Open Science Project:
Final Report

Commissioned by the JISC Research Communications Strategy Project at the Centre for Research Communications – University of Nottingham

Sarah Currier
June 2011

Centre for Research Communications
crc.nottingham.ac.uk

JISC

Table of Contents

0. Executive Summary

0.1 Overview of project

In 2011, the Centre for Research Communications (CRC) carried out a brief snap-shot consultation on open science and citizen science, as part of the JISC-funded Research Communications Strategy (RCS) project, in response to the JISC report *Open Science at Web-Scale* (Lyon, 2009). This consultation, the RCS Open Science Project, ran from April 18th to May 31st, and carried out interviews with seven UK-based open science and citizen science practitioners and advocates, whom it asked for strategic and funding recommendations.

0.2 Overview of open science and citizen science

The RIN / NESTA report *Open to All? Case Studies of Openness in Research* (2010) defined open science as "science carried out and communicated in a manner which allows others to contribute, collaborate and add to the research effort, with all kinds of data, results and protocols made freely available at different stages of the research process". Open science and citizen science sit on "a continuum, which is helpful in positioning the range of behaviours and practices observed in different disciplines and contexts" (Lyon, 2009). This continuum has, at one end, open access journals of peer-reviewed research papers, and at the other, open notebook science, where every thought, step, protocol and method is recorded live on the Web at the same time as the researcher is thinking and doing.

Citizen science is related to open science in two senses: it requires openness of scientific data and publications in order to function; and it represents the ultimate in openness of participation. The JISC report (ibid.) defines citizen science as that where "volunteers who may not have scientific training, perform or manage research-related tasks."

Disciplines outwith the hard sciences are becoming increasingly involved in both open and citizen science, including the humanities and social sciences. The phrases *open research* and *open knowledge* are more inclusive terms sometimes used to acknowledge this.

0.3 Benefits

The putative overarching benefit of openness in access and participation is return on investment, not just for funders and researchers, but for the public who ultimately fund and benefit from research. Predicted positive outcomes go well beyond that measurable by citations of research papers in well-regarded scholarly journals. They include: improving the efficiency of research; promoting scholarly rigour; improving the quality and scope of scientific results; increasing the pace of discovery; improving scientific engagement and literacy for the public; and increasing the economic and social impact of research, including providing new opportunities for innovation.

0.4 Barriers, risks and constraints

Most barriers to wider uptake of open and citizen science lie in the current rewards culture within academia. University funding, research grants and individuals' careers are ultimately tied to publication and citation, which are themselves tied to the publishing industry and its constraints on open sharing. Open and citizen science are still niche activities found in a few areas of science. Lack of openness means that vast swathes of knowledge are inadequately exploited.

There are other constraints which prevent moving forward. A wider constituency could potentially be convinced to take more risks with openness if the these issues were addressed, even without solving the core problem noted above. Concerns include: fears around misinterpretation of data, wrong data being released too early, or being 'scooped'; a lack of usable tools to ease data sharing and curation; a need for new skills, roles and career structures; and more broadly, a lack of a sustainable technical infrastructure not reliant on short-term project-based funding.

0.5 Strategic and funding recommendations

1. For those organisations that do not have a strategy in place, the first recommendation is to decide what their policy on open science is, and to develop and enforce a strategy in support of that policy. At this stage, requiring open access publication of research papers, and open publication of supporting datasets, are the most important steps.
1a. For those funding bodies and other organisations who already require open publication of datasets alongside research papers, ensuring these requirements are enforced is essential.
2. If funding bodies and other organisations are requiring open publication of scientific data, appropriate infrastructure must also be funded. Several participants in this study recommended a return to the idea of sustainably funded national services for storing, serving, manipulating and preserving data. Other ideas include regional consortia or discipline-level provision. Most universities don't have the economies of scale to meet requirements for skills and infrastructure.
3. Continued support for and expanded provision of open source tools for use in laboratories, and within and across research teams. Usability, integration into existing software ecosystems, and support services are essential for uptake.
4. Build on current efforts to examine alternative academic reward structures, including rewarding open sharing and reuse of data, methods etc. alongside citation of formally published papers. Once people's careers are no longer totally dependent on citations of published papers in a few high-status journals, open sharing will become more attractive.
5. For citizen science in particular, but also for more developed areas of open science, provision, at national, regional or discipline level, of advocacy and support services for project development will save on time, resource and expensive mistakes.
6. Develop strategic funding provision in support of appropriate roles and career structures for scientific informaticians, scientific data curation professionals, and Web and software developers with open and citizen science specific interests and skills.

1. Overview of Project

In 2011, the Centre for Research Communications (CRC) carried out a brief snap-shot consultation on open science and citizen science, as part of the JISC-funded Research Communications Strategy (RCS) project. This consultation, the RCS Open Science Project, ran from April 18th to May 31st.

The intention of the RCS Open Science Project was to build upon some of the consultative questions raised in the JISC report *Open Science at Web-Scale: Optimising Participation and Predictive Potential* (Lyon, 2009) (hereafter referred to as the *JISC report*), by investigating the strategic and funding implications of current developments in open science and citizen science, with a particular interest in their relationship with research communications more generally. This report is the major output of the project, alongside a briefing document published as part of the RCS's *Current Issues in Research Communications* series.

1.1 Terms of reference

The RCS Open Science Project was necessarily limited in scope. Both the *JISC report*, and the joint RIN / NESTA document *Open to All? Case Studies of Openness in Research* (2010) (hereafter referred to as the *RIN / NESTA report*) cover current definitions, justifications, barriers and potential solutions for the domain of open science, and in the former publication, citizen science. Both are recommended reading for anyone interested in a thorough overview of these topics; this report does not attempt to cover the same ground in any depth.

The consultative questions raised in the *JISC report* are directed not just at scientists and other researchers, but also to the academic publishing industry, funding bodies, government, learned societies, and higher education and other research institutions; essentially, all those with a stake in science. The *RIN / NESTA report* includes some detail on the case against openness in certain disciplines, notably one case study (of six) where openness was not preferred by the researchers involved. In contrast, this study focused primarily on in-depth interviews with a small number of UK-based open science and citizen science practitioners and advocates. It therefore starts from an inherent appreciation of the potential benefits to be found in open or citizen science.

In addition, during the brief timespan of the project, open science and citizen science started to make their presence felt ever more strongly outwith the more radical reaches of higher education and scientific research, with a number of television programmes, articles and news items in the mainstream media. In this period the EPSRC[1] and the Royal Society[2] also made major announcements on open science and open data, and a number of conferences were held and broadcast via social media on the Web, including the *EC Public Hearing on Access to and Preservation of Scientific Data*[3]. While it has been impossible to fully incorporate and analyse such an upsurge in material, it is hoped that this report will contribute to the discussions growing apace in academia and beyond.

1.2 Questions addressed from *Open Science at Web-Scale*

This project asked seven open science and citizen science practitioners and advocates for detailed opinions, based on their personal philosophies and experiences, around a semi-structured interview plan (see ***Appendices***). The project specification started with three high-level questions, which had already to some extent been covered by the JISC and RIN / NESTA reports:

- who is using open science methods and what do they see as the benefits?
- what are the drawbacks/issues involved?
- what are the implications of open science practice for the development of scholarly communication?

Themes, discussion points and digressions arose organically; it is in retrospect that we can note which consultative questions from the *JISC report* were touched upon:

[1] EPSRC Policy on Access to Research Outputs
http://www.epsrc.ac.uk/about/infoaccess/Pages/roaccess.aspx

[2] The Royal Society launches study on openness in science http://royalsociety.org/news/Royal-Society-launches-study-on-openness-in-science/

[3] http://www.epsiplus.net/news/news/join_the_hearing_on_scientific_information_in_europe_30_may

1.2.1 Open science

"Should research funding bodies be pro-actively supporting open science principles and practice? What are the policy implications? What infrastructure is required?"

- Given the cohort of participants, the answer to the first part of this question is inevitably "yes". Funding, policy and infrastructure questions were all addressed.

"What are the community views on Open Notebook Science? Should these radical methods be migrated across to other disciplines and if so, which other disciplines would benefit? What key ONS development and enhancement issues need to be addressed?"

- Cameron Neylon was the one strong advocate and current practitioner of open notebook science interviewed; Rufus Pollock had some experience with 'open notebook social science'; Carole Goble has created software tools to support some open notebook practices. In contrast, Jason Swedlow, while supportive in principle, viewed open notebook science as more of possible future activity in the context of his research.

"Should open science practices be formally recognised and rewarded as intrinsic elements of scholarly communications? How can this be best achieved?"

- Again, there was a universal "yes" amongst the cohort, with sometimes contrasting ideas on how this could be achieved.

"What are the views of the research community on appropriate incentives and reward structures for data sharing, data re-use and wider participation?"

- This question was discussed in depth by all interviewees.

"How can research funding agencies best support data informatics skills development?"

