A selection of possible themes to be addressed in the Synthesis Report on Digitalisation of Science, Technology and Innovation

1. This document presents information on a series of themes that could be addressed by the Synthesis Report. The coverage of themes described here is not final: other subjects may be added as the work proceeds.

2. The aims of the Synthesis Report are: 1) to summarise the main conclusions of the various projects conducted by the OECD's Committee for Scientific and Technological Policy (CSTP) and its Working Parties on digitalisation of science, technology and innovation (STI), including statistics and data to be included in the relevant analytical parts; 2) to present original analysis on certain questions regarding the digitalisation of STI that cut across the various projects; 3) to present original analysis on certain questions which are important for STI but addressed by none of the projects. The Synthesis Report will also articulate with the OECD's wider horizontal project, titled "Going Digital", which addresses broader issues, some of which have a direct relationship to STI.

3. The material in the Synthesis Report could be organised in the following way: Chapter 1 – Overview and main policy suggestions; Chapter 2 – Digital technology, the practice of science, and implications for policy; Chapter 3 – Digital technology and firm-level innovation: selected themes; Chapter 4 – Digital technology, STEM skills (and the skilled technical workforce); Chapter 5 – Digital technology's effects on selected outcomes in STI; and Chapter 6 – Digital transformation of the policymaking process.

4. An outline of the possible content of the chapters is given below. This content will be added to and possibly revised as the work proceeds.

Chapter 1 – Overview and main policy suggestions

- *Opportunities presented by digitalisation, for example:*
    - ICT's are enlarging the domain of science.[1]
    - New applications of digital technology might enhance discovery and innovation in a world where, as some evidence suggests, ideas are becoming harder to find (Jones *et al*, 2017)?[2]
    - New digital technologies could assist the diffusion of innovations in novel ways.
    - Enhanced opportunities for data sharing can increase the return on investment in science.
- *Risks posed by digitalisation, for example:*
    - ICT allows ever more scientific information to be available to ever larger numbers of people, with some of this information being dangerous. This problem already exists, of course (witness recent debates about the appropriateness of publishing details of the genomes of lethal microbes). But the problem could become more acute (for instance if forms of AI could readily demonstrate how to weaponise scientific information).
- *Policy requirements, for example:*

- Public investment in fundamental research will eventually be needed to take computation past the limits of Moore's Law - via quantum or other forms of computing - and to develop specific-purpose computing.

- Norms in the governance of research systems may need to evolve.[3]

- Assessment of academic merit, and tenure processes, might make use of more suitable metrics (such as page rankings).

- Collaborative human-machine interaction requires a science to reason about the computational and human aspects of the necessary systems.[4]

- The availability of certain types of physical infrastructure (notably data repositories) will become increasingly important for firms and research organisations.

- Regulation of the conditions of access to data may be needed to avoid inappropriate data use.

- Intellectual Property Rights may need to adapt to the digital context

Chapter 2 –Digital technology, the practice of science, and implications for policy

- *ICT and the challenges of scientific information.* The volume of scientific papers is vast and growing. Scientists may have already reached 'peak reading'. And too many research papers are of inadequate quality.[5] Digital technology enabling the development of hype-driven bubbles in scientific research?[6]

- *A framework for Open Science.* The discussion will draw on a framework developed by the OECD's Global Science Forum (GSF), and the STI policy report *Research Ethics and New Forms of Data for Social and Economic Research,* published in December 2016.

- *Open and inclusive agenda setting for science and open research infrastructures*. Within the broader Framework for Open Science (see above) two areas have been selected for preliminary analysis: 1) the inclusion of multiple stakeholders in processes for defining research priorities and agendas; and 2) the use of digital platforms to promote the broader use of research infrastructures. A small number of case studies (5-8) in each area are to be analysed, and challenges and good practices are to be identified.

- *Open data and its policy requirements*. Whilst many governments are mandating open access to research data, the costs of curating this data and ensuring its long-term availability are not necessarily being provisioned for. As the volume of data increases, so these costs are increasing and a number of private sector service providers are starting to compete with publically funded repositories. This section sets out information and assesses the emerging business models across different domains of science. Recommendations and options will also be considered to inform policies for developing federated international data networks, which are an increasingly critical part of the global infrastructure for science.

- *Digitally-enabled citizen science*. Citizen science is increasingly being promoted in certain areas of science as a way of both addressing issues that could not otherwise be addressed and of promoting public trust in science. Digitalisation is rapidly changing what is feasible, enabling new approaches to crowd-sourcing and access to un-tapped intellectual resources to solve problems. However, opening up science raises new challenges in terms of preserving quality, ensuring proper attribution and ethics.

