Rebooting science? Implications of science 2.0 main trends for scientific method and research institutions

Author Details:

Corresponding author:

Dr Katarzyna Szkuta, European Commission, Joint Research Centre, Institute for Prospective Technological Studies (IPTS), C/ Inca Garcilaso No. 3, E-41092 Seville, Spain

Author
David Osimo, Open Evidence Ltd and Open University of Catalonia, ASSBE - Applied Social Science and Behavioural Economics Research Group, Barcelona, Spain

Acknowledgments (if applicable):

This article is based on the results of the study *Science 2.0: implications for European policies on research and innovation study. Scoping study to assist the ERAB board* (SiS 2011) conducted for the European Commission DG RTD and complemented by subsequent own research. The authors acknowledge also the work of Fabio Casati who conducted two interviews that are part of the research data and were used in the analysis.

Structured abstract:

The article aims to analyse a set of converging trends is underpinning a larger phenomenon called Science 2.0 and to assess what are the most important implications for scientific method and research institutions. It is based on a triangulation of exploratory methods which include a wide-ranging literature review, web-based mapping and in-depth interviews with stakeholders. It rejects the notion of science 2.0 as the mere adoption of web 2.0 technologies in science, and puts forward an original integrated definition covering three trends that have not yet been analysed together: open science, citizen science and data-intensive science. It argues that these trends are mutually reinforcing and and puts forward their main implications: enhanced efficiency, transparency and reliability, raise of data-driven science, microcontributions on a macroscale, multidimensional, immediate and multiform evaluation of science, disaggregation of the value chain of service providers for scientists, influx of multiple actors and the democratization of science. It concludes with the identification of three enablers of Science 2.0 – policy measures, individual practice of scientists and new infrastructure and services and sees the main bottleneck in lack of incentives on the individual level.

Keywords:

Science 2.0, web 2.0, open science, citizen science, data-intensive science

Article Classification:

Concept paper

1. Introduction

Historically, science has evolved in response to both internal and external factors. David (2004) shows how academic science was transformed to its modern 'open' form thanks to specific sponsorship relationships in the 17th century. It resulted in a creation of a first scientific journal the Philosophical Transactions in 1665 and enabled a rapid public disclosure of new knowledge. This system of producing accessible, reliable and cumulative stock of knowledge (Dasgupta and David 1994, Mukherjee and Stern 2008) has worked well for centuries. The new communication technologies enabled academic science to become even ‘more open'. The ICT tools made the disclosure faster and more efficient and added additional possibilities – sharing intermediate results as well as underlying data and code. The advent of the Internet transformed the journal business model first by changing it from a paper form to an electronic and secondly by the appearance of open access journals. This second transformation made science more available for all (that have access to the Internet) compared to the paywall-hidden publication in a journal.

Concurrently, another change could be observed, a transition from the so-called older Mode 1—characterized by hierarchical, homogeneous, and discipline-based work, to Mode 2 of research driven by complexity, non-linearity, heterogeneity, and transdisciplinarity of approach as well as incorporation of new stakeholders, namely citizens which are making the demarcation line of science and society less evident to draw (Nowotny et al, 2001).

In recent years, these, by and large, ICT-enabled changes have accelerated towards greater openness, greater collaboration, and greater adoption of data-driven approaches. Those trends are being defined and labelled as data-intensive science, citizen science, crowd science, open science or science 2.0. Both scientific research (Bartling, Frisieke 2014; Davis 2011, Murray-Rust 2008, Neylon and Wu 2009, David 2003, to name a few) and science policies recognised it as an important phenomenon (e.g. recent European Commission Consultation on Science 2.0[1]).

This range of interrelated changes points to an emerging and possibly systemic change in the scientific endeavour. But as for any emerging trend, it calls for a thorough assessment of the significance of the trends and their impact. This paper aims at delivering this assessment by bringing together the evidence available in the scientific literature and deriving implications for both research method and players in the scientific domain.

This article starts with the presentation of the research questions and the method. Subsequently in the section three it proposes an ostensible definition of Science 2.0

[1] http://ec.europa.eu/research/consultations/science-2.0/consultation_en.htm

and compares it to the similar concepts used in the literature. Section four focuses on the main trends of science 2.0 and assesses their importance. Section five discusses the main implications for the scientific institutions and for the scientific method based on the extrapolation of the existing trends. Section six concludes the article by putting forward key issues for further research.

2. Research questions and method

We aim to address some fundamental questions related to these new trends in the modus operandi of science.

Q1. What is the nature of this change? Can we provide a single definition of these new trends? Are they interrelated and self-reinforcing?
Q2. What is the importance of these trends? Are these new approaches widely available and used?
Q3.What are the implications, both positive and negative? What are the main effects of these changes on the science method and the scientific institutions?

We address these questions through a triangulation of exploratory methods. First we conduct a Wide-ranging literature review to identify evidence and data points on these trends (Q1-Q3) which is followed by a web-based mapping to identify and classify related initiatives (Q2) and complemented with ten in-depth interviews with stakeholders from diverse backgrounds (Q3).

The first research question attempts to provide the definitions of these trends - what are the different definitions existing in parallel and if they can provide a coherent framework to map the trends? The literature review results in an initial framework (three macro-trends), which is then populated by the initiatives gathered in the mapping phase. The trends are subsequently analysed to identify possible overlaps and synergies within them, as these can reinforce the justification for a new single definition. The mere parallelism in time between these trends without interrelations would reduce the need for a single definition.
The second research question aims at assessing the expected significance of these trends. As most new trends tend to be overestimated, a more robust assessment of their importance, beyond the fad of a new scientific revolution is being sought. Obviously, because of their emerging nature, little systematic evidence is available on its impact and take-up, and the data is typically fragmented in different domain-specific studies. Both the literature review and the mapping is used to identify possible evidence of its significance, such as:

- the availability of a large number of initiatives beyond the most known ones (i.e. above 100) and spread in different disciplines and phases of the research cycle
- the demonstrated significance of individual initiatives in terms of participation and added-value as compared to traditional methods of research

- the wide availability of aggregated services that enable such initiatives, beyond the level of individual projects
- the level of uptake of these trends by scientists

The third research question deals with the actual implication of these trends for the modus operandi of science. The emerging impacts, both positive and negative are identified. This work does not aim at quantifying them, but at understanding their nature and at relating them to the key challenges that scientific research and institutions are currently facing. It assesses to what extent the implications point to a new architecture of science or rather confirm and reinforce existing structures. The analysis of the implications is built mainly on the findings of the literature review and insights from the in-depth interviews.

The literature review is based on searches based on predefined keywords in Scopus (science 2.0, open science, citizen science, data science, crowd science) looking for those keywords both in titles, keywords and abstracts. This search has been complemented with collaborative references searches on Mendeley and expert recommendations (e.g. through questions in the scientific social networks and Quora.com). The final sample of articles consisted of 190 scholar references mainly from 2008-2015.

The second method used for identifying importance of trends is a large-scale mapping of cases. The mapping was carried out in one of the online bookmarking services[2] and was based on literature review and desk research. The case mapping resulted in 105 cases of science 2.0 projects. See the table 1 for the distribution of cases among the main trends and the scientific domains.

Finally, ten in-depth interviews validate and expand the preliminary list of the implications stemming from the literature review[3]. The list of the interviewees is provided in the appendix 1. The interview questions related mainly to the importance of the main trends, distribution across different science disciplines, implications of science 2.0 as well as to the existing policy framework. Each interview lasted on average around 30-40 minutes. The interviewees' sampling was done through the identification of influential players in the field with the help of the literature review and conferences proceedings on science 2.0 as well as the presence in online communities on science 2.0 and open science. The sample meant to be heterogeneous enough to enable for different stakeholders viewpoints and to validate the results of the literature review and the mapping exercise.

