of London, Queen Mary

Paul Crowther

Department of Physics and Astronomy, University of Sheffield

Melvyn Davies

Lund Observatory, Lund University

Phil Denby

EnSol, Norway

Martin Dove

Department of Earth Sciences, University of Cambridge

Tony Doyle

School of Physics and Astronomy, University of Glasgow

Phil Evans

Department of Physics and Astronomy, University of Leicester

James Ferrando

Department of Physics and Astronomy, University of Glasgow

Daniela Flocco

The Centre for Polar Observation and Modeling

UCL Department of Earth Sciences

Jonathan Flynn

Department of Physics and Astronomy, University of Southampton

Sean Freeman

School of Physics and Astronomy, University of Manchester

Martin Freer

Nuclear Physics Research Group, University of Birmingham

Mitko K. Gaidarov

Institute for Nuclear Research and Nuclear Energy

Bulgarian Academy of Sciences

Neil Gehrels

Astroparticle Physics Laboratory

Goddard Space Flight Center, NASA

Carole Goble

School of Computer Science, University of Manchester

Saskia Goes

Department of Earth Science and Engineering

Imperial College London

Andrea Gottardo

INFN LNL, University of Padova

Anne Green
School of Physics and Astronomy, University of Nottingham

Georgina Hart
Department of Chemistry, University of Oxford

James Hinton
Department of Physics and Astronomy, University of Leicester

Hannah Hogben
Department of Physical and Theoretical Chemistry
University of Oxford

Paul Howes
Department of Physics and Astronomy, University of Leicester

Mike Kendall
Department of Earth Sciences, University of Bristol

Sarah Kendrew
Max Planck Institute for Astronomy, Heidelberg

Ilya Kuprov
Oxford e-Research Centre, University of Oxford

Evipides Kyriakides
Department of Electrical Engineering, University of Cyprus

Andrew Levan
Department of Physics, University of Warwick

Chris Lintott
Oxford Astrophysics, University of Oxford

Ben Lishman
Department of Earth Sciences, University of Bristol

Owen Littlejohns
Department of Physics and Astronomy, University of Leicester

Steve Lloyd
Department of Physics, University of London, Queen Mary

Douglas Macgregor
Department of Physics and Astronomy, University of Glasgow

Karen Masters
Institute of Cosmology and Gravitation, University of Portsmouth

Anastassios Mavrokefalos
Department of Mechanical Engineering
Massachusetts Institute of Technology

Mervyn Miles
School of Physics, University of Bristol

Phillip Moriarty
School of Physics & Astronomy, University of Nottingham

Steve Murray
Department of Earth Sciences, University of Bristol

Dave Newbold
School of Physics, University of Bristol

Edwin Nissen
Department of Earth Sciences, University of Cambridge

Yvonne Nobis
Central Science Library, University of Cambridge

Paul O'Brien
Department of Physics and Astronomy, University of Leicester

Melanie O'Sullivan
Department of Chemistry, University of Oxford

Chong Qi
Royal Institute of Technology, Sweden

Muhammed Qureshi
Department of Physics and Astronomy, University of Leicester

Hardeep Rai
Department of Chemistry, University of Oxford

Akshat Rathi
Department of Chemistry, University of Oxford

Stephen Rhead
Department of Physics and Astronomy, University of Leicester

Nathan Rose
Department of Chemistry, University of Oxford

Mervyn Roy
Department of Physics and Astronomy, University of Leicester

Meg Schwamb
Department of Astronomy, Yale University

Alice Sheppard
Galaxy Zoo

Rob Simpson
Oxford Astrophysics, University of Oxford

John F. Smith
Nuclear Physics Research Group, University of the West of Scotland

Arfon Smith
The Adler Planetarium

Paul Soler
Department of Physics and Astronomy, University of Glasgow

Ben Still
Department of Physics, University of London, Queen Mary

Julian Stirling
School of Physics and Astronomy, University of Nottingham

Adam Sweetman
School of Physics and Astronomy, University of Nottingham

Nial Tanvir
Department of Physics and Astronomy, University of Leicester

Helen Thomas
Department of Earth Sciences, University of Bristol

Kim Tibbets-Harlow
Department of Physics and Astronomy, University of Leicester

Klaus von Haeften
Department of Physics and Astronomy, University of Leicester

Philip Walker
Department of Physics, University of Surrey

Matthew Watson
Department of Earth Sciences, University of Bristol

Daniel Watts
Department of Physics, University of Edinburgh

David Weaver
School of Physics and Astronomy, University of Birmingham

Callum West
Department of Chemistry, University of Oxford

Klaas Wiersema
Department of Physics and Astronomy, University of Leicester

Ian Wright
Planetary and Space Sciences Research Institute, Open University

Jake Yorke
Department of Chemistry, University of Oxford

This document is licensed under a Creative Commons Attribution-Non-Commercial-Share Alike 2.0 UK: England & Wales License

It is available to download at www.rin.ac.uk/phys-sci-case or further hard copies can be ordered via contact@rin.ac.uk

Design and concept by designisGoodland.com



