Genome Medicine

Towards a European Health Research and Innovation Cloud (HRIC)

--Manuscript Draft--

Manuscript Number:	GMED-D-19-00086R2
Full Title:	Towards a European Health Research and Innovation Cloud (HRIC)
Article Type:	Opinion
Section/Category:	I don't know, Editor will decide section
Funding Information:	European Commission (Travel grant to attend the workshop) — Not applicable
Abstract:	The European Union (EU) initiative on the Digital Transformation of Health and Care (Digicare) aims at providing conditions for building a secure, flexible and decentralised digital health infrastructure. Creating a European Health Research and Innovation Cloud (HRIC) within this environment should enable data sharing and analysis for health research across the EU in compliance with data protection legislation while preserving the full trust of the participants. Such a HRIC should learn from and build on existing data infrastructures, integrate best practices, and focus on the concrete needs of the community in terms of technologies, governance, management, regulation and ethics requirements. Here we describe the vision and expected benefits of digital data sharing in health research activities and a roadmap fostering the opportunities while answering the challenges of implementing a HRIC. For this, we put forward five specific recommendations and action points to ensure that a European HRIC: 1) is built on established standards and guidelines, providing cloud technologies through an open and decentralised infrastructure; 2) is developed and certified to the highest standards of interoperability and data security that can be trusted by all stakeholders; 3) is supported by a robust ethical and legal framework compliant with the EU General Data Protection Regulation (GDPR); 4) establishes a proper environment for the training of new generations of data and medical scientists; 5) stimulates research and innovation in transnational collaborations through public and private initiatives and partnerships funded by the EU through Horizon 2020 and Horizon Europe.
Corresponding Author:	Charles Auffray European Institute for Systems Biology and Medicine Lyon, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	European Institute for Systems Biology and Medicine
Corresponding Author's Secondary Institution:	
First Author:	Frank M Aarestrup, PhD
First Author Secondary Information:	
Order of Authors:	Frank M Aarestrup, PhD
	Abdullah Albeyatti, MD
	Willian J Armitage, PhD
	Charles Auffray
	Luca Augello, MSc
	Rudi Balling, PhD
	Nora Benhabiles, PhD
	Guido Bertolini, MD
	Jan G Bjaalie, MD, PhD
	Michaela M Black, PhD

	Niklas Blomberg, PhD
	Petronille Bogaert, MSc
	Marian Bubak, PhD
	Brecht Claerhout, MSc
	Laura Clarke, MSc
	Bertrand De Meulder, PhD
	Gianni D'Errico, PhD
	Alberto Di Meglio, PhD
	Nikolaus Forgo, PhD
	Caroline Gans-Combe, LLM
	Alexander E Gray, BA
	Ivo Gut, PhD
	Alexandra Gyllenberg, PhD
	Georg Hemmrich-Stanisak, PhD
	Lars Hjorth, MD
	Yannis Ioannidis, PhD
	Sonata Jarmalaite, PhD
	Alexander Kel
	Ferath Kherif, PhD
	Jan O Korbel, PhD
	Catherine Larue, PhD
	Mitzi László, MA
	Andrew Maas, MD, PhD
	Luis Magalhaes, MBA
	Isabelle Manneh-Vangramberen, RPH
	Edwin Morley-Fletcher, PhD
	Christian Ohmann, MD
	Per Oksvold, MSc
	Neil P Oxtoby, PhD
	Isabelle Perseil, PhD
	Vasileios Pezoulas, BSc
	Olaf Riess, MD
	Heleen Riper, PhD
	Josep Roca, MD
	Philip Rosenstiel, MD
	Philippe Sabatier, PhD
	Ferran Sanz, PhD
	Mohammed Tayeb
	Gard Thomassen, PhD
	Johan Van Bussel, PhD
	Marc Van Den Bulcke, PhD