Our approach has of course its limitations. The sample was mostly limited to scholarly publications and included only a small proportion of grey literature (mostly governmental and government-sponsored reports on open access and open data). The

---

[2] groups.diigo.com/group/science-20

[3] All of the interviews took place between April and June 2012 via a phone interview (with the except for the interview with Cameron Neylon that was conducted via email). All but two interviews were conducted by one of the two authors. The interview with Cameron Neylon and Roberto Casati were conducted by Fabio Casati.

analysis does not take into account other products of scholarly communications e.g. blogs or online fora. Nevertheless, we tried to overcome those limitations by interviewing stakeholders active in this field – publishers, open science activists, citizen scientists. As far as the interviews are concerned the sample was a bit unbalanced as it targeted mostly the science 2.0, citizen science and open science scholars who are often also activists and therefore more prone to concentrate on positive implications of this systematic change. However, due to the breadth of the phenomenon under study both the case mapping and the interviews were used only to complement a more comprehensive literature review.

3. Defining science 2.0

Several definitions describe recent changes in science methods and rules governing scientific institutions. We have already mentioned Gibbons et al (1994) and Nowotny et al. (2001) who proposed the mode 2 of knowledge production definition. The term of data-intensive science refers to the exploding quantity of data available for analysis, commonly referred to as Big Data, leading to greater use of inductive methods and a so-called fourth paradigm of science (Hey et al., 2003). This new approach combines theory, experiments and simulations, and is often referred to as computational science since it adopts distinctive techniques and technologies. The definition of Big Data does not merely involve the use of very large data sets, but involves also a computational turn in thought and research (Burkholder 1992).
The citizen science or crowd science movement is deeply rooted in the history of scientific research. Amateur scientists have had an important role in scientific disciplines throughout history. For instance, the wide adoption of telescopes suddenly made it popular to watch the sky in the 19th century and led to the explosion of the amateur astronomer and enabled fundamental new discoveries (Hufbauer 2006). The mode of participation has changed however, with the availability of cheap electronics instruments and software that allows for a pervasive remote monitoring of the environment at an unprecedented scale. There are almost 250.000 amateur astronomers in the US only (Kannappan 2001). Collaborative web-based technologies allow for leveraging the individual capacities of thousands of people at once, exploiting the unique human capacity to recognize visual patterns or to report events. Many projects have been launched in the last years, such as FoldIt, Galaxyzoo, Stardust@home, Project Feederwatch, each of which has involved thousands of people. There are already new intermediary services such as SciStarter.com, which aggregates citizen science initiative to let citizens find the most suitable initiative.
In recent years, the term that attracted most attention is certainly “Open Science”. In particular, the public movement for open data and open access to scientific publication has become far more important in recent years and widely recognized also at policy level. Since the advent of web 2.0 this opening-up process has accelerated and extended well beyond open data and open access: scientist use blogs and social networks to collaborate, share ideas and raw material throughout the entire scientific process (Procter et al., 2010). Cooperation happens before papers are finalised: bibliographies, data, experiments, code, annotations and hypothesis are shared during

the scientific process, rather than at the end through the eventual paper publication (Burgelman et al 2010).

Other definition that have been mentioned in the literature include digital science, e-science, in-silico science, science in transition, participatory science, science highway, better science, digital humanities, open research and open scholarship.

The term science 2.0 can be used as an umbrella term for all of these definitions. Building on the work of Burgelman, Osimo and Bogdanowicz (2010) we introduce an ostensible definition of science 2.0 as consisting of the combination of three macro-trends (open scientific outputs, citizen science and data-intensive science) which encompass a set of lower level interrelated trends.

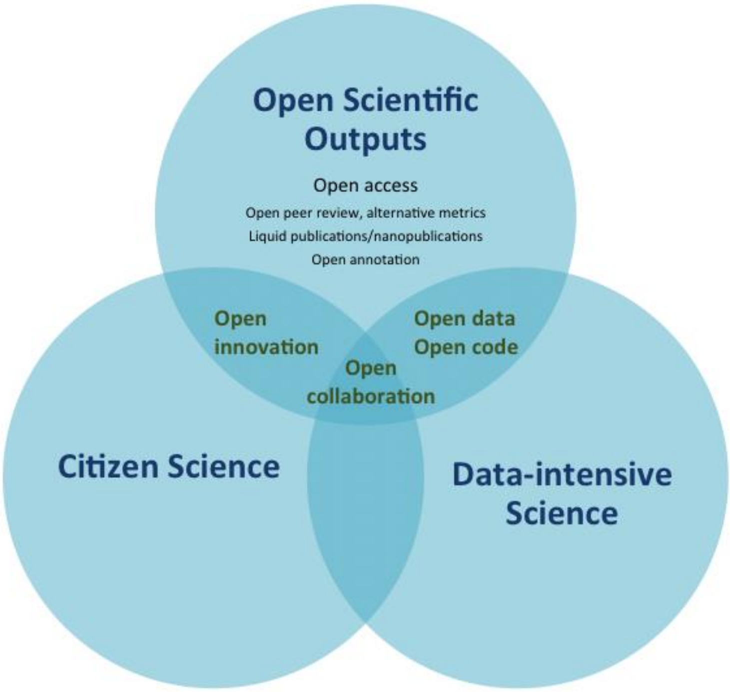

Figure 1: Science 2.0 interrelated trends. Authors proposal drawing on Burgelman et al. (2010

Those three macro trends are interrelated and mutually supportive which favours the choice of a single definition. The large availability of data calls for greater collaboration and involvement of citizens in the analysis (as in the case of the Galaxyzoo project). The greater importance of data supports open data sharing. Open scientific outputs enable the involvement of citizens as well of more extensive collaboration of scientists.

A second aspect of this choice of scope is that those trends cover the full chain of the scientific research process. Too often what is meant by open science is related only or mostly to open access to publications and other scientific outputs (open data and increasingly open code). The review of the underlying trends (presented in the next sections) clearly indicates how pervasive these trends are, from the conceptualisation through the data gathering and peer review process to publication (which is only the last step of that process) (see Figure 2).