- The answer to this specific question was somewhat outwith the expertise of this cohort, however, all were in agreement with the necessity for appropriately skilled informatics and other information and computer science professionals in ensuring adequate uptake and scaling of open and citizen science.

1.2.2 Citizen science

All the respondents, not just the citizen science specific ones (Steven Bamford and Chris Lintott) were strongly in favour of citizen science, saw it as inextricably linked in philosophy and goals with open science, and had things to say on both of the following two questions.

"What are scientist and funder attitudes towards citizen science? What are the societal implications? What role should research funding bodies play?"

"What are the short, medium and long term strategic and policy implications on science practice and outcomes, of a more openly participative research approach which may proactively include the public?"

These two questions were touched on more deeply by Steven Bamford and Chris Lintott:

"What are the financial implications, both in terms of direct and indirect costs, investment in infrastructure and associated benefits? What are the risks? What is the impact on research quality (data, models, outcomes)?"

1.3 Methods

The project was carried out primarily by the report author, who interviewed seven high-profile expert practitioners and advocates of citizen science and open science. Social media were engaged by the author and the RCS team at Nottingham University to promote discussion and information sharing, including: discussion on the JISC-Repositories, Research-DataMan and OER-Discuss JISCmail lists; postings on the CRC's blog; and updates and links shared via a project Twitter account and Netvibes page (see section ***7. Links and References: Other Links***).

Six of the seven interviews were captured on video, with the seventh conducted, by telephone, with Chris Lintott of Galaxy Zoo and The Zooniverse, who was working on secondment in the US. Clips of the videos were posted on the CRC's YouTube channel and blog to inform and promote discussion. In the spirit of open science, the video clips were released under a CC-BY licence, and the interviewees were also given access to the interview videos for their own use.

1.4 Interviewees

The project was lucky to get a good range of coverage at short notice, in terms of the spectrum of openness, in the small number of interviews carried out. Participants were:

1.4.1 Open Science Interviewees

Prof. Peter Murray-Rust: A chemist based at Cambridge University; a developer of open source chemistry informatics software; and an advocate for the most open end of the spectrum. He is a co-founder of the Open Knowledge Foundation, which supports opening up all data with the least restrictive licences possible.

Dr. Rufus Pollock: Also a co-founder of the Open Knowledge Foundation, Rufus is an economist at Cambridge University and past practitioner of open economics, which he also terms "open notebook social science"; like Peter he advocates for maximum openness throughout the research lifecycle.

Prof. Carole Goble: A computer scientist based at Manchester University who has spent a number of years heading initiatives creating open source tools to support open science in several disciplines. She is interested in how sustainability and curation of data and software can be assured. She is also an advocate of the crucial importance of usability of tools for open science, and centring development on the user experience.

Prof. Jason Swedlow: A biologist working in the Wellcome Trust Centre Centre for Gene Regulation and Expression, Jason got his start thinking about openness by developing a much-needed open source tool for managing cell microscopy images, the Open Microscopy Environment. His experience in the current research culture and environment puts him at the 'openly release your data *with* your published paper' end of the openness spectrum; for him, in his discipline, complete openness is an ideal but still a way off being practicable.

Cameron Neylon: A biophysicist and well-known advocate and practitioner of open notebook science based at the Science and Technology Facilities Council, Cameron believes in, practices, and preaches making the entire research process open to the world at the same time as you make it open to yourself and your research team.

1.4.2 Citizen Science Interviewees

Dr. Steven Bamford and ***Dr. Chris Lintott*** are astronomers with much valuable citizen science experience, largely via their involvement in successful large-scale citizen science projects Galaxy Zoo and The Zooniverse. They are also active in encouraging the use of citizen science in any research area where it may benefit the quality of the research produced (including the humanities and social sciences). Along the way they have seen successes and failures, and have developed an understanding of what makes citizen science a success, which they disseminate via the Citizen Science Alliance.

2. Definitions and Relationships

Both open science and citizen science have been discussed and defined extensively elsewhere, however, it is worth reiterating a brief definition for each, expanded somewhat by comments from interviewees. The philosophy behind both practices plays an important role in informing how advocates and practitioners view and promote their causes. It is interesting to note that several of the participants originally stumbled upon open science and citizen science as pragmatic solutions to problems they were encountering; their theory and philosophy grew from there. It is also worth looking briefly at the relationship between open science and citizen science.

2.1 Open science defined

Thinking about science generally and the way the scientific method has worked over several centuries, with scientists testing, confirming, challenging, and building on the work of other scientists, it may be a truism that "If it ain't open, it ain't science" (Coles, 2011).

> "Open Science broadly describes science carried out and communicated in a manner which allows others to contribute, collaborate and add to the research effort, with all kinds of data, results and protocols made freely available at different stages of the research process. Proponents of the approach argue that such collaboration will lead to more efficient research and innovation." -- *RIN / NESTA report* (2010).

It is also important to note that open science and indeed citizen science exist on "a continuum, which is helpful in positioning the range of behaviours and practices observed in different disciplines and contexts" (Lyon, 2009). This continuum may have at one end open access journals of peer-reviewed research papers, and at the other, open notebook science, where every thought, step, protocol and method is recorded live on the Web at the same time as the researcher is thinking and doing.

> "It's kind of fundamental to the conception of science, that we build on other people's work, we learn from others, we then give back to this kind of community, but also to the modern digital era, that open data and open information is what enables a much more distributed, collaborative, decentralised model of innovation, of culture, and of research than was possible before, much in the way that we've been able to build a very distributed peers-based commons production of software for example." -- Rufus Pollock, Interview.

> "I tend to define it as being the movement of people who are interested in improving access to the underlying pieces of science, of the research process. So it's about, it includes things like open access, open notebook science, open process, open publication, but it's bigger than all those things and as a result of that it's very difficult to describe it as a single thing, but as a group of people who are working to open up access to different parts of the research process." -- Cameron Neylon, Clip 1
> http://www.youtube.com/watch?v=U8Psf_4w9XQ

Disciplines outwith the hard sciences are becoming increasingly involved in both open and citizen science, including the humanities and social sciences. The phrases *open research* and *open knowledge* are more inclusive terms sometimes used to acknowledge this.

2.1.1 Open notebook science defined

The *JISC report* and the *RIN / NESTA report* both mention open notebook science, referring to it as "more radical" and "fringe" in the former, and giving detailed case studies of both acceptance of and resistance to its practice in the latter.

Not all interviewees in this study were using open notebook science; the words of advocate and practitioner Cameron Neylon, in defining and describing it, gives some idea why it is still a minority practice even within the open science movement:

> "Open notebook science was a term coined by Jean-Claude Bradley [...], and he has a nice sort of pitch line that tries to explain what the philosophy is, which is "There's no insider information" - the record that exists is what's made available, and that's the end of the story.

> "What that means in practice is that you have a lab notebook that's some sort of Web-based system. Jean-Claude Bradley happens to uses a wiki, we use a system that's based on a blog, others have used PDF documents that they post to the web. The mechanisms doesn't matter but it has to be something that can go on the Web, and for it to be pure open notebook science, in the core sense by which we mean it, it has to be the primary record, made available to the public at the same time as you make it available to yourself.

> "There are various variants of pulling back from that, where people might not make everything available; there are various reasons why that might be the case, from concerns about intellectual property, to concerns about privacy or other personal information; there are people who make an effort to record everything but don't release it immediately, but release it after some form of delay. And those are all things that we see as moving in the right direction but not the full blown approach.

> "It's quite a scary thing to do. I mean you're putting the raw record of the research you're doing online, complete with all the mistakes you make, all the ideas that you haven't thought through yet, and arguably putting up stuff that you haven't thought through yet, and perhaps isn't ready for public consumption. And so when I talk to people about this it's very simple to describe what it is: lab notebook, online, as it's recorded. The rationale behind that people find a little bit more difficult to take". -- Cameron Neylon, Clip 1: http://www.youtube.com/watch?v=U8Psf_4w9XQ

Jason Swedlow works at the opposite end of the spectrum from Cameron; he is in one of the more competitive disciplines to engage in open science. While he is very much supportive of the philosophy of open notebook science, and stated that he "would love" (Swedlow, Clip 1) if his students could get feedback from and interact with colleagues from all over the world, the current funding and publication mechanisms and career structures in his field mitigate against being able to.

2.2 Citizen science defined

The *JISC report* defines citizen science as that where "volunteers who may not have scientific training, perform or manage research-related tasks such as observation, measurement or computation".

Both citizen science interviewees in this study were involved in the prominent success story Galaxy Zoo, and have subsequently taken forward their learning from this into The Zooniverse and the Citizen Science Alliance, in an effort to expand the reach of citizen science, while maintaining its quality and integrity. So, while the usual broad definition as given in the *JISC report* includes citizens offering their computers' CPUs for large-scale data processing, Steven Bamford and Chris Lintott have added three refinements to create a more precise definition: (a) having active participation from volunteers (rather than paid staff or only borrowing people's CPUs); to (b) generate science that's of high enough quality to publish in peer-reviewed journals; and that (c) couldn't be done in any way other than citizen science.

