Science 2.0: the deep unbundling

Final report

Table of Content

Table of Figures

Executive summary

This paper briefly outlines possible futures scenarios of science 2.0, analyses its implications and draws policy recommendations “fit for the future”. Science 2.0 is more than open access: it refers to the emergence of open, data-intensive and citizen science across the full research cycle, from data gathering to reputation management.

Science 2.0 is here to stay and it is already growing well beyond individual projects. On the supply side, an ecosystem of services and standards is emerging. Adoption is growing and becoming mainstream already in some phases such as preprint publication, reference sharing, open access publication. Impact is already visible and will address some of the most burning issues of science, such as the slowness of the publication process and the challenge of reproducing research results.

Based on the extrapolation of existing trends and on analogies from different domains, we anticipate a set of “scenario snippets”:

- The full integration of data, publications and intermediate product will enable reproducibility by default. But adoption of such sharing culture will require time and a new system of incentives based on impact metrics and career structure.
- Evaluation metrics will become multidimensional, granular and instantaneous;
- The work of scientist will change with greater collaboration and independence from institutions.

Overall, we will see an unbundling of services, which are today integrated. Research will be separated from teaching, data collection from data analysis, publication from reputation management. Different specialised service will emerge and displace the incumbents such as publishers and universities. At the same time, the value chain will reorganise through vertical integration around new platforms. These could be built around unexpected positions in the value chain, including electronic reading devices.

In terms of implications, these scenario show opportunities and risks in three main areas.

For the productivity of science, there will be a dramatic increase on the return on investment in science since the publication process will be sped up; data availability and reuse will lead to greater impact from data collection; meta-reviews will become semi-automatic; scientists will have fast access to the specific micro-expertise they need to solve a specific problem. However, the sharing of data and intermediate outputs will not happen naturally; new walled gardens could emerge that reduce interoperability; and making intermediate product ready for reuse will add an additional workload for researchers.

For the quality of science, the scenarios will radically reduce the impact of false scientific claims as these will be quickly uncovered and thereby actively discouraged. The availability of interoperable datasets will accelerate the identification of unexpected correlation thereby facilitating new discoveries. Inductive, data-intensive methods will become more important across all scientific disciplines. However, open reviews (instead than traditional peer review) will make it harder to identify high-quality papers and will favour most popular and catchy findings. The rise of alternative metrics will discourage scientists from undergoing long-term, unpopular and therefore disruptive research path, leading to greater homogeneity and reduction in disruptive discoveries.

For the scientific value chain, new players will emerge, existing incumbents will loose their control of the value chain, and researchers will become more independent. Different specialised services will be used for the different phases. At the researchers level, there will be greater collaboration and subdivision of work through micro-tasking and micro-expertise. This unbundling will be followed by a rebundling, or a vertical concentration of the value chain around new platforms, enabled by technical proprietary standards, personal data ownership, network effects and preferential attachment. Europe has the possibility to lead; however in the absence of the development of a healthy ecosystem it is likely that non-European players will be the owner of the new platforms.

Based on these implications, we formulate recommendations on both the demand and the supply side. These are complementary and necessary to each other because what is needed is a strong European ecosystem. The basic assumption is that science 2.0 is here to stay and that it provides the opportunity to address some of the most important challenges faced by science.

On the demand side, we propose to remove the barriers to science 2.0, and in particular the fact that the career of scientist is still based on the “publish or perish” principles. Alternative evaluation and metrics should be developed. Researchers, projects and programmes should not be evaluated based on publications and patents. The sharing of data and intermediate products can’t be achieved simply by introducing obligations related to funding: there should be a comprehensive “stick and carrot” approach across the overall system. Skills related to data-intensive approaches should be strengthened across all scientific disciplines.

On the supply side, we suggest to adopt a proactive industrial policy characterised by an open, global and interoperable approach. New, platform-based business models should be promoted. At the same time, the emerging new platform should guarantee interoperability not only of scientific outputs, but also of additional services such as researchers' personal data, reputation and identity management.

As we learned from the web 2.0, taking no action will not lead to maintaining the current European leadership, but rather jeopardize it.

