



High Level Panel on Future Data Space

Date: 19. 1. 2022 from 13.00 to 17.00 UTC (14.00 to 18.00 CET)
Register here for participation: <https://ucsd.zoom.us/j/92302933025>

Organisers: George Strawn, Peter Wittenburg (FDO Forum)
Moderators: Christine Kirkpatrick, Dimitris Koureas (FDO Forum)

Panel Participants:

Prof. Dr. Beth Plale, Indiana University, USA
Prof. Dr. Luciano Floridi, University of Oxford, UK
Prof. Dr. Paolo Budroni, TU Vienna, Austria
Prof. Dr. Jürgen Renn, MPI for the History of Science, Germany

Commenters:

Dr. Jean Claude Burgelman, Free University Brussels, Belgium
Prof. Dr. Sarah Nusser, Iowa State University, USA
Prof. Dr. Sabina Leonelli, University of Exeter, UK
Dr. Debora Drucker, Embrapa, Brazil

Purpose

Large sums of money, and equally great ambitions, are already being directed towards developing data/research infrastructures, for science and for industry, and there is no doubt that much additional awareness, new insights and new types of services and tools will emerge. Despite these huge investments there is no agreed view on how the future data space should be organized, what its key pillars should be and how access to data will be managed and facilitated. EOSC [1] for example is based on the FAIR principles [2] and a distributed service landscape, with further specifications currently being derived by expert Task Forces. Similar to the ESFRI process [3], NFDI [4] relies on a process of discipline-driven infrastructure building, while recognizing that this leaves a gap to be bridged with respect to common services and standards. NIH Commons [5] a conceptual framework for a digital environment was designed to allow efficient storage, manipulation, and sharing of research objects. Meanwhile, big industry is defining strategies for offering services on data based on proprietary binding mechanisms, seriously hampering innovation.

This panel will initiate a discussion across initiatives about major organizational principles and key pillars, and is intended as the first of a series of meetings on this topic. Its goal, therefore, is to identify major aspects that need to be considered when examining the emerging future global data space. For this purpose, we invited four “thinkers” whom we know dare to look ahead without being bound by current projects and political considerations.

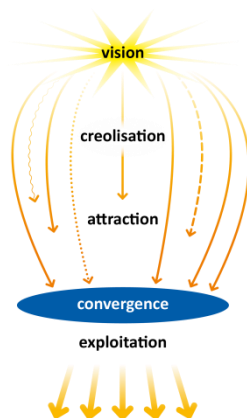
Format

The panelists will present their ideas about essential aspects of the future data space. Then well-known experts from different backgrounds will comment on these ideas and give the panelists the chance to respond. Finally, the floor will be opened to the audience to raise questions, present views and give comments.

This online meeting will take place from 13.00 to 17.00 UTC. including a break. The panel and discussions will be recorded, and all documents will be made available.

Background

In 2018, Wittenburg and Strawn compared the emergence of the Future Data Space with that of other large infrastructures which had revolutionary impacts on all sectors of societies (electricity, the Internet) [6]. When comparing these infrastructures' development steps, we found an overarching pattern which is illustrated in the diagram below. A vision materializes at a certain moment in time and is followed by a phase of creolization where increasing number of experts exploit the space of possible approaches and their implications (technical, sociological, etc.). This creolization has as consequence an increasing fragmentation and heterogeneity which needs to be overcome by a phase of attraction which finally leads to widely agreed convergence towards simple core standards. These agreements then lead to a wave in exploitation to the benefit of societies and economy.



We believe that the Future Data Space will have similar effects on all sectors of societies, especially when we keep in mind that

- data is increasingly important for the sustainable development in all areas of (human) life in the Anthropocene
- cross-disciplinary data integration and processing with AI methods will dominate the scene in research, industry, public in the coming decades
- about 80% of data project time in all sectors is wasted with data wrangling due to non-FAIR data
- many researchers, companies and citizens are excluded from using data for useful purposes, especially in developing countries
- there are no clear views on rights on data and no built-in security mechanisms hampering fair data sharing and reuse.

Consequently, we see that many countries, regions and companies with sufficient financial and human resources are investing large sums in developing comprehensive data/research infrastructures. As examples, we can refer to European Strategy Forum for Research Infrastructures (ESFRI), European Open Science Cloud (EOSC), German National Research Data Infrastructure (NFDI), NIH Commons (now discontinued) and GAIA-X [7] to just mention a few. We realize that all these approaches have different foci, development strategies and interests.