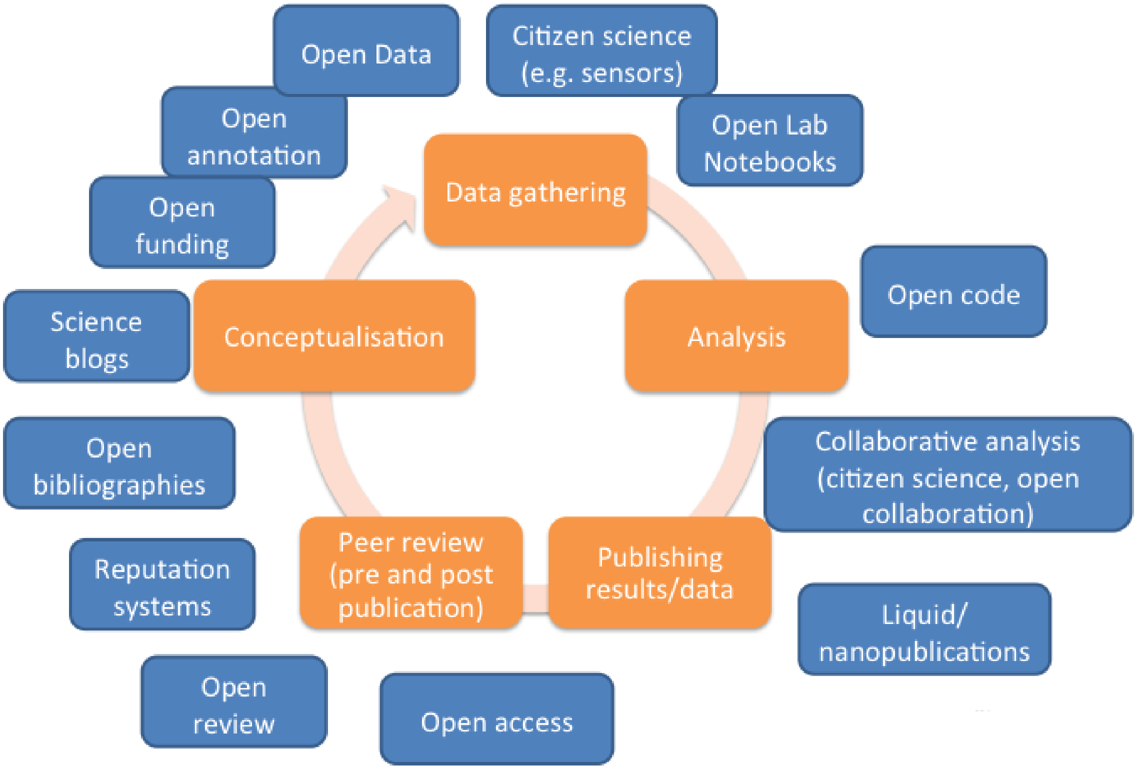

Figure 2 Science 2.0 in the scientific research process. Elaboration inspired by Open to All? Case studies of openness in research, NESTA & RIN report, 2010.

The proposed definition and framework is therefore comprehensive and reflects fully the variety of initiatives identified. It captures additional elements with respect to the other definitions presented.

4. Trends

This section presents the evidence from desk research and mapping of cases that allows for an initial assessment of importance of each of the underlying trends.

4.1 Open scientific outputs

4.1.1. Open access

Open access, understood as the "immediate, online, free availability of research outputs without restrictions on use commonly imposed by publisher copyright agreements" (as defined by Openaire[4]) is by far the most advanced policy and the most adopted trend. Already in 2008 the US National Institutes of Health (NIH) decided that all NIH-funded research is to be made openly accessible within 12 months of publication. All seven UK Research Councils have adopted similar open access policies requiring researchers they fund to make their work openly accessible. ROARMAP

[4] https://www.openaire.eu/

project (The Registry of Open Access Repositories Mandatory Archiving Policies) research indicates that Open Access take up grows especially due to the institutional mandate, i.e. self-regulation by research centres[5]. The number of open journals is growing exponentially - between 2002 and 2015 the number of open journals grew from 32 to 10 450 and the growth in all the countries is significant[6]. Already 44% of the peer-reviewed articles from 2011 were available for free in the Scopus database (Archambault et al, 2013). Currently, open access journals reach the same scientific impact as subscription journals. Initial differences in impact gradually disappear and as from 2000 there are no regional or discipline differences in this regard (Salomon et al., 2013). Most of the main publishers already offer Open Access option in some of their journals (in Nature Publishing Group 36 of its titles are fully OA and e.g. *Nature Communications* with an Impact Factor of 10.742[7]). BioMed Central, now part of Springer, publishes 278 peer-reviewed open access journals[8]. The Public Library Of Science non-governmental open-access editor as of 2011 is commercially self-sustaining, covering all operating costs with its publication fees revenue. Yet, even if all major publishers run OA and hybrid journals and despite of commercial success of PLOS group, only 55% of Welcome-funded researchers comply with OA requirements[9] which shows that the current incentives and requirements are insufficient on individual level. Recent decision of the Higher Education Councils in the UK to conduct the next Research Excellence Framework (REF) exercise (the system for assessing the quality of research in UK higher education institutions) only on the set of open access outputs may introduce the individual level incentives needed to increase the uptake[10].

4.1.2. Open data

One of the fastest growing trends of science 2.0 is open data in science (which is also part of a larger phenomenon of open data). Open data are 'freely available on the public internet permitting any user to download, copy, analyse, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself' as defined by Panton Principles[11]. Sharing enables to re-test the results of other scientists in order to validate their conclusions. It also makes possible to compile many datasets and use them to test for other hypothesis, i.e. more results can be obtained without spending additional money on data gathering. (Piwowar, Vision, & Whitlock, 2011). Open data policies are more and more visible on the political agenda with the pioneering example of Bermuda Principles and more recent requirements of funding agencies (Welcome Trust in the UK, NIH in the US, Horizon 2020 open data pilot). PARSE Insight survey showed that 64% of institutions already have institutional policies on data management[12]. Yet, even though there is some evidence showing positive correlation between publicly available data and the number of citations (Piwowar, Day,

[5] http://roarmap.eprints.org/
[6] https://doaj.org/
[7] http://www.nature.com/libraries/open_access/index.html
[8] http://www.biomedcentral.com/
[9] http://www.wellcome.ac.uk/News/Media-office/Press-releases/2012/WTVM055745.htm
[10] http://www.hefce.ac.uk/pubs/year/2014/201407/
[11] http://pantonprinciples.org/
[12] http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

& Fridsma, 2007), only 25% of researchers make their data openly available (PARSE international survey 2009). The ODE report shows that underlying research data is submitted for less than 20 % of all articles (Reilly et al. 2011). Another survey showed that only 6% of scientists share all their data and 46% do not share their data electronically at all (Tenopir et al., 2011). Fecher et al. (2015) recent work show that sharing data is still impeded by lack of formal recognition (citations, co-authorship), other study suggests that willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results (Wicherts et al., 2011). Fecher et al. (2015) conclude that the fear of data misuse and the missing out on a discovery is higher than the perceived benefits from data sharing. Therefore, the institutional support is very important in this trend (creation of data journals, founders' policies, multi-stakeholders' agreements).

4.1.3. Open code and reproducible science

Data availability is closely linked to another trend - open code and reproducible science. Victoria Stodden defines it as *a publishing standard, which includes analytical tools, raw data and experimental protocols, giving any scientist the possibility of reproducing a colleague's experiment* (Delfanti, 2010). Already 38% of scientists spend more than 1/5 of their time developing software (Merali, 2010) but code sharing is still lagging behind code using. Code sharing involves a cost in terms of documentation and clean up and does not reward the researches with an attribution that is part of the current academic reputation model. This is why code sharing is a growing trend with yet a small uptake. In 2012 only 38% of journals in the field of statistics had a data policy, 22% had a code policy, and 66% had a supplemental materials policy (Stodden et al, 2013).

4.1.4. Alternative reputation systems

Another trend is the growing call for alternative reputation systems beyond the Impact Factor (IF). The 2013 *San Francisco Declaration on Research Assessment* (DORA) underlines the urgent need to change the system under which the evaluations of outputs of scientific research are closely linked or dependent on the journal publishing model [13] and pointing to the IF limitations (skewed citations distribution, field-specificity, exposure to gaming by editorial policy and non-transparency of IF data, no impact measure outside of science publications, also discussed in Bollen, Van de Sompel, Hagberg, & Chute, 2009 or PLoS, Medicine, & Editors, 2006). New alternative measures are being proposed in order to address other activities that long were ignored such as sharing scientific outputs, communicating science to the public, incorporating citizens to experiments, collaborating with industry in an open or closed fashion, teaching (Nicholas et al., 2015).