> "The first is that, you're making use of members of the public. That should be a voluntary effort so if you're employing people to do a project for you, then that's not really citizen science [...]. The people should have an active role. [...] So by active involvement of the participants, I mean that [...] the scientific project actually needs their abilities, rather than their computer's abilities.
>
> "Another aspect that I think is important is that the science [...] should ultimately be of professional quality. It should be of the same quality as being done in universities, and be of a high enough standard to be published in respected journals. So, that doesn't mean that the individual participants within a citizen science project have to have anything like that kind of expert knowledge, it just means that when you aggregate their contributions together, you end up with something that's really worthwhile and is benefiting science as a whole.
>
> "Then the final part is making sure that there was a point to doing that science through citizen science. So you shouldn't have a situation where you're doing something through citizen science, you're involving the public in something that you could've done a better job of without involving them. So if you have an automated method that works better than even

> a professional, an expert, could do something, then you shouldn't be asking members of the public to do it if your automated method would do a better job.
>
> "So: you should be having active participation from the volunteers; you should be generating real science that's of professional quality; and you should be using citizen science because that is the best way to do your science, not because it's currently fashionable." -- Steven Bamford, Clip 1: http://www.youtube.com/watch?v=XjstOkw7OdM

2.3 What is the relationship between citizen science and open science?

The *JISC report* notes "twin aspects" of open science: access and participation, both having relevance for both open and citizen science. Citizen science is one way of providing openness of participation, and it requires scientific data to be open, both so that participants can perform tasks with it, and to ensure that the volunteers who give their time can see the results.

> "In order for us to create a project, it's necessary for the data to be public. So Galaxy Zoo for example used data from the Sloan Digital Sky Survey, and Sloan made a very sensible decision to put all of their data freely accessible online without reserving it for collaboration members, so we were able to piggyback off what they've done. I guess there's also some overlap in the sense that many of our volunteers go on to be quite advanced users of research, so we certainly have people trying to access the literature, and volunteers inspired by what they've done accessing the literature, reading up on the field, and we're building a whole suite of tools to try and make that interface with professional data and professional literature more easy for them, and that clearly, that model clearly doesn't work if I have to buy a journal subscription for everybody." -- Chris Lintott, Interview [audio transcript].

Citizen scientists sometimes also run their own analyses, models or experiments with the data and have been known to provide something unique enough to be named authors on a resulting research paper.

> "There is a similar philosophy behind citizen science and open science, and that is the idea that anyone should be able to find out about the scientific process or contribute to the scientific process. So in citizen science the idea is that anyone can be part of the process in that they are involved in data collection or data analysis, or even in some cases might actually be involved in the further stages, helping with publications and things." -- Steven Bamford, Clip 5: http://www.youtube.com/watch?v=nMvgzdCfrJk

Because citizen science is partially concerned with educating the public, it's not surprising that there is an interest in making scientific results available to the public more broadly:

> "And that extends further to the more general philosophy that science is for everyone, that people should be able to see the results, or the details of results that scientists are telling them. For example, one of the bugbears for me and I know for some other scientists is that if you read in the news or BBC online or something an article about a scientific result, and there's no link to the paper, so it's very difficult to actually verify that.

> "[...] Having that culture of transparency, and it being perfectly acceptable to inspect other people's working, I think would benefit not only science, but the public's perception of science, and just the public's understanding of evidence itself, so even in government policy, the idea that you can pose a question and research it and come up with an answer, and you don't have to accept that answer, you should be able to see how they came up with that answer. I think that encouraging that in the population is a useful thing, so I think educating the public, or increasing the scientific literacy of the public in terms of getting them involved in science, and allowing them to see the products of science, through citizen science and open science is a very good and strongly related idea". -- Steven Bamford, Clip 5: http://www.youtube.com/watch?v=nMvgzdCfrJk

The open notebook science approach is another way of providing openness of participation, the value of which is predicated largely on the value of enabling "serendipitous collaboration". For Cameron Neylon, the unexpected contribution is the whole point of openness, and it doesn't have to come from a professional researcher. He is also a strong advocate of the public's right to know not just the results of scientific research but to see the methods; to have access to the process as well.

2.4 What is the relationship between open and citizen science, and scholarly communications more generally?

Scholarly communications are essentially about the sharing and creation of knowledge. To date the primary area of openness promoted within scholarly communications has been the open access movement, advocating for access to pre-print and peer reviewed research papers via new publishing models and open access repositories. In this way scholarly communications could be seen to sit comfortably on the spectrum of open science.

However, the traditional scholarly communications set-up has much in it that prevents openness; see the ***Barriers, Risks and Constraints*** section further on in this report. There are some moves to encourage the open publication of the data underlying research papers, but even that is still confined to a few disciplines and journals.

Some new issues have emerged within citizen science around including non-professional 'citizen scientists' as co-authors, or otherwise acknowledging their contribution to science:

> "We've thought a lot about ways to give better credit to citizen scientists, so at the minute we stick their names at the end of the paper, and if individuals have made a huge contribution they've gone in as authors because they've made a discovery or run their own independent explanation or whatever.

> "One thing that we're working on is providing a space for them to formally, and in very inverted commas, publish. So if you think about what that means, it probably means I don't want to try and convince a large number of people that they should write stuff that looks like scientific papers, but I probably want a place to collect their paragraph-sized descriptions of what they've done and what they've found. A typical example: "I identified, I saw there was a small round green galaxy in the background of my image. I used the tools provided to measure its colours. I did a database search and got these galaxies which share the

following properties." People are doing that sort of level of research, but at the minute they're discussing it in a very informal space or emailing us and then [...] we write it up as a scientific paper.

"Whereas it would be nice to have a space where they can publish with their name next to it, and that piece of discovery can then be credited to them. And I think that ties nicely with the kind of things that open science people are talking about like micro-citations, and about carrying credit at the level of the data, or all of that stuff. So I think there's an interesting discussion to be had there. I don't think we've made any progress on that yet, and I don't think that there's an obvious system that we can jump into." -- Chris Lintott, Interview transcript (audio).

2.5 Is there a relationship between open and citizen science and other open movements?

2.5.1 The open source software movement

Several of the interviewees were initially inspired by their exposure to open source software; others picked that up along the way; one of the main benefits of open source software is not just the ease of sharing and tweaking software for new use cases and purposes, but also the inspiration and wisdom the methods and practices of open source contribute. For that reason, all of the software tools developed or used by the interviewees are open source.

2.5.2 The open education movement

The open education movement is also to some extent (sometimes problematically) inspired by the open source software movement, and shares many of the putative benefits and barriers with open science and citizen science (the latter perhaps being analogous to learner-generated content). It's possible that open education and open science could learn from each other as well as from open source. There are some specific affordances to be gained by open education from open science as well:

2.5.2.1 Open science angle

The *JISC report* notes a relevant benefit of open science in education: "7. Enhanced opportunities for student learning from open sharing of experimental methods and results data."

As an example, in 2009 Mimas gathered case studies on the use of real data (open or otherwise) in teaching and learning in higher education, mainly at undergraduate level, with some discussion of Masters-level teaching. Known as the *Real World: Real Data: Real Stories* project, one case study utilised open data available via ESDS International, and its particular relevance and usefulness in teaching economics, particularly econometrics[4].

2.5.2.2 Citizen science angle

Citizen science has its own connection to open education: the usefulness of opening up research data so that citizens can be engaged in tasks they find interesting also contributes to the enhancement of scientific literacy for the wider population. Citizen science is an excellent medium for open learning.

[4] http://www.esds.ac.uk/international/casestudies/real-data/real-data.asp

3. Benefits and Justifications

> "Our world will need science to help save it. Science is not the only thing that will save the world, but in the areas of disease, climate, energy, the creation of democracy, all sorts of things like this, what is absolutely key is that knowledge is shared as much as possible, that it's there on demand, that machines can access it and interpret it and so on." -- Peter Murray-Rust, Interview.

3.1 Give the people what they want: the social and political context

As noted previously, this study started from the presumption that open science and citizen science are worthwhile endeavours; benefits and justifications have been covered amply elsewhere. However, it's worth putting the current situation in a wider societal and technological context, in the words of Cameron Neylon. At the government level:

> "What's been interesting and to a certain extent embarrassing is the extent to which the government agenda on data has gone way ahead of what the scientific community has been prepared to take on."

And at the level of the general public:

> "There's certainly a sense in which society's expectations of the transparency of information have jumped a long way. I think the East Anglia climate email hacking is probably one of the strongest indications of that. [...] You talk to people on a train, on a bus about this, and what they were deeply shocked about was that the data was not available. That's the thing that they find incomprehensible. When you see scandals erupting around clinical tests of drugs where the data has been buried; that's what shocks people. **And the simple fact is that the public assumption now is that data that is relevant to the public is public, and when it's not, they're not impressed. And if the research community doesn't grapple with that, and doesn't deal with that fact, that we need to be delivering on that, we're in serious trouble**." -- Cameron Neylon, Clip 2: http://www.youtube.com/watch?v=6Rp1J9qS504

3.2 Benefits of open science

The *JISC and RIN / NESTA reports* provide thorough summaries of the putative benefits of open and citizen science. The primary justification noted by all this study's interviewees is return on investment, going well beyond that measurable by citations of research papers in well-regarded scholarly journals.