“…we never bother to 'publish': we just post our findings on weblogs, and if they get a lot of links, hey, we're the Most Frequently Cited. Tenure? Who needs that? Never heard of it! Doctorates, degrees, defending a thesis — don't know, don't need 'em, can't even be bothered!”

Sterling, B. (2005). Ivory Tower. Nature, 434(7034), 806–806. doi:10.1038/434806a

1. Purpose and method

1.1 Objectives of the report

In this report we present visionary scenarios about the possible future deployment of Science 2.0.

The purpose of such visionary scenario is to open up the policy thinking by introducing new ideas that are currently outside the policy radar. The scenarios are not predictions: they are a methodological instrument to develop out-of-the-box thinking and to derive policy recommendations “fit for the future”.

This report builds on the previous study on the implications of science 2.0 elaborated by Tech4i2 . It is however deeply different in nature: while the previous one aimed at robustly assessing trends and providing the best available data, this one aims to more freely extrapolate on how these trends could play out in the future.

To do so, we elaborated initial scenarios and posted them publicly in order to generate discussion and feedback. The present scenarios have been enriched by online and offline discussions with scientists.

The report presents scenarios of science 2.0 in 2030, about 16 years from today. The date is purely instrumental to stimulate some out-of-the-box thinking while still being able to relate the change to the on-going trends. To appreciate the possible significance of changes by that time, it is worth remembering that 16 years ago, Google did not exist.

1.2 Methodological approach

To develop the elements of the scenario, we use two methods:

- Extrapolation from current trends in the science 2.0 context: for instance, we develop ideas based on the full deployment and adoption of the existing alternative impact metrics such as altmetrics.org

- Analogy from existing visible impact in other domains, such as entertainment, enterprise 2.0 or government 2.0: for instance, the consumer content and technology industry can be considered as an "early adopter" of the 2.0 approach, and it is interesting to foresee the deployment of similar impacts in the scientific domain.

In other words, we first try to imagine future scenarios if the current trends and "weak signals" become mainstream. We then elaborate possible ideas on how trends that have already deployed in similar and more mature domains could play out in the science field.

Through the elaboration of the scenario, we try to develop both positive and negative perspective of the implications of those scenarios. Based on these implications, we draw policy recommendations which aim to make the European Research Area "fit for science 2.0".

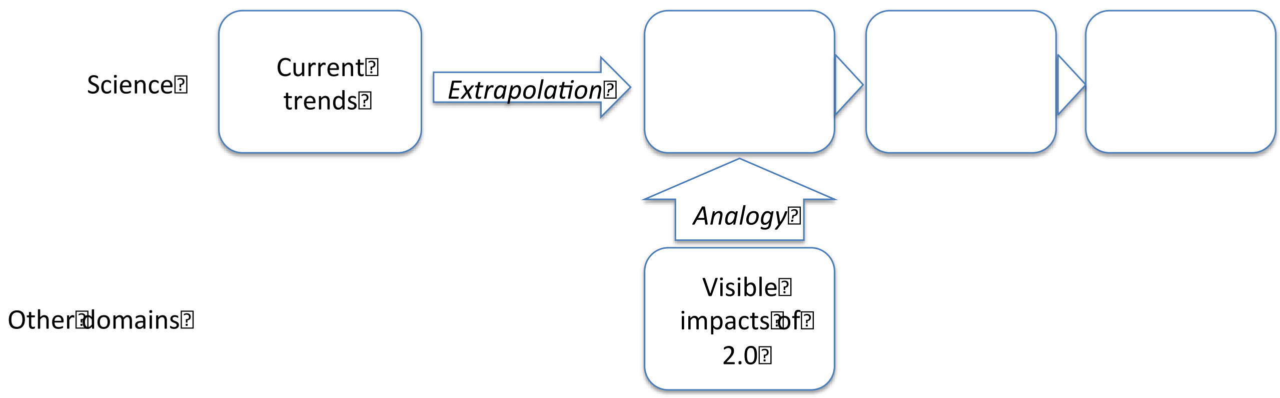

Figure 1: outline of the methodological approach

2. Trends and Scenarios

2.1 Overview of the underlying trends

Science 2.0 encompass three major macro-trends:

1. Open science, including not only open access to scientific datasets but also replicability of scientific discovery (access to methods, tools, data, articles) and collaboration between scientists as from the early stage of the scientific process

2. Citizen science, with the increasing role of citizens and amateurs in the scientific process through "crowdsourcing"

3. Data-intensive science, enabled by the availability of large-scale datasets (petabyte) processed through simulation software and enabled by high performance computing.