For example, altmetrics is a social media-based metrics based on article's visibility on

[13] http://www.ascb.org/dora/

Twitter, Mendeley, Facebook, Connotea, blogs and non-scientific articles on the web (being used by BioMed Central, PLOS or Elsevier). It shows the outreach of science, measuring the impact on communities of practice and general public. Altmetrics do not limit itself to articles, showing popularity and re-use of datasets and other scientific outputs (e.g. slide show presentations). New tools that support that metrics are already available (Research Gate offers RG score based on publications views but also on social activity on the portal, Impactstory measures diverse research outputs including blogposts and slideshow presentations).

4.1.5. Open peer review systems

Another important discussion related to journal-based publishing of the scientific results is the issue of the current peer-review system. It is being challenged by attempts of fully open review and open post-peer reviews (e.g. F1000, open review in Nature pilot[14], minimum threshold publishing of PLOS One and recent PubMed Commons pilot project[15]). Even if open post-peer review systems do not escape from main traditional limitations related to peer review assessment (subjectivity, ´invisible colleges´, ´old boys networks´, etc., Wouters & Costas, 2012), they make the process more transparent. Hybrid approaches combining editor and peer-review filters with interactive and transparent post-publication forms of review already proved to be both cost-effective and assuring the quality of the outcomes (Poschl, 2012). Some journals already implemented different types of open peer-review, e.g. BMJ as from September 2014 publishes full history of the review, i.e. *all previous versions of the manuscript, the study protocol, the report from the manuscript committee meeting, the reviewers' signed comments, and the authors' responses to all the comments from reviewers and editors*[16]*,* some others publish it upon authors approval (e.g. eLife, GigaScience). Other proposals include the open public review of all scientific outputs at all stages of research cycle (e.g. Research Ideas and Outcomes already publishes research proposals[17]).

4.1.6. Open distributed collaboration

Open distributed collaboration of scientists became possible thanks to a myriad of web 2.0 tools. It is the large-scale, remote collaboration of scientists with the use of Internet-based tools similar to open source software collaboration (sharing of outcomes and practices). Scientists can choose, before submitting a formal article, to share their first insights, hypothesis on blogs or scientific forums asking for comments and collaborations almost in real-time. Tim Gowers, mathematician and blogger, sees *the utility of a blog* (for some scientific problems as showed in his Polymath project) *in*

[14] http://www.nature.com/nature/peerreview/debate/
[15] http://www.ncbi.nlm.nih.gov/pubmedcommons/
[16] http://www.bmj.com/about-bmj/resources-authors/peer-review-process
[17] http://riojournal.com/

being a happy medium between journal articles (which build upon each other too slowly) and conversations (which can often be too rapid to solve such complex problems)[18]. Scientists can also work collaboratively and publicly on wikis keeping an open notebook. Finally, there are examples of tools that allow for scientific workflow sharing (e.g. www.myexperiment.org). This distributed effort brings the niche, micro-expertise as well as microcontributions and microcollaboration (Olson et al. 2008) in the mainstream research resulting in increased efficiency and faster solutions to complex problems (ATLAS particle detector project[19] in CERN engaged 2800 scientific authors including 1850 with a PhD and 1200 technicians from 169 institutions from 37 countries, also thousands of industrial relations were built - Hoffman, 2009). Nevertheless, the growth of open collaboration is sketchy and its usage is relatively low. For example, UK researchers have only a passive use of well-known generic tools such as Google Scholar (73%) and Wikipedia (69%) (Procter et al., 2010). Also the usage of web 2.0-based services for novel forms of scholarly communication is relatively low, e.g. only 10% of UK researchers use the web 2.0 to reach non-academic audience (Procter et al., 2010)[20]. What is interesting, a US study points out that there is no evidence to suggest that "tech-savvy" young graduate students, postdoctoral scholars, or assistant professors are those that use the open web2.0 practices more often. (Harley et al, 2010). Still, there are some proved incentives for sharing the research outcomes on the web, e.g. higher citation impact of papers (Franceschet & Costantini, 2010). Compared to the traditional scientific dialogue within articles (Groth & Gurney, 2010) where most citations to publications occur after 3 years (often behind the paywall) and diminish after 10 years, the online conversation has the advantage of being more responsive and as well as more open to outsiders.

4.1.7. Other trends in open scientific outputs

The less prominent trends include beta publishing and content curation, liquid (evolutionary, "collaborative, and composable scientific contributions" [21]), and nanopublications ("smallest unit of publishable information: an assertion about anything that can be uniquely identified and attributed to its author"[22]), open annotation and open bibliographies[23] .

4.2. Citizen science

Citizen science macro trend displays a potential of public involvement in scientific research. It is *a form of research collaboration in which professional scientists engage*

[18] http://f1000research.com/2012/03/02/open-science-and-the-future-of-publishing-a-round-up-of-this-weeks-debate
[19] a complex scientific instrument that will be used to detect and measure subatomic particles in high-energy physics
[20] 12% write blogs at least occasionally; 20% comment on journal articles and 21% comment on blogs, while 17% contribute to a public wiki, and 15% contribute to a private one. Finally, 30% at least occasionally post slides and other content on public websites. (Procter et al., 2010)
[21] http://project.liquidpub.org/liquid-publications-scientific-publications-meet-the-web-1
[22] http://nanopub.org/wordpress/?page_id=65
[23] http://openbiblio.net/

with members of the public in order to accomplish scientific research (Wiggins & Crowston, 2011). Citizen science is also often seen as a part of a wider trend of crowdsourced science or ‘collaboratories ‘where *a large network of people collaborate after an open call for contribution* (Wiggins et al. 2011). Citizen science is not a new phenomenon; amateurs have always participated in scientific activities. Nevertheless, the advent of the Internet have transformed this process into a distributed community of citizen scientists that collaborate with scientists on a diverse set of projects sharing their computer power, time, abilities (human computing) and new perspectives (such as Zooniverse cloud-hosted suite of citizen projects [24] or Experimental Tribe platform[25]). Apart from tapping into a free source of computational power, labour, expertise and finance, citizen science proves a very effective method for formal and informal science education and public understanding of science. It also provides a machine learning assistance (Savage, 2012), which results in even faster computation and analysis of data (Smith et al., 2010, Fortson et al., 2011). Yet, there are not many examples of co-created projects, which are designed by scientists and members of the public working together and for which at least some of the public participants are actively involved in most or all steps of the scientific process (e.g. citizen activism Parris, 1999), DYIbio, crowdsourced health research studies (Swan, 2012). Most of the projects reproduce the hierarchy of the academia and allow only for a narrow intervention from citizens (data gathering or initial part of data analysis). It is also difficult to quantify the uptake of this trend. Intermediary services - SciStarter.com that aggregates citizen science initiatives as of February 2014 listed more than 700 projects citizen science projects. Most popular citizen science project gathered more than 1mio participants (with Zoouniverse[26] having 1.3 mio participants in beginning of 2015 and SETI@home 1.5mio users[27]) but there are many smaller projects with thousands or hundreds of users. This trend shows therefore a steady growth with very high participation in successful projects (top-heavy distribution of participation).