Factors whereby return on investment may be improved include: improving the efficiency of research; promoting scholarly rigour; improving the quality and scope of scientific results; increasing the pace of discovery; improving scientific engagement and literacy for the public; and increasing the economic and social impact of research, including providing new opportunities for innovation.

How the interviewees in this study approach their fellow researchers to interest them in open science is instructive:

3.2.1 Pitching open science to the research community

Each of the interviewees was asked how they would briefly pitch open science to a fellow researcher. Here are some of their answers:

> "[T]here are several aspects to why we should be doing things openly; they range from the ethical, through the moral to the utilitarian. So if we take the ethical arguments: we are funded by the public purse in large part; we do also get funding from other organisations but primarily from the public purse, and from charities, and I think there is a moral imperative to get the results of that out to the community. It's ironic that people, particularly in the medical field, who collect money for medical research, give it to universities, and then the results are only available to people in universities and not to the people who've actually funded it. So that's the moral aspect.
>
> "From the utilitarian aspect, I think there is growing evidence that by making your results openly available it actually advantages you and your institution over people who don't make their results available. If, let us say, 10 times, 100 times more people have access to the results of your work, then you in the long run are going to benefit.." -- Peter Murray-Rust, Interview.
>
> "[O]pen social science is what, I think at least implicitly, we're doing by default, or as the assumed default in science: that we share knowledge, not maybe immediately, but we ultimately share knowledge, we build on other people's work, we collaborate, and that we're also in theory [...] in pursuit of some kind of truth or understanding about the world that's a shared enterprise, and one in which there aren't winners and losers, we're going forward together in some sense. So I think that's implicit, the point I'd say about open science, strictly, that that conception of science goes back hundreds of years, thousands of years, right now the point about open science right now, is it's about making some of that explicit, it's saying in the modern digital age, in which unfortunately lots of people have asserted lots of exclusive rights. [...]
>
> "[Y]ou know you're gonna write publications you want other people to read; making your material open is just going exactly along with the grain of that. The thing I would then say is why aren't we doing that by default? Well not because you don't want to but because of currently the restrictions of the publishing set-up, and I think on data, the slight challenges of publishing data openly and easily." -- Rufus Pollock, Interview.

Carole Goble began by talking about how her own interest came out of her passion for reproducible science, and for scientific results being defendable even if not reproducible. In particular, the sharing of scientific methods was a gap that piqued her interest:

> "A lot of my work is to do with being able to encode and to share and to reuse and to make publicly exposed methods, protocols, techniques. [...] When you're doing science you have materials and methods and results. So we have publishing of results, we are beginning to have publishing of materials with data, say, but publishing of method has been particularly haphazard". Carole Goble, Clip 2: http://www.youtube.com/watch?v=HYebvi_ja2w

To pique the interest of other researchers, Carole suggests asking them how many times they have looked at a paper and wondered if they could reproduce results, but haven't been able to get hold of data or methods or sufficient detail. How many times have they wondered if a colleague has results they haven't used themselves, but would be useful? She points out that open science helps you get hold of results that are underused or not used at all. How many times has a PhD student done their research then left? Having a place to go to find all the research results, a commons, would overcome that problem.

She also talks about the attraction to being a provider of open science, not just a consumer. She notes that researchers can build their reputation, make their results more available, find new collaborators and gain some new insights.

She also mentions the broader benefit of "accelerating knowledge turns", knowledge turns being the speed of returning a hypothesis into a result, and back to a hypothesis. At present knowledge turns are too slow, because of lots of labs producing results independently and competitively. Eventually results emerge so other labs can take results and use them; some results never emerge, because labs hug them to themselves; both result in very slow progress. So if people expose results earlier, accelerate knowledge turns. "That is the biggest driver to open science, the acceleration of knowledge turns". Carole Goble Clip 2: http://www.youtube.com/watch?v=HYebvi_ja2w

Jason Swedlow simply says that "it's the most powerful way to collaborate I've ever seen" (Clip 4). He notes that it is critically important for other people be able to see your data, e.g. in the genomics field, the data producers are great scientists who work very hard, but sometimes notable insights have been missed that external views were able to glean. He also reports that some would say it's better to just reproduce the experiment, so why should they access your data? His response is that most of time in the biological sciences, the data is based on very expensive equipment, processes and expertise; it's very difficult to regenerate all that. If, for instance, somebody wanted to regenerate one of his papers, they would have to acquire four years of experience, cell lines, reagents, all the know-how. He would be happy to drop that on someone else and say 'you do it too', but the first thing that should be done is that data should be available so other people can analyse it and decide if it's necessary to redo it.

3.2.1.1 Pitching open notebook science

Cameron Neylon talked about how to pitch open notebook science to those open to taking on its risks:

> "My view is kind of twofold or maybe threefold. One is that by exposing the working of your lab notebook, it's a powerful motivator to keep a better record, because people might be looking at it. It's also a great way of improving the record, the standards of recording within a group, because you know you can rely on each other to have that information, and again there's that motivation of the group to keep a better record.
>
> "As you step up from that level, then you have the opportunity for serendipitous collaboration, so coming across people by accident, by Google searches, these kind of things, and those are the kind of things you can't predict, can't see in advance, and they're still rare [...]..

> "And then the other thing is that by doing this, by making some sort of commitment, then there is a community around this, and that community is supportive, and people are willing to help because they see that you're committed to an open process. [...]
>
> "So the really simple pitch is, if it's easy to do then you're not losing anything, in most cases, and what you're gaining is the stuff that's unpredictable, that you never would've guessed at, and that's where all the interesting science is, it's always in the bits that you didn't expect.
>
> "The key thing is opening yourself up to unexpected contributions. That's the real value." -- Cameron Neylon, Clip 1 http://www.youtube.com/watch?v=U8Psf_4w9XQ

3.2.2 Pitching open science to funding bodies, policy-makers and university managers

Each of the interviewees was asked how they would briefly pitch open science to those with responsibility for funding, policy and strategy around science and scholarly communications. The responses invariably centred around return on investment and value for money. Here are two representative answers:

> "Very simply, higher education is a trillion dollar industry, it's extremely badly serviced by its information infrastructure, we lose literally tens if not 100s of billions of dollars a year because we cannot manage our information properly. So it is up to the heads of the institutions, up to funders to make sure we do this properly and part of what is necessary is that the whole flow of information from the funder to the researcher to the rest of the world is made as seamless as possible. At the moment there are too many blocks in the way, so that there are people who are actually making money out of restricting the flow of information." -- Peter Murray-Rust, Interview.
>
> "[V]alue for money. In essence we fund public research in whatever discipline, from economics to genomics, we fund it in order to advance human understanding, and also for that material, not just so we know more, but also so technologies can be developed better, better treatments for everything from cancer to regulation of the economy, and the fact is when we don't have open data, and we don't have open information, we're basically kind of cutting off our leg, we are significantly harming ourselves in the efficiency by which researchers go about their business, so we're wasting money in the simplest sense." -- Rufus Pollock, Interview.

3.2.2.1 Pitching open notebook science

Cameron Neylon provided a similar pitch to those above, specifically for open notebook science:

> "The simple pitch is: our key need is to maximise the return on investment. That means effective communication, it means effective archival, and it means supporting the efficient management of information within laboratories themselves. Taking an open notebook approach or a partial open notebook approach delivers on every level. It makes material available even if it isn't eventually formally published. It maximises the archival possibilities at least in the cheapest way you can, by placing stuff on the web, where there are archival mechanisms, it's not perfect, but it's the cheapest way to make progress on that, and it

improves the management of data within a laboratory, by providing a centralised place with a social structure around that that helps to encourage best practice. So you're delivering the return on investment for effective capture of the research that you're investing in, at the laboratory level, at the archival level and at the communication level." -- Cameron Neylon, Clip 1: http://www.youtube.com/watch?v=U8Psf_4w9XQ

3.3 Benefits of Citizen Science

The Citizen Science Alliance notes[5] that the benefits of citizen science are: the ability to more fully utilise extremely large data sets; the ability to gather multiple independent interactions with the data, providing quantitative estimates of error; provision of large, powerful training sets for machine learning; and serendipitous discovery from exposing data to large numbers of users. A secondary, but still important benefit is that citizen science is a powerful tool for education about specific sciences, and about the scientific method itself.