Questions of Relevance

We are listing a few questions that could be addressed during the panel knowing that there will be other important questions as well:

- Is this transitional phase we are experiencing historically unique, or can we learn from history? If so, in what way can we draw conclusions about how the future data space should look like?
- Is the assumption correct that the data space we are imagining is definitively a global one? If so, (1) can we imagine coming to global agreements via evolutionary processes given all the different national and industrial investments or will disruptive steps be necessary? and (2) how will we want to anchor ethical interests, economic interests, etc. in such a global space? (3) if not: what alternative scenarios are conceivable here?
- What kind of agreements will be necessary to enable a phase of cross-border (countries, organizations, companies) use which incorporates as a central concept the sustainability of life and which respects rights on resources and investments?
- Will the overall phenomena of “acceleration” and “rationalization”, which we observe in infrastructure development, play out again, and if so what kind of measures need to be taken to enable data to maintain societal stability?
- Which kinds of mechanisms need to be in place to prevent a “digital dark age,” a phase where paper as basis of our social and scientific memory is being replaced by digital mechanisms?
- “Data Science,” introduced as a 4th dimension [8], will increasingly dominate. What kind of risks and dangers are inherent in the approach, what kind of mechanisms need to be implemented to counter the risks, and how to democratize access to data and advanced tools?
- Previous large cross-border infrastructures were based on simple standards which were agreed upon (50/60 Hz AC, TCP/IP, HTTP, etc.). Do we need yet another simple standard to establish the global data space we envisage and would FAIR Digital Objects as self-standing entities incorporating all relevant information be a mechanism the right level to play this role?
- Is democracy an obstacle or an advantage in the emergence of a global data space?
- Who will be the owner of the Global Data Space and which authority decides on the rights and obligations of citizens with respect to the use of the Global data Space?

Panellists



Paolo Budroni

Since 2019 Dr. Paolo Budroni is senior researcher at TU Wien and Head of the «EOSC and International Liaison Office» based at TU Wien Library. He is also Chair of the e-Infrastructures Reflection Group, as well as Coordinator of the Austrian EOSC Mandated Organisation. He is also member of permanent staff of the University of Vienna (since 1991), and currently on a long-term sabbatical.

He holds a PhD in Philosophy, Art History, and Romance Philology (University of Vienna, 1986). In 1988, he completed his education degree in Foreign Trade at the Vienna University of Economics and Business (WU). He also holds a postgraduate degree in European Integration for Public Administration,

received at the Austrian Federal European Academy in 1996. 2002 – 2004 he was professor for marketing at the graduate level at the Università degli Studi di Perugia (Scienze della Comunicazione).



Luciano Floridi

Luciano Floridi is Professor of Philosophy and Ethics of Information at the University of Oxford, and Professor of Sociology of Culture and Communication at the University of Bologna. His areas of expertise include digital ethics, the ethics of AI, the philosophy of information and the philosophy of technology, topics on which he is an internationally renowned authority and has published more than 300 works. He is deeply engaged with emerging policy initiatives on the socio-ethical value and implications of digital technologies and their applications, and collaborates closely on these topics with many governments and companies worldwide.



Beth Plale

Beth Plale is the McRobbie Professor of Computer Engineering in the Luddy School of Informatics, Computing, and Engineering at Indiana University. Plale additionally serves as the Executive Director of the Pervasive Technology Institute and Director of the Data To Insight Center at Indiana University. Plale's research interests are in AI knowledge representation, safe and trustworthy AI, data management, data provenance, and cloud computing. Plale served at the US National Science Foundation (2017-2021) in a policy position in open science. She is one of the founding members of the Research Data Alliance (RDA), and served as inaugural chair of the RDA Technical Advisory Board. She helped found the IU Center of Excellence for Women and Technology (CEWIT).



Jürgen Renn

Jürgen Renn is a German historian of science, and since 1994 Director at the Max Planck Institute for the History of Science in Berlin. He is honorary professor for History of Science at both the Humboldt-Universität zu Berlin and the Freie Universität Berlin. In addition, he has taught at Boston University, at the ETH in Zurich and at the University of Tel Aviv. He has held visiting positions in Vienna, Bergamo, Pavia, and at CalTech. In 2019 he was Zeeman Fellow at the University of Amsterdam and in 2019/2020 NOMIS scholar with his project Leonardo's Intellectual Cosmos. From 2017–2019, Jürgen Renn served as Chair of the Humanities Sciences Section of the Max Planck Society. Among his most recent publications is *The Evolution of Knowledge: Rethinking Science for the Anthropocene* (2020, Princeton University Press).