	Herman Van Oyen, PhD
Order of Authors Secondary Information:	
Response to Reviewers:	Dear Genome Medicine Editors, I have uploaded a clean copy of the revised version of our HRIC paper, and an annotated version in which the changes and additional references are visible, addressing the nine questions as requested in the mail from your colleague Romina Andrew, also introducing minimal clarifications for the other minor points mentioned by Reviewer 1. We hope you will find this final revised version acceptable for publication in Genome Medicine. With best wishes. Charles Auffray and Nora Benhabiles on behalf of the authors

Towards a European Health Research and Innovation Cloud (HRIC)

Albeyatti A.[1], Armitage W.J.[2], Augello L.[3], Auffray C.[4], Balling R.[5], Benhabiles N.[6*], Bertolini G.[7], Bjaalie J.G.[8], Black M.[9], Blomberg N.[10*], Bogaert P.[11], Bubak M.[12], Claerhout B.[13], Clarke L.[14], De Meulder B.[15], D'Errico G.[16], Di Meglio A.[17], Forgo N.[18], Gans-Combe C.[19], Gray A.E.[20], Gut I.[21], Gyllenberg A.[22], Hemmrich-Stanisak G.[23], Hjorth L.[24], Ioannidis Y.[25], Jarmalaite S.[26], Kel A.[27], Kherif F.[28]., Korbel J.O. [29*], Larue C.[30], Laszlo M.[31], Maas A.[32], Magalhaes L.[33], Manneh-Vangramberen I.[34], Aarestrup F.M.[35], Morley-Fletcher E. [36], Ohmann C.[37], Oksvold P.[38], Oxtoby N.P.[39], Perseil I.[40], Pezoulas V.[41], Riess O.[42], Riper H. [43], Roca J.[44], Rosenstiel P.[45], Sabatier P.[46], Sanz F.[47], Tayeb M.[48], Thomassen G.[49], Van Bussel J.[50], Van den Bulcke M.[51], Van Oyen H.[52]*

Affiliations:

[1]Medicalchain, National Health Service, London United Kingdom, 'abdullah@medicalchain.com'
[2]Bristol Medical School, Translation Health Sciences, Bristol, United Kingdom 'W.J.Armitage@bristol.ac.uk'
[3]Regional Agency for Innovation & Procurement (ARIA), Welfare Services Division, Milano, Lombardy, Italy luca.augello@ariaspa.it
[4]European Institute for Systems Biology and Medicine (EISBM), Vourles, France 'cauffray@eisbm.org'
[5]Luxembourg Centre for Systems Biomedicine, Campus Belval, University of Luxembourg 'rudi.balling@uni.lu'
[6]CEA, French Atomic Energy and alternative energy Commission, Direction de la Recherche Fondamentale, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France 'nora.benhabiles@wanadoo.fr'
[7] Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy 'guido.bertolini@marionegri.it'
[8]Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway 'j.g.bjaalie@medisin.uio.no'
[9]Ulster University, Belfast, United Kingdom 'mm.black@ulster.ac.uk'
[10]ELIXIR, Welcome Genome Campus, Hinxton, CB10 1SD, United Kingdom 'niklas.blomberg@elixir-europe.org'
[11]Sciensano, Brussels, Belgium and Tilburg University, Tilburg, The Netherlands 'Petronille.bogaert@sciensano.be'
[12]Akademia Gornizco Hutnizca University of Science and Technology, Department of Computer Science and Academic Computing Center Cyfronet, Krakow, Poland 'bubak@agh.edu.pl'
[13]Ghent University, Ghent, Belgium 'brecht.claerhout@trinetx.com''
[14]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK 'laura@ebi.ac.uk'
[15]European Institute for Systems Biology and Medicine (EISBM), Vourles, France 'bdemeulder@eisbm.org'
[16]Fondazione Toscana Life Sciences 'gianni.derrico@toscanalifesciences.org'
[17]CERN, European Organization for Nuclear Research, Meyrin, Switzerland 'alberto.di.meglio@cern.ch'
[18]University of Vienna, Vienna, Austria 'nikolaus.forgo@univie.ac.at'
[19]INSEEC School of Business & Economics, Paris, France 'cganscombe@inseec.com'
[20]PwC, Oslo, Norway 'Alexander.Edward.Gray2@ehelse.no'
[21]Center for Genomic Regulations, Barcelona, Spain 'Ivo.gut@cnag.crg.eu'
[22]Neuroimmunology Unit, The Karolinska Neuroimmunology & Multiple Sclerosis Centre, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden 'Alexandra.Gyllenberg@ki.se'
[23]Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany 'g.hemmrich-stanisak@ikmb.uni-kiel.de
[24]Lund University, Skåne University Hospital, Department of Clinical Sciences, Pediatrics, Lund, Sweden 'Lars.Hjorth@skane.se'
[25]Athena Research & Innovation Center and University of Athens, Greece 'yannis@di.uoa.gr'
[26]National Cancer Institute, Vilnius, Lithuania 'sonata.jarmalaite@nvi.lt'
[27]geneXplain GmbH, Wolfenbüttel, Germany 'alexander.kel@genexplain.com'
[28]Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland 'Ferath.Kherif@chuv.ch'
[29]European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany 'korbel@embl.de'
[30]Integrated Biobank of Luxembourg, Luxembourg, 'Catherine.Larue@lih.lu'
[31]'email@mitzilaszlo.org'
[32]Antwerp University Hospital and University of Antwerp, Belgium 'Andrew.Maas@uza.be'
[33]Clinerion Ltd, Basel, Switzerland 'luis.magalhaes@clinerion.com'
[34]European Cancer Patient Coalition, Belgium 'isabelle.manneh-vgb@ecpc.org'