3.3.1 Pitching citizen science to the research community

Steven Bamford talks about two of the immediate, practical benefits of citizen science, for researchers in universities - publications and helping research projects with workload:

"[Galaxy Zoo] produced a completely unique dataset, which I've been using in my work It's been immensely productive. I've had 16 papers at the latest counting in a few years from the Galaxy Zoo data, so it's been a success, and it would not have been possible to do that work with visual morphologies using the established method without citizen science. So that's a prime example of how citizen science can work. So that's why it's been particularly useful both for me, but also for my field, it's contributed that knowledge to the literature." -- Steven Bamford, Clip 2: http://www.youtube.com/watch?v=9-USO9a68mk

"If you have a situation in your department where one of your post-docs or one of your students is spending a lot of time doing a fairly repetitive, straightforward task, instead of focusing on the other areas of their research, then that's probably a prime candidate for trying out citizen science, and a place for citizen science to really change how that project works." -- Steven Bamford, Clip 1: http://www.youtube.com/watch?v=XjstOkw7OdM

Chris Lintott pitches the idea of citizen science from a broader perspective, common to many fields of endeavour today: too much data, not enough eyes (or ears):

"I think researchers in lots of fields are suffering from a common problem, which is that for the first time in a long while we actually often have too much data, so whether that, in astronomy, that suddenly instead of images of 30, I have surveys of millions of galaxies. In ecology you used to send your grad student into the field and they'd come back with a notebook full of observations about animal behaviour. Now you can carpet the forest with webcams, but you've still only got one grad student, who now has to review input from 200 webcams or whatever. It's that sort of jump in scale, is common to a lot of science right now and dealing with it's really hard.

[5] Why citizen science? http://www.citizensciencealliance.org/philosophy.html

> "One solution is obviously to become more technologically advanced, to rely on machine learning techniques and all of that stuff, but there are still places where you just need human eyeballs and human brains to ferret through the data and tell you what's interesting. And so the only solution is to scale the human response as well. We're not going to get funding for 10,000 more scientists or 10,000 more research assistants but we can reach those people over the Web, and so we go looking for projects where people are overwhelmed by the amount of data that they have, where they just, you know, people say, some poor sod has to look at all these images, sort through all these documents, read all of this, transcribe all of that, any of those problems could be distributed to a very willing audience." -- Chris Lintott, Interview transcript (audio).

3.3.2 Pitching citizen science to funding bodies, policy-makers and university managers

As with open science and open notebook science, efficiency and return on investment are the key attractions for funders, policymakers and universities:

> "You've got expert people who you're funding or employing, and you want to use their expertise as efficiently as possible, but you have this situation now [...] where you have lots of very talented researchers doing fairly repetitive, fairly simple tasks, and it's a waste of their ability, it's a waste of their time, and correspondingly it's a waste of funders' money. So citizen science presents the opportunity to remove that repetitive aspect, that workhorse aspect from the researchers' day to day job, and place that in the public domain, and use a large number of members of the public to handle that aspect of the science, and for them to focus on doing the job of coming up with the ideas and writing up the science, producing the high-level science that that process enables." -- Steven Bamford, Clip 2: http://www.youtube.com/watch?v=9-USO9a68mk

Chris Lintott focuses on the educational and outreach aspects of citizen science:

> "This turns out to be a remarkable way to educate people. There isn't a better way of knocking down the idea, that evil idea that science is something done by people in ivory towers paid for by shadowy organisations than by making it a truly collaborative exercise between people, between the public and scientists. Citizen science projects actually provide that, it provides that space in which meaningful collaboration is possible. Rather than standing on stage giving an outreach talk, talking about the wonderful things that I've done, it's absolutely incredible to be able stand up there and talk about what we've done together, the audience, the public and ourselves and that can be a very valuable piece in building support for science and for specific projects". -- Chris Lintott, Interview transcript (audio).

4. Risks, Barriers and Constraints

> "I think it's morally reprehensible to suppress knowledge that has been generated by public funds [...]. And I think it's also morally reprehensible that we have reward systems in universities that reward people for hugging their results rather than distributing their results,

and that's because there is no way of them being able to get credit for doing it, and particularly there's no way for them getting credit for curating their data." -- Carole Goble, Clip 3: http://www.youtube.com/watch?v=3BJASLHAYJ8

4.1 Risks, barriers and constraints in open science

Most barriers to wider uptake of open science lie in the current rewards culture within academia. University funding, research grants and individuals' careers are ultimately tied to publication and citation, which are themselves tied to the publishing industry and its constraints on open sharing. Open and citizen science are still niche activities found in a few areas of science. Lack of openness means that vast swathes of knowledge are inadequately exploited.

> "We're in a sort of Alice in Wonderland thing where the worth of a person is determined by the number of citations that they get. Some authors don't care if their papers aren't read so long as they're cited. And you can get to a stage where you increase your citations if you're a referee by demanding that the author add references to your papers, and that's again not uncommon that referees will say 'You haven't included my work, you should cite it' and so on. That doesn't mean that the author will read it but they have to put it in to get their paper accepted. Again that's not the norm but it's not unknown.

> "So, we have this striving for citations the whole time, which is not the primary purpose of science. The primary purpose of science is to explore nature, to build tools which will explore nature, to conquer abstract fields of thought such as mathematics and computer science and so on, and to communicate that to the world. And in my view it's perfectly reasonable that there is a delay during the thought process in this so that when you announce this to the world, it may depend on the discipline, how likely you are to get positive reactions from the people you release to and how much you're likely to get scooped, but the general purpose is to publish it, whereas at the moment most academics are only read by probably 10 to the minus 7 of the world, so almost everybody on the planet cannot read the results of academic research which is an appalling situation in today's world. -- Peter Murray-Rust, Interview.

There are other constraints which prevent moving forward. A wider constituency could potentially be convinced to take more risks with openness if the these issues were addressed, even without solving the core problem noted above. Concerns include: fears around misinterpretation of data, wrong data being released too early, or being scooped; a lack of usable tools to ease data sharing and curation; a need for new skills, roles and career structures; and more broadly, a lack of a sustainable technical infrastructure not reliant on short-term project-based funding.

4.1.1 Addressing concerns about putting data out there before publication

Cameron Neylon noted in his interview that one common fear that people have about publishing data in any form before peer review is the belief that by putting something on the Web, it won't be able to be published in a formal journal. While his opinion is that for the most part now this is untrue, Jason Swedlow stated in his interview that he has found that it is true in his domain, or at least, is enough of a risk to prevent pre-publication sharing.

There is also the fear of people taking ones data before publication and 'scooping' ones results. Again, where Jason Swedlow notes that this as a reasonably valid fear in his field, which is

particularly richly funded and competitive, the other open science respondents believed that it was not as likely as most people thought. In fact it was said by several interviewees that publishing something on the Web would make scooping less likely as you have date-stamped evidence of your work.

> "The second argument is that you may be scooped if you publish openly during your experiment. Actually this is unlikely to happen if the process is well carried out because it's then clear in public that you've done it. You're actually much more likely at the moment to be scooped by sending your paper to a journal, having a immoral referee who takes your results, rejects your paper for long enough that they can actually do their results in their own laboratory. And this is a not uncommon occurrence." -- Peter Murray-Rust, Interview.

In addition Cameron Neylon also noted that in his experience people who found data on the Web were likely to want to collaborate with the original providers and share credit.

Clearly advice in these areas should be tempered with some understanding of discipline cultures and their concomitant funding arrangements; targeted advocacy would be useful for separating myths from facts.

4.1.2 Addressing concerns about quality of data

Jason Swedlow notes:

> "There have been various experiments with more open publication: open peer review, open notebooks, and by a variety of measures they have not always produced the best work. That doesn't mean that they are fundamentally flawed, it means that we as scientists have to change the way we understand our practice. That's a huge huge change, a huge cultural change, and that will be a long time coming." -- Jason Swedlow, Clip 2: http://www.youtube.com/watch?v=r9SJ2DBNcIl

Carole Goble noted that concerns about quality of data, quality of curation, quality of metadata and so on, were oft-cited reasons for people not wanting to risk the extra labour that goes into openness. There are related fears that released data or methods will not be interpreted or understood correctly. Putting effort into providing raw and analysed data out there is important but time-consuming. She mentioned that one environmental science project even put a tutorial in between users and their data to ensure they understood enough of the scientific method to be allowed to see the data.

4.1.3 Fear of exposure

Despite science's basic remit to test and build upon the work of others, and to expose one's own work to rigorous review, there are underlying insecurities at play for many individuals that prevent willingness to 'go open'. These seem to be related to both fear of scooping and the pressures of the competitive rewards system currently in place. Cameron Neylon and Rufus Pollock talk about the all-too-human insecurity of scientists:

> "This is an underlying one that people don't admit: is the concern about being embarrassed. It's the 'I'm going to look like an idiot'. And that's a really big issue because (a) in my experience most researchers have a pretty severe inferiority complex, always worried about

comparing themselves to each other. The people who succeed are the ones who are seen to have never [made] a mis-step. So you're only allowed to admit that you've ever made a mistake once you get your Nobel prize basically. And so the notion that you'd expose your working is a real concern to a lot of people. They won't admit it in most cases. But that's really one of the main underlying reasons. But I think there's a positive counter to that: it does help you maintain a higher standard of reporting". -- Cameron Neylon, Clip 2: http://www.youtube.com/watch?v=6Rp1J9qS504

"There is, kind of in a sense, a default: why give out data that might help people prove you wrong? I mean, honestly, as scientists we should of course do that, but in people's head, there's that aspect of you know, why give out data, what am I getting out of it? I mean obviously in our kind of theoretical vision of what we should be as scientists we should never think that, we should always be thinking of course we should give out our stuff to be checked, of course we should give it out so that other people can build on my work, but that isn't, you know, we're human, after all." -- Rufus Pollock, Interview.