Commenters



Jean-Claude Burgelman

Jean-Claude Burgelman is professor of Open Science at the Free University of Brussels. He retired on 1-3-2020 from the European Commission as Open Access Envoy and head of unit Open Science at DG RTD. He and his team developed, since 2014, the EC's policies on open science, the science cloud, open data and access.

He joined the European Commission in 1999 as a Visiting Scientist in the Joint Research Centre (the Institute of Prospective Technological Studies - IPTS), where he became Head of the Information Society Unit. In January 2008, he moved to the Bureau of European Policy Advisers (attached to the president of the EC) as adviser for innovation policy. Since 1-10-2008, he joined DG RTD, as advisor and then Head of Unit in charge of top level advisory boards like the European Research and Innovation Area Board, the Innovation for Growth Group and the European Forum for Forward Looking Activities. All these groups contributed to the design of the Framework programs for research and innovation of the European Union.

Till 2000 he was full professor of communication technology policy at the Free University of Brussels, as well as director of its Centre for Studies on Media, Information and Telecommunication and was involved in science and technology assessment. He has been visiting professor at the University of Antwerp, the European College of Bruges and the University of South Africa and sits on several academic journals. He chaired the World Economic Forum's Global Agenda Council on Innovation and was a member of its Science Advisory Committee.

He recently joined the Board of Directors of DONA and became the editor in chief of Frontiers Policy Lab. He acts as an expert for South African's open science policies.



Sarah Nusser

Sarah M. Nusser is professor emerita of statistics at Iowa State University and research professor at University of Virginia's Biocomplexity Institute. She previously served as vice president for research at Iowa State University and director of ISU's Center for Survey Statistics and Methodology. Nusser is actively involved in US efforts to promote open science, transparency, and public access to research data. Nusser serves as chair of the US National Academies Board on Research Data and Information, is senior

fellow with the Association of American Universities on its Accelerating Public Access to Research Data initiative, and serves on the US National Institute of Standards and Technology's Research Data Framework Steering Committee.



Sabina Leonelli

Sabina Leonelli serves as the Co-Director of the Exeter Centre for the Study of the Life Sciences (Egenis), where I lead the Data Studies research strand; theme lead for the "Data Governance, Openness and Ethics" strand of the Exeter Institute for Data Science and Artificial Intelligence (IDSAI); and Turing Fellow at the Alan Turing Institute in London. I am also Editor-in-Chief of the international journal *History and Philosophy of the Life Sciences*, together with Professor Giovanni Boniolo, and Associate Editor for the Harvard Data Science Review. I serve as External Faculty for the Konrad Lorenz Institute for the

Advanced Study of Natural Complex Systems and hold a Honorary Professorship at the School of

History of the University of Adelaide and a Research Fellowship with the Stazione Zoologica Anton Dorhn in Naples. I have been elected Fellow of the Academie Internationale de Philosophie de la Science, the Academia Europaea and the Royal Society of Biology.



Debora Drucker

Debora Drucker is a research data management specialist at Embrapa Digital Agriculture, one of the Brazilian Agricultural Research Corporation Research Centers. She is one of the co-chairs of the Research Data Alliance IGAD Community of Practice (Improving Global Agricultural Data) and Professionalizing Data Stewardship Interest Group and an expert at the Data and Knowledge Task Force of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). She holds a degree in Forestry from University of São Paulo, Master in Ecology at the National Institute for Amazonian Research and Phd in Environment and Society from University of Campinas. She developed data management projects at INPA, Unicamp and Embrapa and is involved with GO CHANGE coordination in Brazil and with the GO FAIR Brazil Agriculture and Biodiversity Thematic Networks.

Reading Material

Here are a few references and pointers to reading material.

L. Floridi: 4th Revolution: How the Infosphere Is Reshaping Human Reality. Oxford University Press

L. Floridi: The Philosophy of Information. Oxford University Press

J. Renn: The Evolution of Knowledge: Rethinking Science for the Anthropocene. Princeton University Press

[1] https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en

[2] <https://www.nature.com/articles/sdata201618>

[3] <https://www.esfri.eu/>

[4] <https://www.nfdi.de/en-gb>

[5] <https://pebourne.wordpress.com/2014/10/07/the-commons/>

[6] P. Wittenburg, G. Strawn: Common Patterns in Revolutionary Infrastructures and Data; <http://doi.org/10.23728/b2share.4e8ac36c0dd343da81fd9e83e72805a0>

[7] <https://www.data-infrastructure.eu/GAIA/Navigation/EN/Home/home.html>

[8] T. Hey, S. Tansley, K. Tolle: The Fourth Paradigm: Data-Intensive Scientific Discovery; Microsoft Research, ISBN: 978-0-9825442-0-4