Science 2.0 embraces the full research cycle, from conceptualisation, data gathering, analysis, publication and peer review, as shown by the figure below.

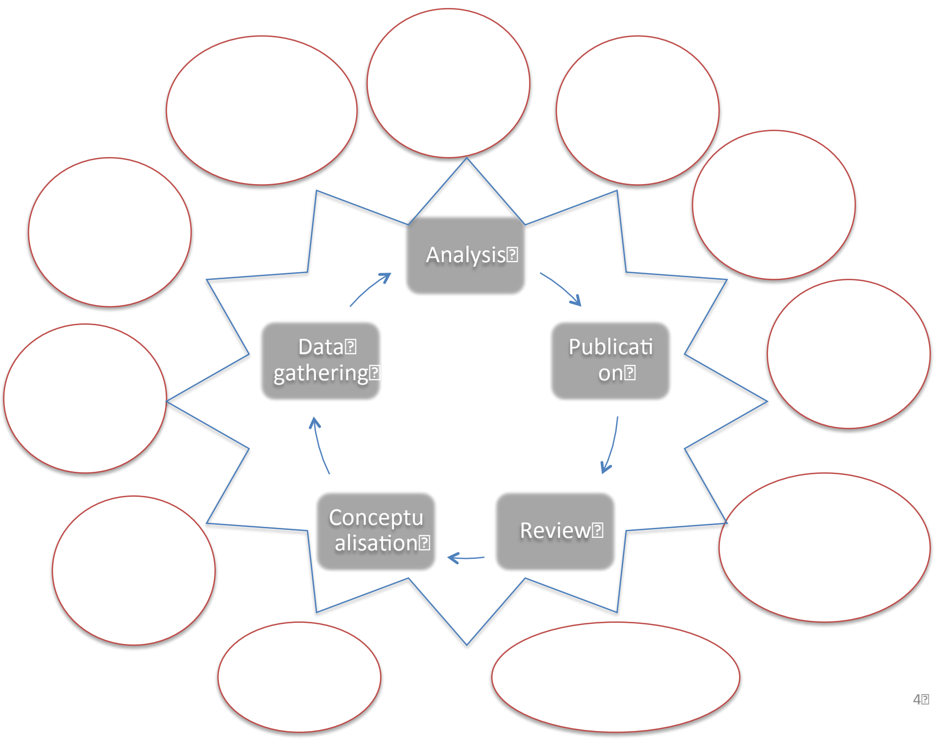

Figure 2: Science 2.0 covers the full research cycle

For each of these trends, we are already able to detect individual projects and success stories. Even more, for each one we are already able to identify fully-fledged services provided to researchers.

For instance, not only we have individual projects for citizens science, such as Zoouniverse: we also have fully fledged services such as Scistarter.com which acts as a crowdsourcing platform for citizens science; not only we have individual tools for Open Annotation such as the possibility to comment in-line in the articles provided by the online Open Access journal PeerJ: we also have the Open Annotation initiative which is developing a standard for annotation.

In other words, science 2.0 is already moving from a collation of innovative but disparate initiatives to a sustainable ecosystem of services and standards.

Yet it is clear that not all trends are developing at the same speed. Some are mature, and some are initial emerging initiatives. In the table below, we summarize the latest data about the development of these trends.

Trend	**Status**	**Data**
Pre-print	Mature	694.000 articles in arXiv (Ginsparg, 2011)
Open access	Fast growing	Exponential growth of OA journals. 8/10% of scientific output is OA (SOAP, 2011)
Data intensive	Fast growing	52% of science authors deals with datasets larger than 1Gb (Sciencestaff, 2011)
Citizen scientist	Medium growth	650K Zoouniverse users 500 similar projects on SciStarter (SciStarter data)
Open data	Medium growth	20% scientists share data 15% journals require data sharing (Beagrie et al, 2009)
Reference sharing	Medium growth	2 Million users of Mendeley reference-sharing tools
Open code	Sketchy growth	21% of JASA articles make code available 7% journals require code (Stodden, 2009)
Reproducible science	Sketchy growth	Individual researches and establishment of the reproducibility initiative (www.reproducibilityinitiative.org)