[35]Technical University of Denmark, Kongens Lyngby, Denmark 'fmaa@food.dtu.dk'
[36]Lynkeus and Public Policy Consultant, Rome, Italy 'emf@lynkeus.com'
[37]European Clinical Research Infrastructure Network, Duesseldorf, Germany 'christian.ohmann@med.uni-duesseldorf.de'
[38]Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden 'per.oksvold@scilifelab.se'
[39]Centre for Medical Image Computing, Department of Computer Science, University College London, United Kingdom 'n.oxtoby@ucl.ac.uk'
[40]Institut National de la Santé et de la Recherche Médicale, Information Technology Department, Paris, France 'isabelle.perseil@inserm.fr'
[41]Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, Greece 'bpezoulas@gmail.com'
[42]Institute of Medical Genetics and Applied Genomics, Rare Disease Center, Tuebingen, Germany 'Olaf.Riess@med.uni-tuebingen.de'
[43]Vrije Universiteit, Dept. of Behavioural and Movement Sciences, Section Clinical, Neuro and Developmental Psychology, Amsterdam, The Netherlands'h.riper@vu.nl'
[44]Hospital Clínic de Barcelona, IDIBAPS, University of Barcelona, Barcelona, Spain 'JROCA@clinic.cat'
[45]Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany 'p.rosenstiel@mucosa.de'
[46]French National Centre for Scientific Research, Grenoble, France 'Philippe.Sabatier@imag.fr'
[47]Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain 'ferran.sanz@upf.edu'
[48]Medicalchain, London, United Kingdom 'Mo@medicalchain.com'
[49]University of Oslo, Oslo, Norway 'gardot@usit.uio.no'
[50]Scientific Institute of Public Health, Brussels, Belgium 'Johan.VanBussel@wiv-isp.be'
[51]Scientific Institute of Public Health, Brussels, Belgium 'marc.VanDenBulcke@wiv-isp.be'
[52]Ghent University, Ghent, Belgium and Sciensano, Brussels, Belgium 'Herman.VanOyen@sciensano.be'
*** corresponding authors**

