REVIEW ARTICLE

SAFEGUARDING EUROPEAN VALUES AND GLOBAL COMPETITIVENESS IN THE DIGITAL SPACE

A Gaia-X Alpbach Lab elaboration for shaping the EU data strategy

Authors:
Wolfgang Grabuschnig, Helmut Leopold, Michael Mürling
AIT Austrian Institute of Technology, Center for Digital Safety Security

From 25 to 27 August 2023, during the European Forum Alpbach, the AIT Austrian Institute of Technology and Gaia-X Hub Austria invited renowned experts from a broad range of disciplines, including IT, data science, economics, law and human sciences, as well as policy makers from international organizations such as the EU Council, the OSCE, UNESCO, and the Slovak think tank GLOBSEC, to an Alpbach Lab for a three-day discourse on data sovereignty. The focus of the event was on the challenging question of how self-determination based on democratic functionality in Europe can be achieved in the digital space, while at the same time strengthening the innovative power of European business.

This dedicated Alpbach LAB comprised three half-day workshops, each with a specific focus. The first day was dedicated to clarifying a common European understanding of data sovereignty and the current challenges for a sustainable, democratic, and globally competitive European data economy. This subsequently served as the starting point for discussions of legal, regulatory, political, social, economic, and technical realities in digital spaces on the 2nd day, and how to establish sovereign digital infrastructures in Europe on the 3rd day.

This article summarizes the highlight of this expert discussion and formulates recommendations as guidelines to be considered for the subsequent data economy strategy development.

Digital sovereignty: Perhaps the greatest civil challenge of our time

Day 1 was spent discussing the term "digital sovereignty" from a wide variety of perspectives, and framing the meta-topic of this Alpbach LAB.

The discussions touched on the different concepts of sovereignty in the digital space between the poles of security and innovation, i.e., holistically, in terms of protecting democratic states, authorities and citizens. For companies, an important element of sovereignty is the protection of intellectual property and trade secrets as a guarantee of economic success.

Another important dimension is individual sovereignty or self-determination in the context of a fundamentally changed media landscape, and in particular the growing dependence on oligarchic, global-reaching social media platforms, the monetization of users, and commodification of personal data, as described by the US economist Shoshana Zuboff in her book "Surveillance Capitalism". The potential political (mis)use of psychometric profiling – as revealed for example through the discovery of the Cambridge Analytica scandal involving the influencing of elections – also underpins this problem. However, the wide range of perspectives on the topic always boiled down to a single common denominator, namely "the absolute need to protect human interests," in the field of tension between safety and individuals' personal liberties (meaning, above all, the inviolability of human dignity and autonomy), as well as human rights and freedoms as foundation of a democratic society.

Starting from the threats to data sovereignty against the backdrop of new challenges in the digital space as highlighted in the discussion, including the power of platforms to shape opinions, the group defined several urgent fields of action for designing tomorrow's digital space, which essentially always boil down to find a necessary balance in digital power dynamics. Important approaches are: (i) making social media platforms more responsible and liable with respect to content moderation and curation with hopes for better enforcement via the Digital Services Act; (ii) overcoming the European Union's legislative limitations in the field of media and data regulation by introducing new governance structures, such as the new Data Governance Act or the upcoming European Media Freedom Act, in order to offer an alternative model to data monopolization and to ensure the diversity and independence of media in the EU over the long term; and finally, (iii) to better educate every digital user in terms of digital, media, and information literacy skills.

Free access to data, the free flow of non-personalized data, and strong digital freedom of expression on the Internet must be constantly balanced against rules to prevent illegal profiling practices, the partial and self-determined release of personal and protected data to participate in innovation processes, and financial support of sustainable stakeholders, to avoid undermining democratic mechanisms and principles in the public digital space or to even establish democratic practices in the first place.

Finally, data sovereignty also means that the rule of law in the digital sphere must be upheld to the same extent as in the analogue world, possibly by a democratically legitimized authority operating in an international context. This is the only way to deal with the phenomenon of disinformation, hate speech or cybercrime and to open possibilities for prosecution. Enforcement of laws and obligations is also a precondition for alternative economic models not relying on data monopolies.

European imperatives: media freedom and security, human dignity by design, lean regulation

On the second day of the workshop, the interdisciplinary panel of experts shed light on how digital platforms and digital media construct digital realities. A particular focus was placed on ethical principles for the design and use of artificial intelligence, as well as the development of Europe's regulatory landscape.