Figure 3: summary of data on Science 2.0 trends

2.2 2030: Reproducibility enabled by full integration of data and articles

Currently, data sharing is the exception rather than the rule. Only 20% of scientist declare to share their data and mainly through interpersonal relations (Wallis, Rolando, & Borgman, 2013) . It is estimated that 90% of the data sits on the researchers' own computer (DeWaard, 2013). Yet there are many examples of data being shared and integrated with articles, providing a Digital Object Identified for both the article and the dataset (see for example https://peerj.com/articles/175/).

We envisage that by 2030, each article will by default be published with access to the underlying data and code. Researchers will expect to be able dig deeper into articles and be able to directly elaborate on the available data, just as changes in the music industry were driven from a change in consumer expectations to be able to access content without limitations.

Data and code sharing will become the norm, rather than the exception. Articles, which won't provide datasets and code, will be automatically considered less reliable and robust.

There will be consistent standardized formats for publishing data, code, workflows, just as today for articles. Most importantly, there will be standardized format for integrated publishing. It will not matter if those material are published alongside the paper, in the same repository. Linked data will make this content discoverable and trackable regardless of the actual hosting. Data, code and articles as well as access metrics will be decoupled but linked.

There are already examples of this approach, from the OECD Statlink tool, which provides DOI for every chart in a report, to the "article of the future" of Elsevier, to the examples presented via the Beyond the Pdf conference and hashtag.

The integration of data with the article will be further expanded to include the actual usage of the scientific evidence in policy-making. In 2030, good governance practice will require politicians and civil servants to make explicit reference to scientific evidence integrated in any policy decisions. Not only scientific articles, but policy documents and politicians speeches will enable readers direct access to the evidence (articles and data) that justify a certain claim, ideally under the shape of meta-reviews. It will be possible to seamlessly browse between the policy documents and the underlying data through open standards, which will further enable the flourishing of fact-checking apps to discuss on the validity of the statements. Even more, this "seamless" process will enable to directly show the stakeholders positions and discussions on the specific topic – expanding towards the policy debate.

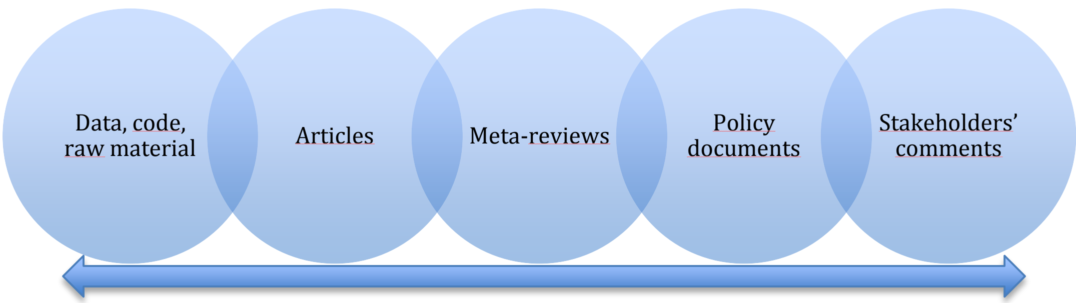

Figure 4: seamless integration of data, publications and policy documents

Moreover, we can envisage that by 2030 reproducibility will be considered a fundamental requirement for scientific publication. Research papers will be expected to contain all the necessary information to reproduce the analysis and validate the results.

Reproducibility will be formalised in a set of rules. A minimum standard of reproducibility will be required for publication (just as PLOS one does); or publications will have a "reproducibility label" assigned by third party services. Reproducible findings will be considered as higher quality, and the label will enable also non experts (e.g. policy makers) to appreciate it.

Making research reproducible is costly, in terms of documenting the experiments and curating the data, especially when it needs to be retrofitted after the research. This is why, increasingly, research protocols and methods will be formalised as templates and through tools that facilitate gathering the necessary information to enable "reproducibility by design".

The very fact of making research reproducible will automatically reduce the amount of false findings (both in bad and good faith), by reducing the incentive to cheat and by introducing more formalised analytical methods.