Abstract

The European Union (EU) initiative on the Digital Transformation of Health and Care (Digicare) aims at providing conditions for building a secure, flexible and decentralised digital health infrastructure. Creating a European Health Research and Innovation Cloud (HRIC) within this environment should enable data sharing and analysis for health research across the EU in compliance with data protection legislation while preserving the full trust of the participants. Such a HRIC should learn from and build on existing data infrastructures, integrate best practices, and focus on the concrete needs of the community in terms of technologies, governance, management, regulation and ethics requirements. Here we describe the vision and expected benefits of digital data sharing in health research activities and a roadmap fostering the opportunities while answering the challenges of implementing a HRIC. For this, we put forward five specific recommendations and action points to ensure that a European HRIC: 1) is built on established standards and guidelines, providing cloud technologies through an open and decentralised infrastructure; 2) is developed and certified to the highest standards of interoperability and data security that can be trusted by all stakeholders; 3) is supported by a robust ethical and legal framework compliant with the EU General Data Protection Regulation (GDPR); 4) establishes a proper environment for the training of new generations of data and medical scientists; 5) stimulates research and innovation in transnational collaborations through public and private initiatives and partnerships funded by the EU through Horizon 2020 and Horizon Europe.

Background

Genomics has brought life sciences into the realm of data sciences - large scale DNA and RNA sequencing is now routine in life-science and biomedical research with an estimate of up to 60 million human genomes being available in the coming years [1, 2]. Recent innovations in medical research and healthcare, such as high-throughput genome sequencing, transcriptomics, proteomics, metabolomics, single-cell omics techniques, high-resolution imaging, electronic medical records (EMRs), big-data analytics and a plethora of internet-connected health devices fundamentally change the infrastructure requirements for health research.

Translating these new data together with clinical information into scientific insights and actionable outcomes for improving clinical care is a major challenge. As life-science and health research datasets rapidly grow larger, with an ever-increasing number of study participants required to detect meaningful but weak signals that may be blurred by a myriad of confounding biological, experimental or environmental factors, the computational resources required to process and analyse this big data increasingly outgrows the capabilities of even large research institutes. The various cloud technologies and services listed in Table 1 are based on shared commercial and private computer and storage resources that can be provided on demand to users from a large number of different institutions conducting or participating in joint projects. They have emerged as powerful solutions to the challenges of collaborating in research on genomic, biomedical and health data. Biomedical and health research has yet to fully enter the big data and cloud computing era. The HRIC, as described in this manuscript, would help facilitate this transition, providing access to larger datasets, cutting-edge tools and knowledge, as envisioned in [3]. For example, the HRIC should ease the incorporation of domain expert knowledge into systems disease maps in a format that can be both understood by all stakeholders (patients and clinicians, scientists and drug developers), and processed by a high-performance computers, thus supporting the development of innovative medicines and diagnostics [4, 5]. Cloud technologies (through e.g. Hadoop applications) makes it possible to collaborate, access and reuse data also in situations when privacy concerns or regulation prohibits remote users from downloading data – an important benefit in Europe where national regulations can differ significantly. Clouds allow bringing algorithms to the data and as such can enable data sharing and joint processing without generating unnecessary copies of the data, which comes with potential benefits for data protection [6, 7]. Clouds, additionally, enable performing computational analyses at a scale that individual institutions would struggle to manage [7]. Consequently, in the last few years the large international cancer and other genomics consortia have created specialized genomics and biomedical cloud environments, each supporting individual projects [2]. These projects have made important advances in connecting health research data across disciplines, organisations and national boundaries. For instance, in research on rare diseases, international collaborations that integrate genomic, phenotypic, and clinical data have introduced new paradigms in diagnosis and care [8]. However, a project-based, fragmented landscape will not enable access and construction of large data cohorts that are required to address novel or broader biomedical questions that were not anticipated in individual projects when collecting inform consent from the participants, nor will it provide adequate data governance and containable cost models.