4.1.4 Can we address the rewards systems in academia?

All of the open science interviewees noted that the current rewards systems in academia are crushing attempts to encourage wide-spread openness. They also all pointed out that openness is most likely from those at the top of the career pyramid, i.e. visionary professors with tenure and nothing to lose, and those at the very bottom, who have a social Web outlook, and have not yet fully committed to the struggle up the pyramid. The squeezed middle, those trying to climb the ladder, are most at risk from the various pitfalls of bucking the current rewards system. Peter Murray-Rust and Jason Swedlow both spoke of their reluctance to ask younger colleagues to take such risks; change must come from a more strategic place:

"There are three or four reward systems, and it's quite clear that reward is absolutely critical in academia. I think it's obsessive at the moment, but both individuals are now very very conscious of their rewards and so are institutions. So, in the modern world we do this through metrics, metrics are imperfect and the ones that we use, which are citations for publications, are among the worst metrics that anybody has devised. They are controlled by monopolists outside the control of the universities, so your metrics, my metrics are done by commercial companies that do not publish their algorithms, and so on. So, nevertheless there is this rush towards metrics, trying to find out which the, quotes, best journals are to publish in, not because they are necessarily the most appropriate but because you get higher metrics from them.

"Now, that culture is now so deeply ingrained that it is difficult to see how to change it unless there's an external catastrophe within universities. So I cannot ask a young researcher to publish in a journal which has a lower impact factor because I think it's the right thing to do, I leave that decision to them. So the only people really who can buck the system are senior people who don't care, and that includes me, and young people who perhaps aren't going on for a career in science. But the people in the middle are desperately driven by the metrics. So the primary metrics are publications, or rather the citations in your publications, which can be several years after you've done the work and are shown to be extremely inaccurate." -- Peter Murray-Rust, Interview.

> "I think almost all the arguments I know of, which are protecting our students [...], basically our students need to be able get competitive papers, and those competitive papers depend on sometimes on keeping results quiet for some time so your competition doesn't [...] Why? Because our current publication structure prohibits me from publishing in a.. every paper must be a significant advance. I can't publish a paper in a prominent journal, and thus my student's name can't be on that paper, and he or she needs that paper for their career, if that paper doesn't exist they can't move forward, the only way I can get that paper is if the results are novel and distinct from anyone else. [...] If the data is open someone else can do the same analysis that we're doing and have the same insights that we have. Full stop." Jason Swedlow, Clip 2 http://www.youtube.com/watch?v=r9SJ2DBNciI

As well as individual career advancement based on citation metrics, other rewards systems include grant income and patents. The latter is often used as a solid economic reason to keep some data closed within universities, however, Carole Goble and Peter Murray-Rust have some things to say about this:

> "I'd say get a grip of your intellectual property regulations; universities are encouraged to suppress results on some vague idea they might make cash out of it later and this often is not the case, and we should be reversing this, we should be thinking about encouraging organisations to be able to much more freely make their results available by rethinking the whole way that we exploit intellectual property in universities. We should give intellectual property to the individual scientists and not to the university. [...] We would rather that it died, suffocated, we would rather suffocate the results than have somebody else benefit. Well how is that benefiting the greater good?" -- Carole Goble Clip 3 http://www.youtube.com/watch?v=3BJASLHAYJ8

> "And on a personal and possibly institutional level there are also patents. Now, probably one out of a hundred patents actually makes more money than it costs, that's just a figure off the top of my head, but almost all universities lose money through their technology transfer or licensing things, and it does great harm to the progress of science if you find that particular steps in the scientific chain have been patented. Some disciplines regard patents as, you know, as very desirable, things like some biosciences where you're coming up with new materials, new assays, materials science where you're making new materials with different properties and so on. Some disciplines such as computer science have I think outgrown this obsession with monetising the results and now make all their results openly available. And I think we will start to see this pass through as people show that patenting in universities is actually counterproductive." -- Peter Murray-Rust, Interview

4.1.5 Who pays?

While open science mavens talk about 'free as in speech, not free as in beer', Carole Goble points out that openness is also 'free as in puppies':

> "Somebody has to pay. In the end, it's not free. It's not, classically, it's free like puppies are free. You get a free puppy, and then you have to walk it, you have to feed it, it craps on the carpet, you have to clean that up, it gets sick, you have to look after it, and then it dies. That is just the same with institutional repositories, with open access publications; somebody

has to look after it: it is free as a free puppy." Carole Goble, Clip 5: http://www.youtube.com/watch?v=NIY0fMe9sUY

4.2 Risks, Barriers and Constraints in Citizen Science

Problematic areas for citizen science include: firstly and most commonly, concerns about the quality of science produced. Secondly, people involved in artificial intelligence or machine learning assuming that humans aren't, or very soon won't be needed for any task. There are also less-commonly expressed concerns about exploitation of volunteers; the sometimes unacknowledged need for extremely careful expert design of protocols, user interfaces and software, including scalability; and the failures of some projects due to misunderstanding by universities of what citizen science is for, i.e. if the focus is on PR and not science, then the project will fail, which is bad for the participants and public perception of citizen science and science generally.

Citizen science also shares some general barriers with open science, namely: the need for good tools; centralised advocacy and support services; and provision of technical skills within appropriate career structures and role allocation.

4.2.1 Addressing data quality

"The first concern that people have is the quality of the data, so as a professional researcher, you're used to having quite some pride in the work that you do, and the idea that anybody off the street could do something better is a bit threatening and also a bit implausible. So the first response is that you're not really going to be getting any useful data. So that's really a case where we need to educate the researchers, explain to them that you're not being so naive as to just be giving their job to random members of the public, but instead you're employing a method where those members of the public can contribute in a way that is useful. So you're isolating the elements of your problem that they can contribute to, collecting their contributions, and then there's still the extremely, very important job to be done by the professional scientists to take that information and turn it into science. And there's additional steps involved there in understanding your population, understanding any biases, and issues around that, so actually in some ways it becomes even more of an intellectual challenge to do, be able to do, science in that way, while taking away the more repetitive simple tasks that actually quite often take up the majority of a researchers time. -- Steven Bamford, Clip 3: http://www.youtube.com/watch?v=rcdedLf1i6w

4.2.2 Addressing the lure of the PR approach

"Really, for its success, even for its success as a public outreach tool, it has to always have science as its focus, it isn't something to just be used as a public outreach stunt." -- Steven Bamford, Clip 3: http://www.youtube.com/watch?v=rcdedLf1i6w

4.2.3 Addressing career paths, skills and roles

"The biggest problem about appointing people with these kind of skills as post-docs is that, you can do it, you can stretch the definitions of roles and find people who would fit them

> who are good at Web development, or the technical side of things, but tick enough boxes on the scientific side that you can appoint them as post-docs, but if what they're working on is the technical aspects, then during that post-doc they're not publishing, or certainly not as first author, they're not leading science, so that's essentially the end of their career on the academic path, so it then becomes much harder to appoint that person for another three years, because they've technically had a post-doc at that point but there's no evidence for what they've done in the literature. So establishing a different career path for those people is important, I think." -- Steven Bamford, Clip 7
> http://www.youtube.com/watch?v=LdyuV34C72k

5. Strategic and Funding Recommendations

Interviewees were asked for their priorities for policy, strategy and funding, at government, funding body and institutional levels. There was broad consensus, with a few disagreements as noted below.

5.1 Developing an open science strategy

For those organisations with a stake in research who do not have a strategy in place, the first recommendation is to decide what their policy on open science is, and to develop and enforce a strategy in support of that policy. This should include specific provision for citizen science.

At this stage, requiring open access publication of research papers, and open publication of supporting datasets, are the most important steps[6]. How these requirements dovetail with specific discipline community constraints should be investigated, but the primary movement should be towards a default of openness, with justifications for being closed in some areas, rather than the other way around.

Even the strongest advocates of open notebook science and other "fringe" approaches did not feel that mandating anything more at this stage in history would be beneficial. That doesn't mean that policy and strategy broadly supportive of investigating more radical approaches would be unwelcome.