2.3 2030: Multidimensional, granular and instant impact factor

Today, impact of research is fundamentally measured by citation in prestigious journals. There are a number of problems about it: mainly that this implies strong delays (up to 5 years) , that *"citations are only a small fraction of how a paper is reused",* and that articles are only one of the kind of scientific outputs [Buschman, M., & Michalek, A. (2010). Are Alternative Metrics Still Alternative ?].

In 2030, impact factor will be multidimensional, granular and real time.

It will consider not just the article citation, but other measures such as actual downloads, views, book holding; number of likes, favourites, "read-later" buttons; mentions on social media and Wikipedia. These measures are already starting to become available through services such as altmetric.com and http://www.plumanalytics.com/metrics.html. It will go even further. It will measures to what extent people are reading and highlighting sections of the articles through electronic readers – as already made visible by kindle.amazon.com.

This will enable everyone to map not only article metrics, but also actual sub-sections of the article as well as of the dataset (granularity).

And all this will be made available in real time. Scientists will be informed in real time about discussion happening on the web about their work. They will be able to connect to other scientists that “think alike” and discover serendipitous connections. The impact factor metrics will become not only a management and reputation tool, but also an actual service to scientists.

Moreover, the reputation of scientists will not only be based on their papers. Reuse of scientific data, for purposes such as replication and meta-review , will be a normal part of the scientist work. By default, scientist will publish their datasets, duly curated, in common repositories for other scientists to access.

These datasets will be published as linked data, which will facilitate reuse, but will also facilitate tracking of this reuse through data citation mechanisms. You will know who has reused the datasets, and what conclusions they have drawn from it.

Tracking of reuse and data citation will enable building new reputation mechanisms beyond impact factor and H-index. These new reputation mechanism will reward scientists who produce datasets that are reused by many other scientists. Data citations will be as worthy as article citations. This will encourage further data sharing as scientists who share data will actually gain from it, in terms of career.

Just as, in the music business, streaming services such as Spotify provide a far more accurate measurement of popularity than the old “top of the pops” by actually tracking the act of listening to a song down at the level of individual, so, in science, data citations will allow for more accurate measures of the reputation of scientists.

2.4 2030: A new social contract for scientists

The availability of data, code, workflows and analysis will lead to greater collaboration. Researchers will normally work on datasets produced by others, and there will be new specialised professions emerging such as “data curator”. Researchers will be activated on specific tasks, even micro-tasks, to work on some parts of the problem to be addressed. Large, collaborative efforts (genome-like) are commonplace. This micro-tasking can already be seen in the solution to the Polymath problem, where individual scientists were contributing with ideas to solve the mathematical problem. Another relevant example is Innocentive, an open innovation platform where thousand of people (not only researchers) bring their own solution to challenging real world problem posted by companies and government. The Open Innovation paradigm will be applied to science, and scientists will outsource part of their work to specialised services such as Science-Exchange. There will be also greater involvement of micro-expertise from amateurs.

Scientists will carry out large part of their work by contributing to other people's research project. And they will be rewarded for it. The researchers' activity on these collaborative efforts will be tracked and measured, adding to their reputation. Alternative metrics platform such as ImpactStory will also include the reputation of the scientist on Innocentive.

Reputation management system will provide the fundamental incentive for collaboration between scientists, not only by rewarding the best but also by indicating the micro-expertise needed. Because reputation will include non-scientific work such as Innocentive-like platforms, there will be greater scope for non-career scientists to develop an academic career by being appointed for specific micro-courses in universities, for example using Massive Open Online Courses.

2.5 2030: Unbundling and rebundling of the scientific production

We previously argued that science 2.0 is more than open access: it affects the full research cycle, from literature review, to hypothesis creation, data collection etc. Moreover, there are today available tools and standards for most of these activities.

One of the implications of these emerging ecosystem is the decoupling and unbundling of these services.

Today, services from data repository, to data curation, to paper publication, to access metrics, are typically vertically integrated and mostly managed by the publishers. Data are published (if ever) alongside the article; metrics are provided in the same website through proprietary system.