- *What are the likely/possible developments in applying big-data and AI in science over the medium term, and what policy or institutional issues might this raise*? The challenges of automating science are enormous, requiring the effective incorporation of machines in observation, hypothesis generation and experimentation, with all three steps seamlessly feeding each other.[7] Substantial recent progress has been seen in experimentation, for instance in the pharmaceuticals industry. Some, but slower, progress has occurred in automated hypothesis generation.[8] And some limited progress has taken place in large-scale observation and data collection (natural language-processing has advanced to the point at which it can automatically extract not only relationships but also context from scientific papers).[9] Such developments raise a host of questions:

  - *How will/should science cope when AI-derived research is not intuitively obvious, even if empirically correct*?[10]
  - *How will/should science cope when algorithms, even when transparent, cannot be understood by human researchers (for instance when they run to hundreds of pages)?*
  - *If successful, could recent efforts to develop AIs which explain their own decision making be an important part of the way forward*?[11] Or is it inevitable that researchers will ‘fly blind’ more often in the future?
  - *What are the possible consequences of a world in which a growing share of scientific research is performed in silico?* (for instance, does experimentation have epistemological priority or superiority over simulation, and if so, under what conditions?).

- *What untapped potential exists for using digital technologies in science and its governance*? For example:

  - Blockchain technology has potentially important uses in science (for example, among a number of conceivable uses, the authenticity of research data could more readily be verified, as changes to a database would leave a trace).[12]
  - ICT might be used to engage the networked public in novel forms of discovery. For instance, in 2011, players of an online protein-folding game - Foldit - outperformed scientists by discovering the structure of a protein involved in the Mason-Pfizer monkey virus. This discovery was facilitated by complex software that permitted visualization of protein shapes, allowing the employment of shape recognition and modification skills by persons not necessarily trained in biochemistry.
  - ICT could assist in organising, sharing and analysing huge volumes of negative scientific results.[13]
  - Opportunities for massively collaborative authorship could grow. The noted futurist Kevi Kelly has observed that science might entail experiments involving tens of authors, including machines, and be ongoing, constantly evolving (akin to Wiki entries).
  - Opportunities may also exist to use ICT to lever the collective insight of the scientific community in ways that improve the allocation of public research funds. A possible method for doing this is described in Bollen *et al* (2013).[14]

- *What barriers need to be removed, and/or what policies need to be designed and implemented, to make optimal use of the new digitally-enabled possibilities?* For example:

  - Reforms required to aspects of intellectual property rights (for instance in connection with text mining and copyright)?

Chapter 3 – Digital and firm-level innovation: selected themes

- *The role in production of transformative digital technologies including supercomputing and advanced modelling.* What approaches are being used across countries to expand access to supercomputing?
- *Does the growing criticality of supercomputing have policy implications for countries at different levels of economic development*?[15]
- *Do data localisation policies have impacts on R&D intensive firms different from / greater than, impacts on other firm types*?[16]
- The role of data access and sharing, especially the interplay between public access to data from public research, proprietary data of firms, and hybrid forms resulting from the co-operation between public and private actors.
- *How is digital technology affecting open innovation, and what are the associated policy implications*? This part of the chapter would consider, among other questions:
  - *In what ways are business models changing with digital innovations and how are these changes affecting innovation ecosystems? What role are digital platforms playing? What are the implications for large firms, SMEs, and start-ups? What are the implications in terms of core skills and capacities that businesses need to succeed in this evolving context? In what ways are public research institutions supporting the new requirements for innovation*?

Chapter 4 – Digital technology, STEM skills (and the skilled technical workforce?)

- *How is the development of digital technology changing the skills and educational needs of science*? This is a topic that many countries are concerned about.[17] Some preliminary scoping work is underway to identify from the published literature and the STI Outlook 2016 questionnaire responses what is already known and what approaches countries are taking to develop digital skills for scientists. It appears from this early work that there may be a capacity building need for skilled data managers/stewards in some countries.
- *What is good practice in fostering interdisciplinary education and research*? Digital technology augments the need for interdisciplinary education and research by connecting different fields of knowledge. At the same time, scientific breakthroughs are typically achieved by small cross-disciplinary groups working without significant hierarchy (Heinze *et al*, 2009). CSTP's recent work on the next production revolution highlighted the growing importance of interdisciplinary education. But interdisciplinary research and teaching may be impeded in some institutions by policies on hiring, promotion and tenure, as well as funding systems that privilege traditional disciplines.[18] *Does digital technology offer new opportunities for delivering STEM education*?

Chapter 5 – Digital technology's effects on selected outcomes in STI

Overall outcomes

- For instance, what are the impacts of enhanced data access on discovery?

Differential outcomes

- A huge volume of literature is currently concerned with the distributional impacts of digital technologies, in both labour and product markets. Evidence exists for technology-driven growing polarisation in labour markets and winner-take-most concentration effects in some product markets. The question arises as to whether digital technology is having comparable differentiating effects on performance and income in science and in the scientific workforce.