4.2.1 Open innovation and open funding

4.2.1.1. Open innovation

Open innovation (Chesbrough 2003, 2006), which is closely linked both to open collaboration and citizen science, attracted a lot of interest from industry and policymakers (e.g. the current strategic priorities of the European Commission in the area of Research, Science and Innovation are Open Science, Open Innovation and Open to the World[28]). This approach transforms public and private R&D activities by making the boundaries more permeable in order to capture knowledge from outside (i.e. users, subcontractors, clients, interested citizens). As of April 2014 the network of InnoCentive Solvers (contributors) comprised more than 300,000 registered users from nearly 200 countries[29]. Research on InnoCentive challenges shows that open approach to research solved one-third of a sample of problems that large and well-known R & D-

[24] https://www.zooniverse.org/about
[25] http://www.xtribe.eu/
[26] https://www.zooniverse.org/
[27] http://boincstats.com/en/stats/0/project/detail
[28] http://ec.europa.eu/research/openscience/index.cfm
[29] http://www.innocentive.com/about-innocentive/facts-stats

intensive firms had been unsuccessful in solving internally (Lakhani et al., 2007). Still, these results may probably be improved by completely opening up the submission process since Innocentive challenges solutions are not opened for public view. Bücheler and Sieg (2012) research suggests that performance may be increased by turning crowds of individuals into swarms of teams (based on collective intelligence and variety of backgrounds).

4.2.1.2. Open funding

Open funding is a growing trend of crowdsourcing funding for research projects and inducement prizes in funding research. Crowdfunding platforms such as Kickstarter and IndieGoGo, well established already in the innovation, policy, culture and charity sphere start to be used for funding research. Also new science-focused platforms are being launched either by commercial parties or research institutions (e.g. Science Starter in Germany). At the same inducement prizes (open innovation challenges) are developed by government agencies that use this instrument to fund challenge-based research (e.g. in the European Commission Horizon 2020 research programme the Inducement Prize Contests look for solutions to a set of societal and technological challenges, e.g. collaborative spectrum sharing).

There are also attempts to use alternative methods for the peer-review used for selecting research proposals in project funded research. Recently, Bollen et al. (2014) proposed a system of collective decision-making and pooling of research funds driven by algorithms and mathematical models which would in their opinion drastically reduce the current very high costs of both peer-review of research proposals and the time the scientific community spends on writing them instead of researching.

4.4 Data-intensive science

One of the key and visible trends is the data-intensive science stemming from the growing amount of data availability (Stodden et al 2009, Burkholder, L, ed. 1992, Donoho, 2009). Data-intensive science involves the use of datasets that cannot be stored, captured, managed and analysed by the mean of conventional database software. The new Large Synoptic Survey Telescope observing program will produce about 30 TB per night, leading to a total database over the ten years of operations of 60 PB for the raw data, and 30 PB for the catalogue database[30]. Large Hadron Collider at CERN already produces 25 PB of data annually which is analysed by a special grid of 150 computer centers[31] (the Worldwide LHC Computing Grid). This data explosion is not only driven by large scientific observatories. Lazier et al. (2009) foresee quantitative revolution in social science due to the abundance of social network data. An increasing amount of data is being produced by the increased pervasivity of sensors in non-scientific objects. For example, today's smartphones have Wi-Fi, GPS, accelerometers, camera, microphones, and gyroscopes. Citizens increasingly leave a "data shadow" (Coleman, 2008) as a result of their online and offline behaviour. This data is crucial for the understanding of human behavior.

[30] http://www.lsst.org/lsst/
[31] http://wlcg.web.cern.ch/

This radical growth in data availability leads to modifications in the nature of the scientific method. Gray (2009) pointed out that ICT is changing the very nature of science. Information technology has not only affected the three traditional paradigms of science (experimental, theoretical and computational), but it has also fostered the emergence of a fourth - data intensive - paradigm. This new approach, which Gray called eScience or data-intensive science, unifies theory, experiment and simulation.
Due to data availability and augmentation of computing power brand new fields of science are being created: computational chemistry, biology, economics, engineering, mechanics, neuroscience, geophysics, computational social science. Lazier et al. (2009) foresee quantitative revolution in social science due to abundance of social network data.

There are today companies dedicated to the application of big data in science, such as InSilico DB and Ugentec, mainly in the field of genomics and web platforms such as Kaggle.com provide "big data research as a service".
Yet at the same time, there is little data on big data usage among scientists. In a 2011 survey of Science peer-reviewers 20% of scientists declared that they deal in their research with datasets larger than 100 GB, and 52% larger than 1 GB (Science editorial, 2011).

4.5 Summary of trends

The analysis of trends as well as the mapping of cases (see Table 1) shows that there is already a large number of initiatives (more than 100 described in the literature) spread in different phases of the research cycle and among different disciplines. The most interesting and studied initiatives provide added value compared to traditional methods of science (transparency, openness to new actors, efficiency and reliability) and show increasing participation of scientists (although with different intensity across trends).
The mapping results however point to the relative advancement of natural science compared to social sciences and humanities (only a few cases found for all three macrotrends). Possible explanations include the fact that the latter usually require smaller teams and therefore less advanced collaboration methods, the datasets used are usually smaller; there is a larger prominence of qualitative studies and therefore less need for citizen support.

Table 1 Mapping of science 2.0 cases

	Open science	Citizen science	Data-intensive science
Total	44	40	21
All	105 cases		

In all trends there is a wide availability of services/infrastructure that supports the uptake (elaborated web platforms and software as a service solutions). The level of uptake of these trends by scientists varies between trends with the highest adoption in open access (across domains), citizen science (domain specific) and data-intensive science (in selected disciplines).

5. Implications

The final part of discussion presents a set of transversal implications for the science method and scientific organizations based on literature review and validated through the interviews. The cross-cutting implications of those changes sustain the reason for keeping a single definition for the science 2.0 trends.

5.1 Enhanced efficiency, transparency and reliability

As discussed, increasingly, articles are being made available together with underlying data, code and supplementary information - notes, interactive graphs, etc. Very often articles are linked to institutional repositories where the data is available for downloads (as also as part of the journals requirements). Open access to scientific outputs enhances the transparency and reproducibility of the scientific process (Ioannidis 2005). This may result in tighter control over scientific fraud. Recent cases in data manipulation or error (Diederik Stapel fraud, Duke microarray case or Reinhart and Rogoff) show how data availability is important to eliminate badly conducted research. Moreover this compendium type of disclosure (Stodden, 2010) is important for sharing part of the practical knowledge, especially if part of the practice is being shared – lab notes, workflows, code. As a result, it enables the other researchers to replicate results, test the data and mine the articles content (Murray-Rust 2008). Finally, it further strengthens the cumulative character of the scientific endeavour by producing larger datasets (GeneBank) and better code.

Yet, the mere availability of data will not eliminate incentives for manipulations of scientific outputs (which are deeply embedded into the *publish or perish culture*). Examples of predatory open access journals and hijacked journals[32] (Bohannon 2014, Bealls' list[33]) show that the new model of publication can also give rise to other forms of unethical behaviours. Publication of underlying data and code will have to be funded either by funding agencies or from other sources as the cleaning of data-sets and code are resource-intensive. Creation of the sharing culture demands not only financial resources but also recognition, since current reputation model does not reward individual scientist for opening up those research outputs. In other words, sharing of intermediate results benefits the community but not the individual researcher who takes the decision to publish it (Scheliga, Friesike, 2014). Also, there is a limit to the extent the more tacit knowledge of a scientific process can be shared without personal contact.

[32] Predatory open access journals are journals abusing the business model of open access journal where the author pays the publication fee without providing strict editorial rules and reliable peer-review process. Hijacked journals are esteemed scientific journals which reputation was used to create a fake internet website offering open access publishing for a fee.