This is not an accident, but part of the fundamental business model of publishers. It is telling that one of the justifications for Elsevier to take down articles posted by academics says the following:

"One key reason is to ensure that the final published version of an article is readily discoverable and citable via the journal itself in order to maximise the usage metrics and credit for our authors, and to protect the quality and integrity of the scientific record. The formal publications on our platforms also give researchers better tools and links, for example to data"

Findability, reputation and metrics – as well as access to data – are mentioned as key services provided by publishers. It will not be like this in 2030: there will be different providers by different services, which will interoperate through standards (possibly open standards).

This unbundling should not be seen simply from the perspective of the publishers, but also of the individual scientist. By 2030, it will not normally be the same researcher who creates the datasets, builds the programme and publishes the results. Scientists will reuse and build on the datasets and code (as well as other intermediate information) of other scientists. The gatekeeping role of the researchers will also be reduced: according to a study, 90% of the data still reside on the researchers own computers.

However, the unbundling of science (data production and publication separated from articles and from reputation measurement services; individuals vs. institutions; articles vs. journals) could not turn outas a liberation of current lock-ins where data will flow freely from researcher to researcher and from papers to data repository through interoperable open formats. There will be a vertical disintegration of the value chain, new players will enter the market. However, this openness will not last forever. The new players will try to lock-in customers in a similar way, and services will be re-aggregated around new players. For instance, it could be that data-publishers will be the new gatekeepers, which will also provide access to publications and metrics.

The recent history of the web economy shows how this re-bundling could play out. Surely, unbundling weakened the current gatekeepers, such as telecom providers, newspapers, music labels. But rather than a fully anarchical, interoperable economy based on open standards and open API, new gatekeepers and walled gardens emerged under the name of “platforms”: Apple, Facebook, Google, Amazon. Some even said that “the web is dead“. The recent demise of the interoperable RSS format by Google in favour of the proprietary platform GooglePlus is a reflection of this trend. As Wolff puts it in the same feature, “chaos isn’t a business model“.

Even when interoperability is ensured technologically, lock-in could be ensured by network effects, preferential attachment or personal data ownership. The Internet, the World Wide Web, citation networks, and many social networks all are scale-free networks showing power law distribution. In other words, the rich get richer very fast.

So it is possible and likely that either for natural development or for the invisible hand of managers, future science 2.0 will not be totally unbundled and fully interoperable. Instead, it will be divided into walled gardens. Already now, we see platforms such as Mendeley, Google Scholar, Researchgate and Figshare extending their services to what could be considered a kind of vertical integration. For instance, they all try to gather the researcher's publications in one place and act as the researcher's academic identity.

Just as the web has severely hampered the financial sustainability of newspapers, so openness will weaken existing publishing powerhouses, which by the way are one of the European strengths. New players will outcompete the "European Champions" of scientific publishing.

Future Science 2.0 will then be platform based. New players will integrate the value chain and build walled platforms. It could for example be that Amazon will build a platform around the electronic book reader Kindle for scientific publishing, including reputation management. Based on the unique data they have from what people read and highlight, they will be able to lock-in researchers and provide finely grained real time reputation based on what people download, read and highlight. They will enable direct publishing (they already do) and even provide scientific crowdsourcing platform for citizen science based on the Mechanical Turk. To put it in a provocative statement: the Kindle will do to scientific publishing what the iPod did to music.

Of course, we are not predicting the name of the winner, but simply emphasizing that the new platforms could emerge from unexpected positions in the value chain. It could be data publishers such as FigShare, or reference management systems such as Mendeley. In any case, the lock-in will be based on ownership of the personal data of researchers: what they read, cite, highlight, what data they gather, what they analyse and publish. Different platforms will provide different, competing reputation measures and identity. Data publication service will provide services to the researchers such as: "Researchers who analysed this dataset also analysed these others".

As a possible result, scientific reputation will become less reliable; existing publishers will disappear or be bought (imagine, in 2030 Mendeley will control Elsevier); data interoperability will be reduced because of different standards.

3. Implications and recommendations

In this section, we summarize our analysis of the implications (both positive and negative) of the possible full deployment of the science 2.0 paradigm. We then present the final policy recommendations based on this analysis.

3.1 Implications for the productivity of science

The most immediate reflection on the scenarios outlined above is that the full deployment of science 2.0 is likely to dramatically upscale the return of investment in research by removing many of the existing bottlenecks.