- What is the relationship between digital technology and the progress of research in developing countries? Are the effects likely to amplify differences in performance or could digital technology be an equalising force?

- Some evidence suggests that scientific superstars have a unique impact on how science evolves, by being leaders, prescient and disruptive.[19] The visibility of 'superstars' is amplified by digital technology. Might the superstar effect be more pronounced as a consequence of ICTs?[20]

- Paula Stephan calculated the Gini coefficient for science and engineering faculty members at US doctorate-granting institutions, using salary data from the US National Science Foundation's Survey of Doctorate Recipients. The Gini coefficient more than doubled between 1973 and 2006 in most fields, with the largest increases in the life sciences. In the United Kingdom, the salaries of the highest-earning professors have been increasing faster than the average since the late 1990s.[21] But to what extent, if at all, is digital technology a determinant?[22]

Chapter 6 – Digital technology and policymaking

The digital science and innovation policy/governance (DSIP) project aims to provide CSTP with the means to make an informed assessment of the transformational potential and possible pitfalls of DSIP, and to identify possible good practices and make recommendations. A major part of this chapter would distil the main findings of the DSIP project, thereby covering the following:

- *Snapshot of the overall DSIP landscape.*

- *DSIP's promises, uses, risks and challenges* (based largely on survey and interview findings): What specific roles are DSIP infrastructures supposed to play?; What does the wider institutional and policy landscape look like?; What's happening in practice?; and, What are the main challenges for developing, utilising and sustaining these infrastructures?

- *Standards and interoperability*: What is the interoperability challenge? Unique persistent and pervasive identifiers (e.g. ORCID, ISNI); Other standards: CASRAI, CERIF, VIVO, standards specific to national admin infrastructures; Accessing proprietary data / IP issues; What can be learned from the standards literature more generally?

- *DSIP and research assessment*: What potential impacts could be had on accountability and learning? The promise of altmetrics? Limits, dangers and pitfalls, and guidelines on good practices.

- *The roles of the private sector in DSIP*: Main roles performed, and recent developments; Proprietary systems, emerging platforms and competition issues; Business models and their sustainability; Future prospects for private sector involvement, considering the benefits, risks and challenges.

- *Future prospects and scenarios for DSIP scenarios* (also incorporating the project's international aspects): Drawing together the main drivers of change in the DSIP area and forecasting their evolution and impacts over the next decade; Outlining 3-4 future scenarios (10-year time horizon) for the DSIP landscape that could be explorative or normative; The international aspects and policy options in each scenario.

In addition, Chapter 6 will draw on the main findings of ongoing work on the implications of digital technologies for innovation policy. Accordingly, the chapter will review:

- What innovation policy responses are needed to foster innovation in the context of the digital transition? How, if at all, should innovation policy respond to more rapid changes brought by digital technologies, increased risks and uncertainties, the global nature of markets for innovation and the increased role of digital platforms?

- What types of instruments have been put in place by different countries to create vibrant digital innovation ecosystems? What role do legal and regulatory changes (IPR, competition, standards, etc.) play? What measures have been adopted to foster industry-science knowledge diffusion in the context of the digital transition?

---

[1] For instance, advances in computing enabled the decoding of the human genome. In another example, the first trillion-frame-a-second camera was created, in Japan, in early 2014. Called STAMP, for Sequentially Timed All-optical Mapping Photography, the new camera technology could permit new ways of exploring complex ultrafast phenomena (such as chemical explosions). Read more at: https://phys.org/news/2015-04-trillion-frame-per-second-camera.html#jCp. More generally, Technical development in the tools of science - scientific instrumentation – has been critical to advances in science. In his book 'Imagined Worlds' the physicist Freeman Dyson observes that there have been seven concept-driven revolutions in science during the past 500 years. During roughly the same period there have been around 20 tool-driven revolutions. Today, ICT represents the prime example of an evolving general-purpose technology that also serves as a scientific tool.

[2] For instance, ICT is allowing the focusing of collective intelligence and knowledge in pursuit of scientific problem solving in novel ways. For example, Polymath is an open forum in which professional and amateur mathematicians collaborate to solve long-standing problems in mathematics.

[3] For instance, new arrangements might need to be found for code review. Code review as part of peer review prior to publication is reported to be problematic today, as it is time-consuming and many researchers lack the skills required. Arriving at workable norms around code review will likely become more important as code becomes more central to science.

[4] Delivering this science is the core research objective of the Oxford-based research programme ORCHID.