[33] http://scholarlyoa.com/publishers/

5.2 Raise of data-driven science

The advent of big data – both by accumulating research data (Genome Project) and producing new data from the social digital activities demands a more computational approach to analysis. But if the usage of computational methods will be far greater, it will demand more storage space (e.g. cloud infrastructure), computing power (grids and distributed computing), data skills and a lot more of citizen scientists, in other words: infrastructure, personnel and skills. What is more, data-intensive science presents many risks, such as data reliability, privacy violations when databases are merged, limitations linked to the usage of social media data (e.g. bias towards the presence stemming from poor archiving and search functions (Crawford, 2010). Data-intensive science may also magnify problem with confirmation bias (Ioannidis, 2005) and augment the number of spurious correlations stemming from data dredging.

5.3 Microcontributions on a macroscale

Researchers are facing public expectations to solve ever more complex problems, which demands the presence of interdisciplinary teams. With the advent of the Internet those communities of scientists (e.g. CERN) grow larger and larger and are supported by citizen scientists in some of the (more menial) tasks. Those endavours allow for more serendipity because of the scale of contribution, the myriad of different actors and wealth of data produced and analysed. Thus, unforeseen discoveries were already made that went far beyond the original research goals (e.g. Hanny's Voorwerp during the project Zooniverse, Christian et al., n.d.). Nevertheless, the scientific process in many of those cases is cut into small contributions similar to Amazon Turk. This brings the issue of hyperspecialisation of scientists and the question of missing the bigger picture of a given research problem. What is more, the microcontributions are linked to a wider phenomenon of freelance work and are posing issues related to researchers' precarity. Also, management of large collaboration groups generates higher transaction costs. Finally, the push for engagement of citizen scientists may result in further gamification of science (von Ahn, L., & Dabbish, L. 2008) and distort for the wider public the research objectives.

5.4 Multidimensional, immediate and multiform evaluation of science

Impact factors and peer review have been the cornerstone of the assessment of the quality science outputs and researchers' excellence. In the future, we can expect the co-existence of diverse instruments, more open methods by a wider variety of evaluators, and faster feedback to account for the increased cycle of science publishing. There are experiments with new evaluation and reputation models such as data citation and data journals, altmetrics and post-publication reviews. Nevertheless, most of those alternative reputation systems offer a complementary solution to traditional citation measures by showing the outreach of science (i.e. impact on communities of practice and general public) registering mostly the views and downloads of scientific outputs rather than distinguishing between different roles of scientists (peer-reviewer, teacher, micro-contributor, etc). Altmetrics can on one hand identify and credit "silent" authors (read but not cited), which have important influences for the field (Shadbolt, Brody, Carr, & Harnad, 2006) but on the other hand has the downside of measuring mostly social reach rather than real impact of the scientific work. What is more, usage-based metrics can privilege controversial content driven by 'clickability' rather than scientific soundness (which already happens with the scientific article press releases – Sumner et al. 2014). Finally, the networked community Internet is even more prone to the Matthew effect (Merton 1968) so the inclusion of usage-based metrics may augment the presence of star scientists (Moody 2004).

Different experiments with pre and post-reviews (open peer-review review Nature's 2006 trial, F1000 post-review) are one of the alternative or complementary methods to filter out the increasing number of publications. The completely open peer-review process increases the accountability of reviewers and editors, enhances transparency as well as minimalises influence of competing interest. Yet, if we open the process of publishing scientific outputs, the issue of peer-review will have to be extended to data and code (e.g. are journals dedicated to datasets). Also the open peer-review has to offer incentives for reviewers that will counterbalance the negative reactions linked to the disclosure of the reviewers' names (van Rooyen et al., 2010). Apart from the opening up of the peer-review process, there are some of the research outputs that are published in their non-finalised form, in line with the *publish and filter* way. Sharing scientific results in a draft form enables collaboration, updates and quick reaction from the community compared to a long time till retraction or a new updated version of a research article. The question is what are the validation measures apart from the number of re-tweets. What is more, some research suggests that sharing research ideas at an early point may even stifle some innovation by directing it in one direction. Sharing the early ideas or intermediate outputs lowers the incentives to carry out similar research and discourage efforts to participate as well as decreases chances of testing other solutions to a problem. However, the incentives-versus-reuse tradeoff literature shows that the early disclosure enhances the knowledge transfer and creates convergence hence a critical mass (Boudreau, Lakhani 2014).

5.5 Disaggregation of the research value chain

Altmetrics solutions are just one of the examples of disaggregation of the scientific publication value chain. Others include pre-publication archiving -arXiv, data archiving, collaborative research platforms, social networks for scientists than enable self-archiving, crowd collaboration tools, etc. First changes were brought by the rise of new open access journals 'researcher pays' business model. Nevertheless, the OA journals, traditional publishers and new players still experiment with new business models (Gold access, membership plans[34]). Similarly to Red Hat[35] start-ups and academic spin-offs are already offering specific services based on open collaboration (e.g. in biomedicine - Transparency Life Sciences[36] is a drug development company based on open innovation and designing clinical test with the use of crowdsourcing).

[34] *PeerJ journal offers a membership plan that covers the publishing fees of one to an unlimited number of papers annually*, see more at http://peerj.com/

[35] http://www.redhat.com/

[36] http://transparencyls.com/

However, those new tools and platforms are offered both by non-profit and for-profit organisations, which brings the risk of new lock-in (similar to those of Apple Store, Amazon or Facebook) and 'walled gardens'. This enhanced competition and instability is likely to have a similar impact to the one seen in other content industries, with new players emerging, opening up of new opportunities for business and for scientists, and at the same time the risk of disruption of existing systems and less focused on long-term and large-scale research. Thanks to the data on researchers and search patterns they possess, those players will be able to offer more individualized services (similar to Trip Advisor or Amazon) in return for the ownership of researchers' personal data. There is also a question of long-term stability of those tools (that can be discontinued or bought by larger players as it happened to Mendeley and the burst of the application bubble). In other words, the so often criticised scientific disclosure system in a scientific journal created by the exigency of noble patronage rather than needs of science still has the advantage of stability (David 2004).

5.6 Arrival of new actors and the democratization of science

Another faction of the disaggregation of the research value chain is the advent of new actors (Nowotny, 2001). Citizens are more involved in the scientific process by having more access to scientific outcomes (through open access, social media), through citizen science and by crowdfunding of scientific activities. The crowdfunding tools (such as Kickstarter[37]) enable the public to vote for the most needed research and therefore re-define research priorities. More informed citizens would probably understand better the need for research funding in the public budget.

Yet, this change towards demand-driven science based on bottom-up funding and citizen-researcher collaboration may strongly privilege applied science with foreseeable applications and be driven by particular interests (e.g. crowdsourced health research) rather than pure science objectives. Also, the most prominent citizen science projects are still very hierarchical, i.e. citizens are members of the science teams only at the level of data gathering or initial data analysis.

Table 2 Summary of the implications for the science methods and research institutions

Implications for the science method	*Implications for research institutions*
Enhanced efficiency, transparency and reliability	Multidimensional, immediate and multiform evaluation of science
Raise of data-driven science	
Microcontributions on a macroscale	Important impact on the whole value chain of service providers for science

[37] http://www.kickstarter.com/

	Arrival of new actors and the democratization of science

6. Conclusions

The analysis of the trends shows that the ICT-driven trends are transforming the method and the institutions of science (see the summary table 2) but the drive towards a more open science is only augmenting some of the characteristics of science that existed already in the 17^{th} century, i.e. based on the public disclosure of knowledge (David, 2003). The advent of the Internet and web 2.0 tools have changed the way of doing research but at the same time enforced the scientific values as proposed by Merton based on communal ownership, universalism, disinterestedness and organised scepticism.