One bottleneck, for instance, is the long time needed to obtain a publication. In the future, the limitation of editorial review to the assessment of a "minimum quality threshold" will dramatically speed up the publication process, accelerating the "time-to-publication" and thereby providing further incentives to scientific publication also for new authors who are typically discouraged by the long duration of the process.

The data availability "by default" will dramatically increase the use and impact of scientific work. In particular, publication of data from failed scientific experiments will speed up the solution-finding process. The open the availability of workflows and scientific notebook is expected to "cut down the time it takes to go from lab to medicine by 10-15 years with Open Notebook Science".

Large-scale open collaboration, enabled by sophisticated reputation management systems, will enable researchers to reach out and quickly identify the specific knowledge needed, thereby dramatically reducing the "time-to-problem-solving".

However, the sharing of data and other intermediate outputs will be challenging to obtain. Introducing a cultural shift towards openness and sharing will require more than funding or rigid rules: a new system of incentives and career progression will have to be in place.

There is a concrete risk that the vertical concentration or re-bundling will create walled gardens that will limit this productivity growth by limiting the reuse of data, workflows and other intermediate products. Even the simple lack of interoperability between data, annotation, workflows, metrics, references would severely hamper the productivity gains.

Furthermore, publishing data, workflows and intermediate products could prove to be an excessive workload for researchers who would have to move resources from actual research to documentation of the research.

3.2 Implications for the quality of science

It is likely that science 2.0 will have a deep impact on the quality of scientific production.

On the positive side, it will certainly be easier and faster to uncover false claims, when data are published. Cases such as the Reinhart-Rogoff error in calculation will be uncovered in a much shorter timeframe thanks to the public availability of data and code, thereby limiting also the misusing of science.

Researchers will be more careful and rigorous in making claims, because they know that any high-profile results will be scrutinized and subject to immediate reproduction. Forgery will diminish simply because of the “possibility” of reproduction and replication of research, even if few research results will actually be reproduced.

Metareviews will be easier and almost automatic, leading to a faster progression of science by building on each other results.

The large availability of interoperable datasets (including text-based datasets) will make the uncovering of unexpected correlations almost seamless and instantaneous, thereby leading to unprecedented rates of discoveries and accelerating the capacity to react to specific problems. Data-driven discovery will become the norm and inductive approaches will radically increase their importance across all scientific disciplines.

However, the difficulty in ensuring high quantity and quality of open assessment of research (because of the demise of traditional peer review) will reduce the signal to noise ratio and will make it far more difficult for important publications to emerge and have an impact.

Moreover, the emphasis on open reviews and alternative metrics such as number of tweets will reward the most “popular” articles. Researchers will tend to focus on “popular” topics and refrain from developing new research paths. Truly disruptive research, which typically needs long time spans (5/10 years) will suffer from the introduction of instant feedback loop and reputation measurement, thereby leading to a decline in the quality of long-term disruptive research.

3.3 Implications for the scientific value chain

The institutional setting of science will face a radical reorganisation.

Academic institutions will loose their power towards individual academics.

There will be more and more temporary position in academia and circulation between business, government and research positions.

The job of the researcher will change. Teaching and researching will become increasingly separated, with the increasing role of Massive Online Courses for teaching by a selected number of science superstars.

Journals will become less important than the single articles. They will loose the control over the full research phases, in particular with regard to reference management, data storage, data analysis and annotation, publication support, peer review, reputation management. Competition will open up between players at different levels of the value chain and between different business models. Data storage services (e.g. DataDryad) will expand their services to include publication, collaboration and reputation management. Reference management software (e.g. Mendeley) will provide reputation services and collaboration spaces. Devices (e.g. Amazon's Kindle) will provide data collection, collaboration, data storage, publication, and reputation services.

Peer review will be divided between the "assessment of publishability in terms of minimum quality levels", sufficient for publication, and the "importance of the publication",

At the individual level, researchers will no longer vertically integrate the work, from data collection to analysis to publication. Datasets and code will be used and reused by third parties. Researchers will be recognized and rewarded for their intermediate outputs to be used by others.

