Strategic development: Well-managed data and advanced technologies in service of health research and innovation – CSC's view on Horizon Europe health research and strategic planning 2025 – 27

Summary

The Covid-19 pandemic highlighted the importance of health and healthcare for Europe's resilience more than ever. By supporting the underlying research and innovation, Horizon Europe can greatly contribute to citizens' health and wellbeing and Europe's ability to successfully tackle unforeseen crises in the future. Particular attention must be paid to enabling the necessary data sharing, and increasing capabilities and capacities to host and operate the data used as raw material of health research and innovation. Horizon Europe therefore must support:

- Aligned investment with a focus on building on existing, well-functioning national and European health data infrastructures, moving away from project-based, short-term and fragmented solutions
- Secondary use of health data ensuring quality, reusability and interoperability of health data spanning from healthcare to research and innovation across Member States
- Smooth integration of various data types into the European Health Data Space, curating important data sources for research
- Federated data access technologies and secure processing environments in order to foster research in areas like genomics in compliance with European legal and privacy standards
- Data altruism in practice e.g. by exploring enabling technologies
- Advancing the use of high-performance computing and artificial intelligence in the processing of health data and other sensitive data, including by prioritising security-graded operational services
- Researchers' competencies in data, HPC, AI and cybersecurity, and training them in future EHDS services, such as secure processing environments, and in international data sharing standards and practices, ensuring complementarity with the advanced digital skills priority in the Digital Europe Programme
- Facilitating the use of research infrastructures by contributing to tailor-made support and services
- Public-private partnerships around health data infrastructures such as EHDS or GDI
- Open but data-privacy-protected global partnerships with like-minded actors to pool the best expertise, infrastructure and know-how available worldwide
- Ensuring European ownership of valuable datasets by developing publicly funded research infrastructures that enable European research and innovation in the field

Secondary use of health data contributes to strategic autonomy and resilience

The Covid-19 pandemic revealed that EU's ambition to strive for open strategic autonomy is more connected to health and health and life science research and data generated in such work than first thought. The Horizon Europe Strategic Plan 2025 - 2027 must recognize that in the field of health, the foundation of the EU's resilience and open strategic autonomy lies in research excellence, world-class and well-managed health data available for innovation ecosystems, cross-border collaboration, as well as competitive education and training. By fostering excellent science, Horizon Europe can greatly contribute to citizens' health and wellbeing and Europe's ability to successfully tackle unforeseen crises in the future. One of the urgent actions is increasing capabilities and

capacities to host and operate management services for valuable data that can be used for improving the wellbeing of European citizens through research and innovation.

For example, many infections are caused by micro-organisms interacting with the human body and its immune system. With the existing research infrastructures such as ELIXIR, it has been possible to collect comprehensive COVID-19 virus data into a well-organised data portal that can be used to understand the molecular basis of infection, and inform decision-makers about the evolution of the virus variants when it interacts with the European population. More than 15 million COVID-19 viral sequences can be analysed with supercomputers, in order to increase know-how and preparedness for the next pandemic threats.

Consolidation of a long-term European e-infrastructure ecosystem for secondary use of health data

Streamlined funding opportunities and seamless coordination between European Commission DGs (especially CNECT, RTD, SANTE) are of the essence, as the next HE strategic plan must have as an underpinning priority the strengthening of long-term sustainability, accessibility and resilience of secondary use health-related data infrastructures and their services. Member States and the European Commission must focus on jointly aligning their investments with a focus on developing existing, well-functioning national and European health data infrastructures, moving away from project-based, short-term and fractured solutions. One way forward is developing coherent legal frameworks and governance models for increased interoperability[1] with the ultimate goal of creating a European single market for the secondary use of health-related data.

The European Health Data Space (EHDS) is a promising initiative towards such goals and sustainability. Sustained funding is required for key EHDS elements, such as improved data interoperability, data access technologies and secure processing environments. For best coherence, Horizon Europe should also consider supporting the EHDS board's participation in infrastructure development for the benefit of research and innovation, and their closer involvement with the framework programme and its stakeholders.