The arguments propounded by the open science and science 2.0 activists point to a vision of science that is:

- less hierarchical by the higher level of inclusion of citizens (citizen science, crowdfunding, open innovation) - universalism
- open by sharing data, methods, tools and outcomes - communal ownership
- transparent by the ex-ante and ex-post open and inclusive process of peer-review - disinterestedness
- replicable thanks to share-able methods, tools and data - organised scepticism

The Mertonian description of science endeavour is of course ideal typical in the Weberian sense (Weber, 1949) since science is a human activity and its rules were and are socially influenced (peer review, grant process, institutional politics, etc.). Therefore in the previous discussion on the implications we have shown that the emerging trends have the potential of enhancing openness and transparency of scientific research but embedded incentives system as well as career trajectories have important influence on individual scientists' behaviour.

Our research shows that these disparate trends can be brought together under a single definition of Science 2.0, which encompasses open science, citizen science and data-intensive science. These trends are complementary and mutually reinforcing and cover the full research cycle. The proposed framework reflects well the full variety of initiatives identified all along the research value chain.
Secondly, these trends are important, potentially transformative and deserve attention. While their importance varies depending on a trend and a discipline, what is clear is that we are no longer referring to isolated individual projects. Not only the number of initiatives increases, but they cover all the phases of the research cycle with however very uneven distribution as regards scientific disciplines. Moreover, the number of initiatives is continuously growing because of the emergence of a whole ecosystem of services. For every phase of the value chain identified in Figure 2, we can find "science-as-a-service" solutions that enable a scientist to launch a "science 2.0" initiative at minimal cost.

The potential implications of these cover a wide range of issues and address some of the key challenges of science today, both in a positive and a negative way. Yet, the evidence about the impact is clearly lacking so that at this stage we can only speculate on the typology of potential implications but not assess their significance.

We can however see three different enablers of this change – policy measures, individual practice of scientists and new infrastructure and services. The individual scientific practice is lagging behind in many trends (e.g. open data) but is very important for emerging trends to underpin the need for a new policy intervention. Needless to say that infrastructure and tools are a fundament for building a stable environment for those new practices. Therefore, the policy action is also important when the infrastructure and tools offered are fully private to avoid situation of recreating non-transparent systems and walled gardens.

The evidence on positive impact is not yet strong enough to justify the promotion of a universal uptake of science 2.0 across all disciplines. Yet, what is clear is that the current system of incentives for researchers actively discourages the uptake of Science 2.0 by not giving them credit for most of the science 2.0 activities. The main policy action to be taken, at this stage would be the work on removing the barriers. It would entail implementing a more diverse set indicators measuring research quality and impact (e.g. in the performance-based funding models such as the UK Research Excellence Framework exercise) and supporting science 2.0 public infrastructures (e.g. data repositories).

There is also a need for further research on comparable data on uptake of science 2.0, robust evidence of the impact (both positive and negative) of such initiatives, either at individual or systemic level, assessment of policy impact in this domain as well understanding of implication of business involvement in the research practices (especially important for reputation management) and more exploratory work on weaker trends.

References

Archambault E, Amyot D, Deschamps P, Nicol A, Provencher F, Rebout L , Roberge G (2013). Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011. Science-Metrix, retrieved from http://www.science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf .

Bartling S, Friesieke S (2014) ed. Opening Science. Springer Open. http://www.openingscience.org

Bohannon J. (2013). Who is afraid of the Peer Review? Science 342: 6154, pp. 60-65. doi: 10.1126/science.342.6154.60.

Bollen J, Van de Sompel H, Hagberg A, Chute R (2009). A Principal Component Analysis of 39 Scientific Impact Measures, PLoS ONE 4:6/ e6022. doi:10.1371/journal.pone.0006022.

Bollen J, Crandall D, Junk D, Ding Y, Börner K (2014). From funding agencies to scientific agency. Collective allocation of science funding as an alternative to peer review. EMBO reports 15: 2, pp. 131–133.

Boudreau K., Lakhani K. (2014). 'Open' Disclosure of Innovations, Incentives and Follow-on Reuse: Theory on Processes of Cumulative Innovation and a Field Experiment in Computational Biology. Research Policy 44: 1, pp. 4–19. doi: 10.1016/j.respol.2014.08.001

Burgelman J.C, Osimo D. Bogdanowicz M. (2010). Science 2.0 (change will happen…). First Monday 15 7 -5.

Burkholder, L. (Ed.). (1992). Philosophy and the Computer. Boulder: Westview Press

Coleman, S (2008). Foundations of digital government. In: Digital government. E-Government Research, Case Studies, and Implementation. Springer US. pp. 3-19.

David P A (2004). Patronage, Reputation and Common Agency Contracting in the Scientific Revolution: From Keeping 'Nature's Secrets' to the Institutionalization of 'Open Science'. Discussion Papers 03-039, Stanford Institute for Economic Policy Research.

Dasgupta P, David PA (1994). Toward a new economics of science. Research Policy 23:5. pp. 487-521 doi: 10.1016/0048-7333(94)01002-1.

Davis P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. FASEB journal: official publication of the Federation of American Societies for Experimental Biology 25: 7, pp. 2129-34. doi:10.1096/fj.11-183988.

Donoho D.L. et al. (2009). Reproducible Research in Computational Harmonic Analysis. Computing in Science and Engineering 11:1, pp. 8-18.

Delfanti, A. (2010). Open science, a complex movement. Editorial. Journal of Science Communication, 9.

Fecher B, Friesike S, Hebing M (2015). What Drives Academic Data Sharing? PLoS ONE 10:2 e0118053. doi:10.1371/journal.pone.0118053

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. Journal of Informetrics, 4:4, pp/ 540–553. doi:10.1016/j.joi.2010.06.003.

Gray, J (2009). Jim Gray on eScience: A transformed scientific method. In: The fourth paradigm: Data-intensive scientific discovery ed. by Hey T, Tansley S, and Tolle K. Miscrosoft Research. Redmond, pp. XVII-XXXI

Groth P, Gurney T.(2010). Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study, In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, Raleigh, NC: US.

Groth, P., Gibson, A., Velterop, J. (2010). The anatomy of a nanopublication. Information Services and Use, 30: 1, pp. 51–56. Retrieved from http://iospress.metapress.com/content/ftkh21q50t521wm2, last accessed in April 2015.

Hey T., Tansley S., Tolle K., (2009). The fourth paradigm. Data intensive scientific dicovery. Microsoft Research, Redmond, Washington.

Hoffman HF (2009). Open, global collaborations in particle physics, presentation, http://communia-project.eu/communiafiles/Conf%202009_S_%20Hoffmann_KS.pdf , ppt retrieved on 28.02.2014

Hufbauer K. (2006). Amateurs and the Rise of Astrophysics 1840–1910, Berichte zur Wissenschaftsgeschichte, 9:3, pp. 183–190, 1986.

Kannappan S, (2001). Border Trading: The Amateur-Professional Partnership In Variable Star Astronomy, MA thesis, MIT, retrieved from http://user.physics.unc.edu/~sheila/entirethesis.pdf on April 20 2015

Kotarski R, Reilly S, Schrimpf S, Smit E, Walshe K (2012). Report on best practices for citability of data and on evolving roles in scholarly communication. ODE report: retrieved from http://www.stm-assoc.org/2012_07_10_STM_Research_Data_Group_Data_Citation_and_Evolving_Roles_ODE_Report.pdf , last accessed in April 2015.