This unbundling will not lead to a "flat world". Instead, value creation and exploitation will be re-bundled through the emergence of new players and new platform. These new players could come from unexpected positions in the value chain, such as from the device level or from the reference management software or from the provision of MOOC. Key in the strategic advantage of new players will be the capacity to accumulate data from individual researchers, in terms of what they read, collect, annotate, analyse, visualize. The competitive advantage of new players will also be guaranteed by the innovativeness and solidity of their business models, as traditional business models will be destroyed.

3.4 Recommendations: how to govern the deep unbundling?

Based on the implications described above, we advance a set of recommendations for European stakeholders as a whole.

First and foremost, it is clear that science 2.0 is here to stay. The trends are deeply entrenched and fast growing, and are a reflection of more general trends that we have already seen deploying in the context of web 2.0. It is worth remembering the scepticism that surrounded the emergence of web 2.0 players back in 2005, while in the meantime these players have become some of the largest global companies and growing a healthy start-up base is being recognized as a key policy priority in all OECD countries.

Moreover, in the specific context of science, these trends and their impact directly address some of the most profound challenges faced by science today, such as the difficulty in reproducing research results.

In other words, science 2.0 deserves policy attention because of the implications it brings, and because it is more than a passing trend.

We here divide the recommendations in demand and supply side measures. By demand-side, we refer to the changes in the modality of doing science by researchers. By supply-side we refer mainly to the different players in the value chain such as publishers, start-ups, and service providers.

On the demand side, there is the need to remove the barriers for science 2.0 adoption to be possible. In particular, the sharing of data, code and intermediate products that is crucial to the advancement of science will not happen under the current system of incentive and career progression for researchers, purely founded on “publish or perish”. It is not sufficient to require beneficiaries of research funding to share data: it is necessary to introduce systemic changes at the institutional level. A stick and carrot approach is necessary in order to achieve the necessary changes in the overall institutional setting, not just in the funded programmes. Therefore, we recommend that the overall evaluation system (from scientists career to research funding) start including intermediate products (such as datasets, code, workflows) at part of the evaluation metrics. Any evaluation system founded simply on patent and publications will not simply fail to promote, but actually actively discourage a collaborative and sharing behaviour by scientists.

We also recommend supporting experimentation in developing, applying and evaluating new forms of research evaluation metrics that experiment with different indicators and incentives systems and assess their implications.

Specific initiatives for awareness raising and knowledge sharing between the existing disparate science 2.0 initiatives would also be very beneficial.

The development of science 2.0 should be seen as complementary, rather than alternative, to the existing system. It is far from demonstrated that an open, crowdsourced approach to research evaluation would function better than the existing system based on peer review and publication. Appropriate pilots and experiments should be carried out in order to support the design of the appropriate mix of traditional and open evaluation.

On the supply side, it is clear that there cannot be a defence of the status quo. Science is changing, new players are coming in, and the value chain is being disrupted, whether we want it or not. Europe is a global leader when it comes to scientific publishing (Elsevier), it has some of the most innovative start-ups such as Mendeley and Researchgate, and has therefore all the ingredients to become a leader in Science 2.0.

Europe should embrace change. It is necessary to ensure a thriving base of science 2.0 start-ups if we don't want US-based companies to draw the financial benefits. It should promote the adoption of "platform" approaches by these start-ups, enabling them to build an ecosystem together with other players in the value chain.

Experimentation of new business models should be actively promoted. This is not to say that Europe should deliberately choose to promote open access to scientific publication, but that it should ensure that European players emerge with new business models that go beyond the current "single business model" of protecting publications behind a pay wall.

At the same time, it should ensure that these emerging platforms and ecosystem do not become walled gardens and create further fragmentation. Open standard and interoperability should be actively promoted and closely watched, not only at the level of publishing data and articles but also (and crucially) when it comes to alternative metrics, reputation management and identity of scientists.

Europe should not shy away from the concept of adopting a proactive industrial policy when it comes to scientific publishing, provided that such policy embraces the values of global openness and interoperability. Another lesson we can learn from web 2.0 is that taking no action and maintaining the status quo would only lead to endangering the European leadership in the domain of scientific publishing, with negative implications for the European Research Area as a whole.

In this perspective, demand and supply-side measures are to be considered as complementary and necessary to each other. To reap the full benefits of science 2.0, it is necessary to build a strong ecosystem in Europe, which can and should become lead market of science 2.0 with both the main suppliers and the greatest adoption levels.