According to UNESCO, as outlined in its Internet4Trust initiative, global guidelines are needed to counter today's dissemination of disinformation and hate speech on social networks, and to protect whistle-blowers and journalists in the digital space. The core principles of content moderation include compliance with human rights standards and a high level of freedom of expression. It is the declared position of the OSCE that security depends on free media, and that we must protect the benefits of an open, global Internet against prevailing fragmentation tendencies if we are to achieve healthy and democratic online information systems with a positive impact on our societies in the future.

In terms of regulating the digital space in the European Union, it can be said that the EU takes a positive and exemplary position when it comes to the importance of mastering digital technologies. But the continual increase in system complexity and overlapping legislation should also be noted. It is vital to boost skills and technical capabilities in the EU over the long term. This will ensure the required scope of freedom for all stakeholders in the digital space within the framework of necessary regulation. Furthermore, this will guarantee that the authorities, the economy, and everyone in a free market has the freedom to act and can exercise choice on many levels.

The example of artificial intelligence (AI) clearly shows us the scale of our responsibility as a society in the face of current technological developments. There is always a need to balance innovative business models with the protection of individuals and democratic processes. In addition, there is a need for transparency and explainability in the function of the algorithms and the data used, a compliance structure, democratic legitimization, and a clear allocation of responsibilities. As a tool for technology development, human dignity by design should be a guiding principle for organizations. This requires a comprehensive consideration of human beings and their rights and needs, using instruments such as fundamental rights impact assessment and following ethical guidelines.

Furthermore, risk assessments and principles of responsibility, such as documentation, logging, quality management, transparency, possibilities for redress, correction, and deletion, are all important mechanisms.

All in all, digital realities must always achieve a balance between sovereignty, justice, and freedom. Here, the relationship between AI and responsibility, legitimization and sustainability are particularly important, as is the key question of what or whom we want to protect and control. This must be asked and answered repeatedly, and from case to case. Considering the polarized opinions flooding today's digital space, we must once again learn and convey that opinion is not the same as fact or truth if we are to counteract the media's loss of control in depicting real events.

Open information infrastructures, digital commons, data spaces, Gaia-X

The workshop on the final day of the Alpbach LAB placed the spotlight on sociotechnical infrastructures as the key prerequisite for fundamental data sovereignty.

After a brief insight into the anatomy of AI, the discussion turned to the importance of open science in providing access to multilingual scientific knowledge and the exchange, sharing and reuse of data, information and knowledge, and its fundamental importance for knowledge generation in general.

The conclusion was drawn that, without the open-source principle and "digital commons", the development of a sociotechnical infrastructure such as the Internet would not have been possible. The Big Science initiative BLOOM was presented as a means of explaining open cultures in the field of AI. BLOOM is an interdisciplinary research initiative to which US big tech companies have contributed human resources and software, and France its public Jean Zay supercomputer infrastructure. It involved open-source volunteers from all over the world, and culminated in the creation of ROOTS, one of the largest multilingual datasets with 46 natural languages and 13 programming languages, which was then used to train BLOOM.

What makes ROOTS special is that in selecting their datasets the Big Science research community determined the efficiency of its LLM (Large Language Model) not merely by its scale, but also by how it takes account of important social considerations. By developing an ethical charter for science based on core values such as openness, inclusion, and accountability, it avoided excessive reliance on automatic curation in terms of data source selection, governance, preparation, and analysis of data, as this fails to adequately examine the biases it has itself helped to generate. In contrast to proprietary "walled garden" models of generative AI, the multilingual dataset ROOTS used for implementing the BLOOM LLM using statistical models based on neural networks is thus designed with ethics in mind and with extensive human data curation.

BLOOM can also be used to examine the results of other generative AI tools for possible bias. The open-source approach also holds great potential for using and adapting the tool – primarily designed as a chatbot – to specific business requirements. Not to mention its potential as a polyglot LLM for use in machine translation.

Today, global sociotechnical infrastructures are crucial for the positive development of humankind in a global context. To shape our infrastructures in the way we desire, we need new and alternative approaches to policy and regulation, global standards, and informed decision-makers.

Finally, the key question of data use and the fundamental rights aspect of personal data was discussed from a sociological perspective. Discussions reflected on, what data we can possess, data as inherent expression of individual self-determination and human dignity and the responsibility that arises from data processing.