Supporting secondary use of data in the European Health Data Space (EHDS) – interoperability, reusability and secure federated access

From healthcare to research: interoperability throughout data lifecycles

The next HE Strategic Plan must enhance the conditions for cross-border health and biomedical research activities within the European Health Data Space. Existing projects, practices, policies and infrastructures related to cross-border use of health data form a practical starting point for its foundation.

On the path towards a European health data space, reusability for research purposes of data generated in healthcare needs to be ensured. Equally important is to accelerate the integration of different data types into EHDS – health data types that can be used for development of algorithms include images, genomic data, drug efficacy and toxicity data, molecular structures, and also general social and statistical data. Adding new data types into the future EHDS will

[1] Four layers of interoperability: legal, organisational, semantic and technical; see European Interoperability Framework

increase its complexity, which inevitably calls for more standardisation using open global standards. However, some technical functions like data access control technologies are shared between the prospective data types. The EHDS involves a large number of various stakeholders, and interoperable and reusable high-quality data through higher education institutions, hospitals, biobanks etc. a prerequisite for the buildup of an ecosystem of future health research, for example in training of AI algorithms properly. Reliable data resources should thus be a particular focus area for investment and Horizon Europe should support curation of health research data sources that are interoperable from the beginning, and throughout the data lifecycle.

Genomics and sensitive data as key enablers – secure federated access required

Genomic data is the digital basis of all life forms, including humans. Science and research on genomics translates into the development of the European Genomic Data Infrastructure (GDI)[2]. Areas such as personalised cancer therapies and cell-cell simulations require controlled and secure access to Member State genome data as defined in the EHDS. Aligned Member State legislations will be required to unlock the full potential of genome discovery.

Particularly in the case of health-related data, accessing it in different locations in Europe must be done in compliance with data protection rules and other legal and ethical norms. More resilience can be sought through cybersecurity including by leveraging existing solutions of secure processing environments for sensitive data, which form a key topic to explore deeper in the coming years.

At the same time the balance between security and functionality must be right – the access should also be smooth to enable research and innovation. Federated data access technologies can provide a solution fulfilling these conditions, and by supporting such tools and technologies Horizon Europe could increase the potential of the EHDS for research.

Technologies for data altruism

Additionally, the idea of data altruism needs to be implemented in practice, and Horizon Europe should support research and concept development on how to do it by means of technology. Initiatives like Europe-wide consent forms are necessary to explore. This must be done in line with MyData principles that aim to empower individuals by improving their right to self-determination regarding their personal data and sharing of it based on trust. This is likely to further facilitate the availability of research data.

Speed up convergence of data, HPC and AI in health research

To unleash the full potential of data-enabled research and innovation in the health field and to achieve maximum impact and benefits for citizens and health systems, the Strategic Plan 2025 – 27 must prioritise greater convergence of data, high-performance computing (HPC) and artificial intelligence (AI).

Utilising large amounts of diverse data aimed at complex applications such as a digital twin of the human body also require strong computational capacities. Systematic linking of data closer to HPC, excellence applications, and capabilities like handling sensitive data in HPC environments

[2] https://gdi.onemilliongenomes.eu/

CSC – TIETEEN TIETOTEKNIIKAN KESKUS OY
Keilaranta 14, PL 405, 02101 Espoo, puh. 09 457 2001,
Y-tunnus 0920632-0, www.csc.fi

CSC – IT CENTER FOR SCIENCE LTD.
Keilaranta 14, P.O. BOX 405, FI-02101 Espoo, Finland, Tel. +358 9 457 2001
VAT number FI09206320, www.csc.fi

are among key future developments for EU priorities like personalised medicine and EU Mission Cancer. Meanwhile, technologies like AI, already identified as a priority in the 2021 Strategic Plan, still present great potential for health research and must remain as a focus area.