5.1.1 Enforcing openness policies

A number of funding agencies already mandate some degree of open publication of papers and data. A number of universities and other organisations have some requirements for open publication, and/or are offering repositories or other services. For any organisation with openness policies, remaining vigilant on enforcement of these policies will be a great support to those researchers and projects who are taking the lead in sharing their research outputs. However, noting the difficulties of complying, and offering practical provision to overcome these difficulties, as noted below, is also recommended.

5.1.2 Citizen science policy and strategy

[6] It should be noted that participant Cameron Neylon disagreed with mandating openness by funding bodies, and preferred to encourage conversations about how best to maximise impact in given disciplines and domains.

The most important thing to remember for citizen science policy and strategy is that all citizen science activities must have good science at their core. Any outreach benefits must come from the science. Experience has shown that thinking of PR or marketing first does not result in good citizen science. Nor does avoiding the costs of commercial quality technical developments, including overall design, scalability and user interface in particular.

5.1.3 Policy and strategy at the government level

Peter Murray-Rust sums up the interviewees' feedback on government-level policy and strategy:

> "I would advise government that all public funding of research required that the results of that are publicly available. I think that's generally accepted but there are strong lobbies against it. About 5 years ago, Ian Gibson, MP for East Anglia, had a select committee in the House of Commons where he addressed the question of open access and he got compelling support from the Commons to make science publication open access, he was frustrated by not having information in the medical field. Unhappily that got bogged down at the Lords. So I would say that government has to be clear that it must work by making science information universally openly available. I think there will be an economic return to that because some countries are going to do that and the countries which develop the infrastructure to support that will also develop the wealth creation industries based on managing modern technical and scientific data so there must be a spin-off in terms of wealth creation for the countries that get there first. So I would say to the government: publish your science openly, support information creation and dissemination industries and you won't regret the investment." -- Peter Murray-Rust, Interview.

5.2 Providing appropriate and sustainable infrastructure

> "Strategically fund sustainable infrastructure that enables scientists to be open". -- Carole Goble, Clip 3: http://www.youtube.com/watch?v=3BJASLHAYJ8

If funding bodies and institutions require open publication of scientific data, appropriate infrastructure to allow such publication must also be funded. Storing, serving, manipulating, curating and preserving scientific data are non-trivial in terms of the expense and skills required, and individual institutions do not have the economies of scale needed to take this on on their own. Likewise, academic publishers are unlikely, at this stage, to fund such enterprises.

> "I would ask for a straightforward way for scientists to be able to at least deposit their results somewhere of all different kinds, just park them. So sort of a public commons or ways of getting hold of a public commons. I would ask funders to fund the proper management of data, and the proper management of its quality, the husbandry behind data and to respect that that is part of the scientific method and not to just cut that bit because it's computing and therefore doesn't matter.
>
> "So that would be to properly resource it, and to properly provide the resources so when they say there is a data management obligation, that there's a mechanism for people to be able to do that. So just saying it is not good enough." -- Carole Goble, Clip 3: http://www.youtube.com/watch?v=3BJASLHAYJ8

Scientific data produced from public funding, or produced in the public interest, should be publicly available; responsibility therefore resides with the ultimate funders and beneficiaries of the research.

Several participants in this study recommended a return to the idea of sustainably funded national services. Other ideas include regional consortia or discipline-level provision. This study does not offer a definitive solution; further investigation must be carried out. It is important to note that for most interviewees, "sustainably funded" means treating such services as long-term enterprises rather than as short-term research projects in their own right.

> "The reason people don't put things out may not be because they don't want those results to be available, it's because there's no mechanism for them to do it: there's no resource to put them in, there's no money allocated to tending them; preparing data for public access is incredibly expensive; that's the difference between something being available and something being usable." -- Carole Goble, Clip 3: http://www.youtube.com/watch?v=3BJASLHAYJ8

Peter Murray-Rust also put forward a strong case for developing open science infrastructure such that it will be compatible with the Semantic Web via linked open data; in his experience however, much of the potential held in this idea is currently blocked by the restrictive IPR policies of publishers:

> "We for example have developed a tool called OSCAR which is able to read chemical journals. And we could read 10 million reactions every year if we were allowed to, but the publishers explicitly forbid this and if I were actually to use this tool which would advantage science, then probably I would be sued, the university would be cut off subscriptions to these journals, and nobody would win out. The publishers don't win out because they're not doing it, they're simply stopping us doing something which is a reasonable thing to do because they think it protects their business model. So, I would like to see this swept away, there's an awful lot of modern informatics that cannot be done because of restrictive practices.
>
> "So, I think if you like at the first level I'd like to make all this information open, the term that Tim Berners-Lee uses is linked open data, I'd like to see every discipline contributing to the linked open data cloud. So that's the infrastructure that I would like. [...]
>
> "I would like to see that academia felt itself to be a major player in the Semantic Web so that academia saw the outputs in the linked open data cloud as beneficial, and indeed that might be one of the most achievable metrics that we could have, so if you got brownie points for having a linked open data set on the Semantic Web as a measure of your contribution to the world [...]." -- Peter Murray-Rust, Interview.

5.2.1 Supporting tools for researchers

Part of the technical infrastructure noted above is the provision of tools for knowledge management, communications, data manipulation and so forth, for use in laboratories, and within and across research teams. In some cases generic tools can be used or easily adapted across

disciplines; in some cases highly specialised discipline- or sub-discipline-specific requirements exist.

Participants in this study identified a number of open source tools, mainly ones that they themselves had had a hand in developing. Ongoing funding for several of these has been forthcoming already. Continued support for and expanded provision of such open source tools is recommended.

However, it is important to note that development, curation and user support of open source tools is also non-trivial. Carole Goble has found that Agile development methodologies centralising the user experience, multi-disciplinary teams, integration of tools into researchers' workflows, and provision of support services around tools is vital to their wide uptake and success. This can be expensive; her team develops and supports a small number of niche scientific tools, and she requires £1.5 million in funding each year to keep these working effectively.

5.2.2 Tools for citizen science

Tools for citizen science have some special requirements. They are usually built on existing products, so they are not usually funded as innovative computer science projects. In addition, they have particularly stringent requirements for usability, to ensure maximum participation from the public, and robustness and scalability to handle the load of data and participants.

5.2.3 Open bibliography as part of institutional infrastructure

Peter Murray-Rust has noted some particular infrastructure requirements around institutional bibliography that he believes would be extremely helpful in assisting with scholarly communications:

> "Now, if you had the open bibliography of all of the papers that had been published, you would have at a lag of no more than 6 months a complete map of who your academic staff were collaborating with, but you can't do that at the moment because bibliography, if you try and open up bibliography, you will be resisted by the publishers and you will be met with total apathy from the institutions. They don't realise that bibliography is actually part of the seed-corn for the wealth that they generate, and they give it away and then buy it back with appalling restrictions." -- Peter Murray-Rust, Interview.

5.3 Changing the research culture

> "I would ask institutions, funding councils, promotion bodies, publications, to respect other forms of scientific output other than a finished PDF in a peer reviewed and expensive journal, so, to understand that there are multiple ways that you can produce things, and those multiple ways to demand credit. And there should be mechanisms of credit, so that's a technical problem and also a social problem. [...] That's happening in the environmental sciences because of 'climategate' [...] Let's not lurch from crisis to crisis; we should really have this as kind of a principle." -- Carole Goble, Clip 3: http://www.youtube.com/watch?v=3BJASLHAYJ8

The idea of changing a culture of research and publication that has been entrenched for a very long time, and is supported by a number of powerful stakeholders, is daunting. However, there are a number of initiatives already afoot investigating possible alternatives to the strict mandate of: peer review - publish - citation - metrics - research assessment - funding - career advancement.

It is therefore recommended that funding bodies and other interested organisations continue to support current efforts to examine alternative academic rewards structures, particularly around sharing and reuse of scientific data, methods, results, etc., alongside citation of formally published papers. Once people's careers and employers are no longer dependent on citations of published papers in a few high-status journals, open sharing will become more attractive.

> "We in the United Kingdom are not accelerating our knowledge discovery fast enough, because we're not sharing fast enough, or at all, and we're never going to share until the cultural and credit environment of science and academia changes: so what are you going to do about that? [...] I would suggest that there is as great a credit for publishing your data as there is for publishing your papers, that you invest in attribution and reputation-protecting infrastructure, for publishing the results we're producing, so that your scientists can get some credit from this; you start enforcing the policies your funding agencies have invented but don't bother to enforce, you start a dialogue with the publishers so the publishers really understand that a dead piece of PDF is not the be all and end all of science." -- Carole Goble, Clip 3: http://www.youtube.com/watch?v=3BJASLHAYJ8

5.4 Advocacy and support services

There are a small number of high-profile examples of successful open science and citizen science in the UK. The people who have pioneered in these areas have gathered much experience about what does and doesn't work. They are already forming organisations, such as the Open Knowledge Foundation and the Citizen Science Alliance, to attempt to provide some organised advocacy and support for those who wish to move into open and citizen science. These organisations are pulled together from voluntary effort and various bits of funding and would be seriously stretched if there was any kind of upsurge in open and citizen science.