Discussions further focused on the specific concept of digital commons, which refers to the ownership and sharing of data resources within a group. Such an approach poses a particular challenge of where to draw the line: How do you strike a balance between data protection, exclusive use of data and data accessibility? Digital commons can work well with clear design guidelines, agreed rules, self-observation of member behaviour, and a graduated sanction system.

Data spaces and Gaia-X as the infrastructural basis for a sovereign data economy

In recent decades there has been a significant shift towards platform-based, data-driven applications in the economy. Following the growth of digitalization, the development of communication infrastructures, and the worldwide Internet as a fundamental communication platform over the last 30 years, we are now entering a new era of digitalization. This era is characterized by the simple, comprehensive exchange of data and its processing within data-driven ecosystems. Data spaces are a new concept for ensuring the sovereign, interoperable and trustworthy exchange of data between companies in multi-stakeholder systems. These data spaces facilitate free markets which support innovation, thereby counteracting monopoly structures that are questionable from a democratic perspective, while protecting fundamental rights and freedoms and ensuring compliance with legal obligations. In essence, data spaces create the framework for a modern data economy.

The data spaces themselves are determined by technical mechanisms designed to create trust between participants within a data marketplace by adhering to defined and agreed rules for the use and processing of data. These rules include laws, standards and agreements which must be upheld when exchanging data, thereby establishing trust between organizations and towards new stakeholders. This makes it easier for organizations and users to carry out interactions and transactions, and to share 'valuable' data with other participants. Digital identity verification and the management of access to data are other important features which ensure the trustworthy operation of the system, without the need for a central authority to act as an intermediary in all interactions between participants. This establishes trust among all data space participants and provides customers with easy and scalable access to data and services in federated marketplaces. By ensuring data sovereignty, data security and data integrity, data spaces can achieve the scale required for data-based value creation.

Open system architecture such as Gaia-X and new EU data regulation can become a model for a global data marketplace

Gaia-X is actively developing technical open-source solutions and establishing open specifications to initiate and enable the standardized development of a system architecture for federated data exchange platforms. The aim is to create effective data spaces in a variety of market segments which are designed to support seamless, sovereign, and trustworthy data exchange for new data-driven business cases. As the basic principles and processes are uniform across all data spaces, Gaia-X can also create trust between stakeholders across data space boundaries. This creates the foundations for the rapid implementation of new and innovative business models.

Nowadays, a variety of EU innovation programs support a range of data space pilot projects, proof of concept implementations of federated data marketplaces, and the testing of new, data-driven business models in a wide variety of markets and domains. Data spaces for production, for the automotive sector, for digital product passports and for energy, as well as federated

marketplaces with sensitive data for production machines or airport management are just a few examples of the intensive activities underway in the EU.

New data regulations in the EU, including the new Data Act and the Data Governance Act, rely on technical concepts and mechanisms for implementing the regulatory requirements. Initiatives such as Gaia-X offer the corresponding solutions for implementing the necessary technical functions effectively and efficiently based on uniform specifications. Gaia-X develops specifications and rules and offers open-source solutions to support companies and help them enter the new data marketplaces faster and much more cost-effectively.

Gaia-X has joined other initiatives such as the Big Data Value Association (BDVA), FIWARE, and International Data Spaces Association (IDSA) as part of the Data Spaces Business Alliance (DBSA), forming a critical mass of initiatives and expertise needed to implement modern data spaces and related open data marketplaces. As a result, Europe will be a competitive location for business and society in tomorrow's global data market and play an active role in shaping this environment.

Summary

The Alpbach LAB offered deep insights into the complex and intricate interplay between digital infrastructure and data sovereignty. The commodity concept, the definition of "digital commons" and the Gaia-X initiative are shining examples of the data landscape that is currently evolving. However, while the value of data is undeniable, its social implications are not as clear. It will be up to us to constantly explore and define the correct balance between data access and responsible data processing in line with social values.

Gaia-X Hub Austria and AIT thank all experts participating in this Alpbach Lab in summer 2023 for their valuable contribution and open discussion. The list of participants (short biography and portrait photo) at the Alpbach LAB run by the AIT Austrian Institute of Technology and Gaia-X Hub Austria can be viewed at the following link: https://www.gaia-x.at/efa23/