At present, European large-scale e-infrastructures emphasize performance and scalability. As a result, very few national or European HPC systems offer security-graded operational services for processing sensitive data. However, duplicating data processing and computing service ecosystem for high-performance computing services in Europe for sensitive data is not feasible in terms of expert availability and financial cost. In order to respond to the data management needs for e.g. genomics and EU Mission Cancer, trusted research environments for sensitive data processing are needed. The future e-infrastructure needs to be able to support analysis of data subject to the data protection regulations.

All this calls for continued support for the technologies and applications themselves but also for a stronger coordination between Horizon Europe, EU4Health and the Digital Europe Programme, including EuroHPC, in order to align strategic objectives and avoid risks of siloed funding.

Boosting research and innovation through competence development and researcher-centric infrastructures

Advanced technologies also require advanced skills by those utilizing them in order to foster research and innovation. Competences in the abovementioned areas of data, HPC, AI and cybersecurity are crucial. There is a shortage of competences and skills like computational biology, scientific software analysis and development, and Horizon Europe should try to find ways to remedy this for the benefit of industry and academia in a joint manner with the Digital Europe Programme. Training for researchers will also be needed to make a most effective use of the future EHDS services, such as secure processing environments, and in international data sharing standards and practices.

Systematic competence development should not be considered a separate track, but an integrated part of advancing the actual research infrastructures. It must be planned with current and future generations of RI users in mind, as well as the scientific and technical staff in charge of the infrastructures. At the same time, human-centric approach in the design of research infrastructures also includes facilitating the use of research infrastructures by providing tailor-made support and services.

Private sector development and international partnerships

European strategic resilience will also require strong private sector health research. A large majority of the industrial health research relies on open access to data, such as ChEMBL[3] for drug development or ENA[4] for viral pathology sequence data. Sustainable innovation of European pharmaceutical applications and therapy developments can be leveraged with strategic pan-European data resources. The COVID-19 Data Portal[5] is a recent example of a critical EU-funded resource. The European Genomic Data Infrastructure is expected to deliver controlled researcher

[3] https://www.ebi.ac.uk/chembl/
[4] https://www.ebi.ac.uk/ena/browser/home
[5] https://www.covid19dataportal.org/

access to genome data. Such data infrastructures have a potential to unlock faster and more sustainable European industry development. Strengthening the support for public-private partnerships around health data infrastructures such as EHDS or GDI is advisable. Such initiatives will deliver for growing European industrial ecosystems that will further reinforce the research data infrastructures. However, the ownership of the data must be kept in Europe, underpinned by publicly funded research infrastructure as Europe's strategic asset.

CSC is equally convinced that Europe's open strategic autonomy and resilience is strengthened by global partnerships with like-minded actors. In the health area this translates into open but data-privacy-protected international partnerships to pool the best expertise and know-how available worldwide, to access world-class research infrastructures and to leverage national, European and global efforts and investments in research and innovation areas that contribute to citizens' health and well-being.

About CSC

CSC – IT Center for Science ltd. is a Finnish center of expertise in information technology providing high-quality ICT infrastructure and expert services for research, higher education institutions, culture organisations, public administration and companies. This includes e.g. development of audited, secure data processing solutions, and acting as the Finnish node of ELIXIR, a distributed infrastructure bringing together life science resources from across Europe. ELIXIR Europe consists of 23 countries under a single legal framework. It enables coordination large pan-European efforts such as European Genomic Data Infrastructure project, one of the Digital Europe pathfinders towards European Health Data Space together with the EU 1+ million genomes initiative. Together with ten other European countries, CSC also hosts at its datacenter in Finland Europe's first pre-exascale supercomputer LUMI, acquired through the EuroHPC Joint Undertaking. CSC's research infrastructures are utilised for research related to strategic policy areas of the European Union, such as climate and environment, as well as personalised medicine and health.

CSC – TIETEEN TIETOTEKNIIKAN KESKUS OY
Keilaranta 14, PL 405, 02101 Espoo, puh. 09 457 2001,
Y-tunnus 0920632-0, www.csc.fi

CSC – IT CENTER FOR SCIENCE LTD.
Keilaranta 14, P.O. BOX 405, FI-02101 Espoo, Finland, Tel. +358 9 457 2001
VAT number FI09206320, www.csc.fi