[5] In 2016 there were more than 1.2 million new papers published in the biomedical sciences alone, bringing the total number of peer-reviewed biomedical papers to over 26 million. However, the average scientist reads only about 250 papers a year (http://www.nature.com/news/scientists-may-be-reaching-a-peak-in-reading-habits-1.14658). And by some measures the quality of scientific literature has been in decline. Some recent studies have found that the majority of biomedical papers were not reproducible (http://www.nature.com/nature/journal/v483/n7391/full/483531a.html).

[6] Pedersen,D.B., V.F.Hendricks (2013), "Science Bubbles", Philosophy & Technology (November 2013). http://link.springer.com/article/10.1007/s13347-013-0142-7

[7] Ahmed Alkhateeb, "Can Scientific Discovery be Automated ?", The Atlantic, April 25th, 2017, https://www.theatlantic.com/science/archive/2017/04/can-scientific-discovery-be-automated/524136/

[8] Ibid.

[9] https://nlp.stanford.edu/pubs/gupta-manning-ijcnlp11.pdf

[10] See Steven Strogaz (on the 'end of insight'): https://www.edge.org/response-detail/11385

[11] DARPA is funding 13 different research groups, working on a range of approaches to making AI more explainable. See Knight, W. (2017), "The US Military Wants its Machines to Explain Themselves", MIT Technology Review, March 14th. https://www.technologyreview.com/s/603795/the-us-military-wants-its-autonomous-machines-to-explain-themselves/?set=603851

[12] See Bartling, S (2017) 'Blockchain for Open Science and Knowledge Creation' https://zenodo.org/record/401369#.WS1qxE2weos.

[13] Much published scientific data results from experiments that are not reproducible. Some hold that this is the result of bias towards positive results in science. This problem can be addressed without further development of

---

ICTs – there already exists a Journal of Negative Results in Biomedicine – but information on negative results will be vast, and its treatment enhanced by advanced ICT.

[14] Bollen, J, D.Crandall, J.Damion, Y.Ding and K.Börner (2014), "From Funding Agencies to Scientific Agency: Collective Allocation of Science Funding as an Alternative to Peer Review", EMBO Reports 15 (1): 1-121. DOI: 10.1002/embr.201338068

[15] Even for developing economies, might it be worth having a backbone network of supercomputers? The question arises of whether there will be sufficiently sophisticated applications in production. But, the supercomputers could find initial applications in research and science, and then find applications later in industry. This is what happened in Japan, where the first industrial uses were in aerospace, automotive and nuclear. Substituting supercomputers with the cloud, some experts contend, only works when applications are only needed occasionally. If the industry or scientific applications are regular or continuous then the cloud is likely to be too expensive.

[16] https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost

[17] In the United Kingdom, there is an overall shortage of students in STEM subjects and acute shortage of teachers of computer programming and engineering (http://theconversation.com/maths-anxiety-is-creating-a-shortage-of-young-scientists-heres-a-solution-58889). The UK government in 2015 offered financial incentives for students to become teachers in maths and physics. It is reported that 52% of New York schools cannot teach physics owing to teacher shortages. The UK Government has been able to recruit only 70% of the required number of computer science teachers into the profession (https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/270/270.pdf).

[18] Solutions to a mismatch between evolving research modes and institutional norms are likely to emerge from individual institutions themselves and the effects of inter-institutional competition. However, policy may also play a role. For instance, peer review practices have a bearing on the way that public agencies allocate funding for multidisciplinary research. In this connection, therefore, it could be valuable to collect data on and assess: the peer review procedures and mechanisms in scientific agencies and how well these support inter-disciplinary science; and, the challenges that inter-disciplinary science poses to current peer review processes (for instance in recruiting a large enough pool of reviewers while avoiding conflicts of interest). More broadly, beyond just peer review, value could be had in assessing how practices adopted across research institutions, private and public, enable or hinder effective interdisciplinary research. These practices could operate at the level of teams, departments or entire institutions, and range from tenure-related conditions, to organisational structures, to the role of ICT infrastructure. Among other things, it could be valuable to document the distinctive practices of institutions that have proven successful in fostering inter-disciplinary collaboration, such as Stanford's Bio-X.

[19] http://www.enid-europe.org/conference/abstract%20pdf/Klavans_Boyack_superstars.pdf

[20] See also Xie (2014), who contends that "…the basic structure of inequalities in science has remained unchanged, their intensities and mechanisms may have been altered by recent forces of globalization and internet technology." Xie, Y. '"Undemocracy": inequalities in science' *Science* 23 May 2014: Vol. 344, Issue 6186, pp. 809-810 DOI: 10.1126/science.1252743

[21] http://www.nature.com/news/science-s-1-how-income-inequality-is-getting-worse-in-research-1.20651

[22] And is income inequality a bad thing in a world where talented scientists can switch to the business sector if remuneration is deemed inadequate ?