Citizen science in particular would benefit greatly from coordinated approaches to advocacy and support. The stakes are high in terms of public perception, when, for instance, participants number in their hundreds of thousands; not repeating past mistakes is vital.

Good advocacy for citizen science and open science needs to be supported with appropriate evidence. One kind of evidence, suggested by Peter Murray-Rust, is research into the economic benefits of openness, building on work previously done by Rufus Pollock:

> "I think that they actually need an economic analysis of the benefits of them regaining control of their scholarship and their scholarly information. Rufus Pollock, whom you'll talk to, did this for maps and showed that by opening up map information the net wealth to the community was greatly increased. I think that a good economic analysis of university income and benefits, measuring tangibles and intangibles as outputs in the university, would I hope show that there was a compelling economic case for doing that. So I think that we need metrics for intangibles, metrics other than papers, metrics other than grant income and so forth." -- Peter Murray-Rust, Interview.

5.5 New skills, roles, careers, professions

All the interviewees talked about new requirements for highly specialised skill sets to support open and citizen science. As leading edge practitioners, they have all attempted various methods to provide employment on their projects for people with the appropriate mix of scientific knowledge and technical skills. However, the emergence of new roles, and possibly new career structures and professions, was noted as something that was either under way or should be.

For example, employing someone with an appropriate PhD as a post-doc researcher to do a project's technical tasks, thereby denying them the chance to publish on the actual research (when perhaps they don't want an academic career anyway); then being unable to employ them again as a post-doc as they have no publication record, was cited several times as an unsustainable approach.

The final recommendation is therefore around development of strategic funding provision in support of appropriate roles and career structures for scientific informaticians, scientific data curation professionals, and Web and software developers with open and citizen science specific interests and skills.

It is unclear at present whether these roles should be funded within individual project grants; in existing university services such as libraries or technical support departments; in newly formed departmental services; shared across consortia of several universities within a given subject area; outsourced to private companies; or provided as national services. Investigation into such models is also recommended.

5.6 Funding citizen science

It's worthwhile to look at specific issues around funding citizen science that haven't been covered above. There is huge scope for citizen science to make a real contribution to human knowledge across many disciplines, but not much is being funded. Current funding tends to come from sources outwith the central research funding bodies, e.g. the Leverhulme Trust. This is partly an advocacy issue and partly an issue whereby citizen science doesn't fit easily within traditional scientific funding models. It is also related to the previously mentioned issue with funding new roles and finding new career structures:

> "It's actually really hard to get these projects funded in new areas. Once we've found with astronomy and now climate change transcription and various others, you can see the same process: once we develop a project and we can show what we mean and show that the results are good, then it gets adopted by a large part of the community within any area of academia. And we've been good at getting funding for education off the back of projects for advanced tools to build on projects, but it's been relatively hard to fund the actual core software team to build new projects. And that's because we're not computer science, because we're not using any particularly advanced technologies, we're using off-the-shelf commercial tools, very well I think, but we're not a computer science research project in that sense, but nor do we come in looking like a standard science proposal, because we tend to need money for developers and so on, and so there's a problem in infrastructure funding there from my point of view, which is compounded by the fact that there's no space within academia for people who can build tools, Web tools. At the minute most of my team pretend to be post-docs, because most of them have PhDs and they're at that level, but the question is where's their career structure? If I build a fantastic website that produces a

> catalogue that's used by 2000 astronomers, but I don't write the papers, then what do I do next? And so, finding a home for those people is a big problem." -- Chris Lintott, Interview (audio).

One of the reasons for this is the long funding cycles for grants within universities and other research institutions. Sometimes it only becomes clear during a research project that citizen science could make a huge difference:

> "At the moment it's quite common for a research project to have been funded, to be under way, and then they realise this would be a fantastic opportunity to use citizen science; it would make this project either have much more robust results because we'd have access to a larger set of data, or it would go faster, but they can't employ citizen science because they don't have money for that purpose. If they could go on a weekly, ideally, or a monthly timescale, go to a funding agency and say "We've got this situation, we've written up how that would help", and have that reviewed quickly and get the money on a timescale that would allow them to get that up and running on the same timescale on their current grant., that would obviously be a help. At the moment you've got the situation where because you'd have to spend a year to get that money, if you waited that year that PhD student would be a third further through their PhD." -- Steven Bamford, Clip 8: http://www.youtube.com/watch?v=RotxHwglL3s

6. Links and References

References:

Coles, P. (2011). *If it ain't open, it ain't science.* Blog post: http://telescoper.wordpress.com/2011/05/16/if-it-aint-open-it-aint-science/

Lyon, L. (2009). *Open science at Web-scale: optimising participation and predictive potential: consultative report.* JISC, DCC, UKOLN. Available: http://www.jisc.ac.uk/publications/reports/2009/opensciencerpt.aspx

Research Information Network and NESTA (2010). *Open to all?: case studies of openness in research: a joint RIN / NESTA report.* Available: *http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies*

Other links:

Participant quotes are from interviews which are openly available on the CRC YouTube channel: http://www.youtube.com/user/CRCNottingham

For resources discovered during the project's brief lifespan, see the report author's Diigo bookmarks: http://www.diigo.com/list/morageyrie/RCS-Open-Science-Project

For a wider, dynamically updated collection of resources related to the project, see the project Netvibes page: http://www.netvibes.com/rcs_open_science

For blog posts from the Research Communications Strategy team, see: http://rcsproject.wordpress.com/

For project Tweets, see: http://twitter.com/rcsopenscience

7. Acknowledgements

The author would like to thank the interviewees (and their administrative teams who helped set up our meetings) for their time, and of course their openness.

Admiration and gratitude is also due to the authors of the *JISC report* and the *RIN / NESTA report*, both of which are considerable achievements and made my job much easier.

Thank you to Mat Todd and Jean-Claude Bradley, open science advocates extraordinaire from beyond these shores, who both offered to be interviewed; so sorry there wasn't time.

And thank you to the JISC and the Centre for Research Communications for funding this timely investigation.

Appendices

Appendix 1: Interview Plan: Open Science

* Can you give your name and a brief summary of your own work in open science, for the video.

* What does ‘open science’ mean to you? How do you define it?

* How would you define it to a fellow researcher/academic whose interest you were trying to pique? How would you define it to a university manager, government policy-maker or funder, if you had their attention for 5 minutes?

* Why are you involved in open science? What do you see as the benefits, for you personally and professionally, for your discipline, for your academic institution and research community?

* What groups/disciplines/sub-disciplines are currently using open science methods? What do they see as the benefits?

* What groups/disciplines/sub-disciplines are not currently using open science methods, to your knowledge? Any ideas about differences between those areas and those who are using it?

* What arguments do people make against open science? How would you respond?

* How do you see the relationship between open science as a philosophy and practice, and scholarly communications more generally (e.g. research papers, conferences, journals; citation, reputation and rewards; open access to research publications (via repositories or otherwise)).

* How would you describe the relationship between open science and citizen science?

* What do you think would improve the practice of open science methods for your work and your research community? What do you think would improve the uptake in your discipline? What about more broadly?

* What would you like to do or achieve with open science that you aren’t or can’t currently achieve? What’s getting in the way, in your opinion?

* What would you like to see happen at government, institutional and funding body levels to improve the lot of open science? What would you ask for? What would you advise (and to whom?).

* If you could have 1 wish granted to improve the lot of open science (no matter how unreasonable it may seem right now) what would yours be?

Appendix 2: Interview Plan: Citizen Science

* Can you give your name and affiliation(s), and a summary of your own work in citizen science- just a few sentences, for the video.

* What does ‘citizen science’ mean to you? How do you define it?

* How would you define it to a fellow researcher/academic whose interest you were trying to pique? How would you define it to a university or research institution manager, government policy-maker or funder, if you had their attention for 5 minutes?

* Why are you involved in citizen science? What do you see as the benefits, for you personally and professionally, for your discipline, for your academic institution and research community?

* What groups/disciplines/sub-disciplines are currently using citizen science methods? What groups/disciplines/sub-disciplines are not currently using citizen science methods, to your knowledge? Any ideas about differences between those who are and aren't using it?

* Do people make arguments against citizen science? If so, what are they, and how do you respond?

* How do you see the relationship between citizen science and open science as philosophy and practice? What about their relationship with scholarly communications more generally (for instance: research papers, conferences, journals; citation, reputation and rewards; open access to research publications (via repositories or otherwise)).

* What would you like to do or achieve with citizen science that you aren't or can't currently achieve? What's getting in the way, in your opinion?

* What do you think would improve the practice of citizen science methods for your work and your research community? What do you think would improve the uptake in your discipline? What about more other disciplines?

* What would you like to see happen at government, institutional and funding body levels to improve the lot of citizen science? What would you ask for? What would you advise (and to whom?). NB: Focus is on UK context.

* If you could have 1 wish granted to improve the lot of citizen science (no matter how unreasonable it may seem right now) what would yours be?