SUMMARY REPORT

Collaborative yet independent: Information practices in the physical sciences

DECEMBER 2011

research information network

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

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 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.

This short summary provides an overview of the key findings and recommendations of the study. The full report can be accessed at **www.rin.ac.uk/phys-sci-case**

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.

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.

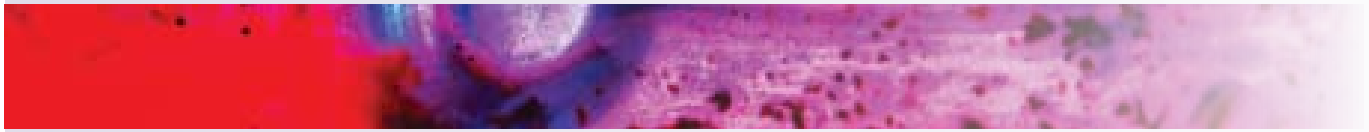
FIELD: Nuclear physics

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.

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.



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.

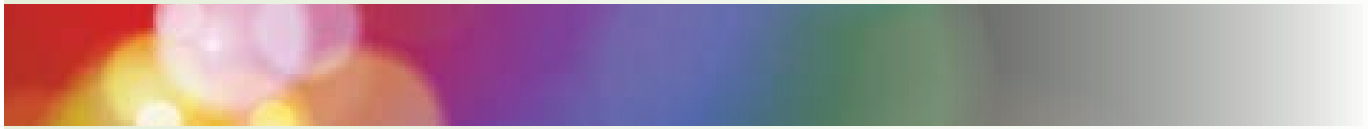
INTERDISCIPLINARY FIELD: Earth science

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.

Conclusions

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 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.

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.

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.

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.

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.

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.

The following resources are mentioned in this report:

ADS: Astrophysical Data System	http://adsabs.harvard.edu/index.html
arXiv	http://www.arxiv.org
arXiv astro-ph	http://arxiv.org/archive/astro-ph
CERN CDS (CERN Document Server)	http://weblib.cern.ch/
CERN INDICO (INtegrated DIgital COnference)	http://indico.cern.ch
National Nuclear Data Center	http://www.nndc.bnl.gov
SciFinder	https://scifinder.cas.org
SPIRES: Stanford Public Information Retrieval System	http://slac.stanford.edu/spires
T2K (Tokai to Kamioka) experiment	http://jnusrv01.kek.jp/public/t2k
Zooniverse	http://www.zooniverse.org

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