Lakhani, K. R.; Panetta, J. A. (2007). The Principles of Distributed Innovation. Innovations: Technology, Governance, Globalization, 2:3, pp. 97-112. doi:10.1162/itgg.2007.2.3.97.

Lazier et al. (2009). Computational Social Science, Science 323:5915, pp. 721-723, doi: 10.1126/science.1167742.

Lyon, L. (2009). Open Science at Web-Scale : Optimising Participation and Predictive Potential Consultative Report. JISC.

Merali Z (2010). Computational science: ...Error, Nature 467, pp. 775-777, doi:10.1038/467775a.

Merton, Robert K. (1973). The Normative Structure of Science. in Merton: The Sociology of Science: Theoretical and Empirical Investigations, Chicago: University of Chicago Press.

Murray-Rust P (2008). Open Data in Science, Serials Review, 34: 1.

Moody, J. (2004). The Structure of a Social Science Collaboration Network. American Sociological Review. 69, pp. 213-238.

Mukherjee A, Stern S, (2009). Disclosure or secrecy? The dynamics of Open Science. International Journal of Industrial Organization, 27, pp. 449 – 462.

Neylon C, Wu S (2009). Article-level metrics and the evolution of scientific impact. PLoS biology 7:11, e1000242.

Nowotny H., Scott P, Gibbons M (2001). Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty. Polity Press.

Nesta and RIN joint report (2010). Open to All? Case studies of openness in research. http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies , last accessed in April 2015.

Nicholas D, Herman E, Jamali H R (2015). Emerging reputation mechanisms for scholars. JRC Science and Policy Report. JRC- IPTS.

Nielsen, M. (2011). Reinventing Discovery: The New Era of Networked Science, Princeton University Press.

PARSE (2009). Insight into digital preservation of research output in Europe, retrieved from: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf , last accessed in April 2015.

PLoS, T., Medicine, & Editors (2006). The impact factor game. It is time to find a better way to assess the scientific literature. PLoS Medicine, 3:6 e291. doi:10.1371/journal.pmed.0030291.

Piwowar, H. A., Vision, T. J., & Whitlock, M. C. (2011). Data archiving is a good investment. Nature. 473, pp. 285. doi.org/10.1038/473285a.

Piwowar, H., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PloS one*, *2*(3), e308. doi:10.1371/journal.pone.0000308

Priem J, Piwowar H, Hemminger B (2011). Altmetrics in the wild: An exploratory study of impact metrics based on social media, Metrics 2011: Symposium on Informetric and Scientometric Research. Vol 12.

Poschl M, (2012). Multi-Stage Open Peer Review: Scientific Evaluation Integrating the Strengths of Traditional Peer Review with the Virtues of Transparency and Self-Regulation. Frontiers in Computational Neuroscience, 6:33.

Olson J. S., Hofer E., Bos N., Zimmerman A., Olson G. M., Cooney D., and Faniel I. (2008). A theory of remote scientific collaboration. in G. M. Olson, A. Zimmerman, and N. Bos (Eds.) Scientific Collaboration on the Internet. Cambridge, MA: MIT Press.

Parris T. M. (1999). Bytes of Note: Connecting with citizen science. Environment: Science and Policy for Sustainable Development, 41:10, pp. 3-3.

Procter R., Williams R., Stewart J., Poschen M., Snee, H., Voss A., Asgari-Targhi M. (2010). Adoption and use of Web 2.0 in scholarly communications. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, *368*(1926), 4039-4056.

Savage N. (2012). Gaining wisdom from crowds. Communications of the ACM, 55:3, pp. 13. doi:10.1145/2093548.2093553.

Scheliga K, Friesike S (2014). Putting open science into practice: A social dilemma? First Monday 19:9.

Shadbolt N, Brody, T Carr, L and Harnad S (2006) The Open Research Web: A Preview of the Optimal and the Inevitable. In, Jacobs, Neil (ed.) Open Access: Key Strategic, Technical and Economic Aspects., Chandos.

Smith A., Lintott C., Alliance T. C. S. (2010). Web scale citizen science. Proceedings of the Royal Society Discussion Meeting Web Science: A New Frontier.

Solomon D, Laakso M, Björk B-C (2013). A longitudinal comparison of citation rates and growth among open access journals. Journal of Informetrics, 7: 3, pp. 642-650.,doi:10.1016/j.joi.2013.03.008.

Science editorial (2011). Dealing with data. Challenges and opportunities. 331:6018, pp.692-3.

Stodden, V. (2010). Massive Data , the Digitization of Science , and Reproducibility of Results, presentation. CERN Colloquium Geneva, Switzerland retrieved from: https://web.stanford.edu/~vcs/talks/VictoriaStoddenCERN2010.pdf , last accessed in April 2015.

Stodden V (2009). Enabling Reproducible Research: Open Licensing for Scientific Innovation, International Journal of Communications Law and Policy 13.

Stodden V. et al. (2013). Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals, PLOS One, 8:6 e67111. doi: 10.1371/journal.pone.0067111

Swan, M. (2012). Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem. J Med Internet Res 14:2 e46. doi:2196/jmir.1988.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. (2011), Data Sharing by Scientists: Practices and Perceptions, PLOS One, 6:6 e21101. doi: 10.1371/journal.pone.0021101

Ziman J. (2000). Real Science: What It Is, and What It Means. Cambridge: Cambridge UP.

Wicherts JM, Bakker M, Molenaar D (2011). Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. PLoS ONE 6:11: e26828. doi:10.1371/journal.pone.0026828

Wiggins A., Crowston K. (2011). From Conservation to Crowdsourcing : A Typology of Citizen Science. Proceeding HICSS '11 Proceedings of the 2011 44th Hawaii International Conference on System Sciences. doi:10.1109/HICSS.2011.207.

Wouters P, Costas R (2012). Users, narcissism and control: Tracking the impact of scholarly

publications in the 21st century. Utrecht: SURF foundation. Retrieved September 20, 2012 from: http://www.surffoundation.nl/nl/publicaties/Documents/Users%20narcissism%20and%20control.pdf, last accessed in April 2015.
von Ahn, L., Dabbish, L. (2008). Designing games with a purpose. Communications of the ACM, 51:8, pp. 57. doi:10.1145/1378704.1378719.

Appendix 1

List of interviewees

Name	Profile
	Open access to scientific outputs
1. David Hoole	Publisher - Head of Marketing, Nature Publishing Group
2. Barbara Kaloumenos	STMS, Federation of Publishers
3. Bettina Goerner	Springer, Open Access Manager
	Open Funding
4. Suzanne Iacono	Senior Advisor and Programme Director of the Global Environment for Network Innovations programme of National Science Foundation, US
	Open Innovation
5. Simon Schneider	Innovation Evangelist, Head of Innocentive Grand Challenges
6. Anthony Williams	Expert, Co-author of Wikinomics
	Open Science
7. Roberto Casati	Senior researcher at CNRS-EHESS-ENS Institut Nicod.
8. Cameron Neylon	Senior Scientist, Science and Technology Facilities Council Didcot, United Kingdom, Director of Advocacy at PLoS
	Citizen science
8. Francois Grey	Coordinator, Citizen Cyberscience Centre, CERN+UNITAR+UNIGE Professor of Distributed Scientific Computing, Tsinghua University, Beijing

9. Muki Haklay	Professor of GIScience, UCL London Extreme Citizen Science (ExCiteS) research group
Open and distributed collaboration	
10. Markus Nordberg	Scientist, CERN ATLAS resources co-ordination