



High Level Panel on Future Data Space

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

Organisers: George Strawn, Peter Wittenburg (FDO Forum)
Moderator: Prof. Dr. Beth Plale (Indiana University)

Panel Participants:

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. Robert Jan Smits (Eindhoven University)
Prof. Dr. Sarah Nusser (Iowa State University)
NN (tba)

Purpose

Large sums are already being invested in developing data/research infrastructures in the research area and in industry with great ambitions and there is no doubt that much additional awareness, new insights and new types of services and tools will emerge. However, 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 organized. EOSC [1] for example is based on the FAIR principles [2] and should be based on a distributed service landscape. Further specifications are planned to emerge from expert meetings. Similar to the ESFRI process [3], NFDI [4] relies on a discipline driven infrastructure building process, however, recognised in the meantime the gap with respect to common services and standards. Under the leadership of Phil Bourne, NIH Commons [5] a conceptual framework for a digital environment was designed to allow efficient storage, manipulation, and sharing of research objects. Also, big industry is defining strategies for offering services on data based on proprietary binding mechanisms, seriously hampering innovation.

This panel is meant to initiate a discussion across initiatives about major organizational principles and key pillars, i.e., it is a first action item on an intended series of meetings about this topic. Its goal is therefore to identify major aspects that need to be considered when brainstorming about the emerging future global data space. For this purpose, we invited 3 “thinkers” whom we know dare to look ahead without being bound by current projects and political considerations.

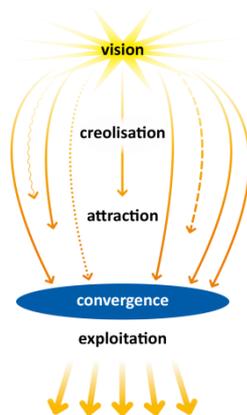
Format

For this panel we will ask the panelists to present their ideas about essential aspects of the future data space. Then we will ask some well-known experts of different background to comment on these ideas and give the panelists the chance to respond. Finally, we will open the floor to the audience to raise questions, present views and give comments.

The meeting will be an online meeting from 14.00 to 18.00 CEST including a break. The meeting will be recorded, and all documents will be made available.

Background

In 2018 Wittenburg & Strawn compared the emergence of the Future Data Space with that one of other large infrastructures which had revolutionary impacts on all sectors of societies (electricity, Internet) [6]. When comparing their development steps, we found an overarching pattern which is illustrated in the diagram. 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 is finally leading to widely agreed convergence to 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) exploitation which has as central concept the sustainability of life and which respects rights on resources and investments?
- Will the overall phenomena of “acceleration” and “rationalization” which we can observe be stressed 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 “dark digital 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 are dreaming of and would FAIR Digital Objects as self-standing entities incorporating all relevant information be the right level?
- 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). Paolo Budroni has worked in the field of cultural heritage and architecture of research information systems and knowledge management for the last 30 years. His long-term involvement in digital asset management and the provision of aligned services in the scientific community have provided him with a thorough knowledge of technical systems and the requirements of the academic world. Paolo Budroni has led as Managing Director the development of the digital archiving system for RDM of the University of Vienna, which also serves as Institutional Repository of the University (Phaidra, 2007-2016). Phaidra is now used in 5 countries and at 17 Institutions.

He was involved in the coordination and co-organisation with the European Commission and the Austrian Presidency of the Council of events like the official “Launch of the EOSC” and the Austrian Presidency event “EOSC – Austria takes initiative”. He is co- author of the „Vienna Declaration on European Open Science Cloud“. Among his areas of expertise are research data management, life cycle management, the architecture of digital asset management systems, current research information systems and cost development in life cycle management of data. Paolo Budroni is of board of the Austrian RDA National Node. He is also advising research institutions in Europe and Africa (Ethiopia) in data management policy development and alignment. As staff member of the University of Vienna and on behalf of the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Paolo Budroni has participated in February 2018 to the Launch of the Ethiopian National data Centre (NADRE <https://nadre.ethernet.edu.et/>) and has further coordinated and led the “Training seminar for research data stewardship and e-infrastructures for Ethiopian Stakeholders at Gondar University” in August 2018.

Projects acquired or involved in acquirement: OpenAIRE, Europeana Libraries, TEMPUS Project in Western Balkan Countries, LEARN, e-Infrastructures Austria, Open Education Austria, EOSC-Pillar, EOSC Secretariat Project, CO-VERSATILE (Covid urgent action)



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.



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). He holds memberships at several national and international scientific and editorial boards, among them the National Academy of Sciences Leopoldina (since 2005) and the Deutsches Archäologisches Institut (since 2019). In 2021 he was elected foreign associate at the class of physical, mathematical, and natural sciences at the Istituto Veneto di Scienze, Lettere ed Arte in Venice.

Together with his group, he undertakes research on the structural changes in systems of knowledge. Their aim is to develop a theoretical understanding of knowledge evolution, taking into account its epistemic, social, and material dimensions. As groundwork for such a theoretical approach to the history of knowledge, he has been studying some of the great transformations of systems of physical knowledge, such as the origin of theoretical science in antiquity, the emergence of classical mechanics in the early modern period, and the revolutions of modern physics in the early twentieth century. In addition to this longitudinal perspective on the evolution of knowledge, he and his collaborators have developed a transversal approach, studying dissemination and transformation processes of knowledge across cultural boundaries, as well as processes of globalization and the historical origins and co-evolutionary dynamics leading into the Anthropocene. The nature and structures of these transformative and accelerating mechanisms and their direct impact on Earth system capacity requires highly integrative research efforts. Moreover, digitalization affects the way science is done in myriad ways. Novel forms of synoptic analysis, a new conceptual framework, new research tools, and new research practices will be required to interpret and to help mitigate and steer the grand transformations *underway*. What is needed is free access to digital data and to the structures in which this data is stored and organized, as well as to the tools and methods for analyzing this data which allow research independent of private companies and market interests. Such research will have to overcome traditional borderlines, also between the natural sciences, the social sciences, and the humanities. Many transversal connections between knowledge domains are needed to grasp the present situation and the interconnectedness of phenomena that we face. Jürgen Renn and his team have developed an approach to this research, namely, geanthropology.

Among his current projects are a project at the Berlin Institute for the Foundations of Learning and Data (BIFOLD)—a merger of the Berlin Big Data Center (BBDC) and the Berlin Center for Machine Learning (BZML): “Pathways of Exoplanet Research” will utilize machine learning techniques to comprehensively document and algorithmically analyze open access research data in order to produce an open access live machine model of exoplanet research. Together with Carsten Reinhardt and Jürgen Kocka, he heads the research program *History of the Max Planck Society* since 2014. In the course of this project, the team investigates innovative methodological approaches in the digital humanities and new tools to manage the vast complexity of research activities and big data.

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