Scaling and sustainably managing such solutions to support all European life scientists thus requires a coordinated action from science policy makers, funders and other actors in this complex ecosystem. Connecting Europe's health data to advance the understanding of life and disease requires that research data and analysis tools, standards and computational services are made FAIR – i.e. findable, accessible, interoperable, and reusable for researchers across scientific disciplines and national boundaries [9]. Truly enabling personalised and digital medicine across Europe and beyond will require a connected digital infrastructure for Europe's health data that supports systematic openness and integration of research data with real-world datasets (e.g. environmental monitoring data) generated inside all healthcare systems, government agencies, foundations and private organisations that will adopt it.

On the 13th of March 2018, the Health directorate of the Directorate-General for Research and Innovation of the European Commission of the EU organised a workshop to explore the possibility and challenges involved in establishing a cloud for health research and innovation, accessible by researchers and health professionals throughout Europe, in line with recommendations for a European Innovation Council and the Horizon Europe 2021-2027 framework programme [10, 11]. The cloud computing environment proposed in this manuscript builds upon the European Open Science Cloud (EOSC) initiative developed in the last few years by the European Commission [12], with a focus in the life science and medicine field. The EOSC aims at developing a trusted, open environment for the scientific community for storing, sharing and re-using scientific data and results. Overall, the authors feel that the cloud described in this manuscript, providing the biomedical and health research community with the technical infrastructure and services listed in Table 1 to support the development of innovative diagnostics methods and medical treatments, should become an integral part of the EOSC. The workshop gathered a broad range of experts from multiple biomedical research disciplines, health care, informatics, ethics and legislation, including representatives of more than 45 FP7 and Horizon 2020 (H2020) EU-funded collaborative projects. The participants explored requirements and developed a set of recommendations for a European "*Health Research and Innovation Cloud"* (HRIC) to connect researchers and health data sources in Europe [13] such that clinical data, software, computational resources, methods, clinical protocols, and publications, can be more widely and securely accessed and reused following the FAIR principles [9] than is currently possible with existing European research infrastrucures such as ELIXIR that form a network of heterogenous national nodes. For example, the HRIC infrastructure would benefit from the aforementioned advantages of cloud computing for the archiving and dissemination of health data. This paper summarises the main conclusions from the workshop and highlights five recommendations and action points to the EU and national stakeholders (Table 2). The recommendations are key issues that need to be addressed in order to link biological, clinical, environmental and lifestyle information (from single individuals to large cohorts) to the health and wellbeing status of patients and citizens over time, while making this wealth of data and information available for European health research and innovation in clinical care.

The HRIC should be built on established standards and guidelines, to foster European-wide medical research

Rationale

Sharing of data, information and knowledge represents the most important functionality in the context of a HRIC. High-level standardisation, common exchange mechanisms, interfaces

and protocols, and semantic interoperability form the foundation for widespread adoption of the FAIR principles [9] in health research. Data shared collaboratively in such health-related cloud projects are now largely standardised for the processing of genomic DNA read and genomic variant calling files. By comparison, the sharing of highly sensitive clinical and health data has thus far been much less developed hence representing a key future focus area, and numerous challenges remain with regard to sharing these data in a meaningful way.

Existing standards and guidelines

Many individual projects, in Europe and globally, have demonstrated the opportunities and the added value presented by connecting and exchanging data across countries via standardised protocols. Table 3 lists recent European projects that have developed towards clinical and health data exchange via cloud-based solutions. All those projects have developed and implemented worthy ideas that should be included in the HRIC. However, we envision the HRIC as a disease-agnostic environment, and on a larger scale than the platforms mentioned in Table 3. Moreover, the HRIC should not be tied to a single project or consortium, but should rather be under the governance of an independent body. International exchange of health research data holds tremendous potential in disease research by facilitating better investigation of disease causality and linking of genotypes and phenotypes, such as demonstrated in the Pan-Cancer Analysis of Whole Genomes (PCAWG) project, for example. An important aspect of the cloud-based data governance is that it allows sharing of data outside a consortium through a data request mechanism and a governance infrastructure that track participant consent, and data access. The cloud-based audit trail capabilities of the cloud-based research analyses that can be implemented at both data and infrastructure levels is a direct benefit to the data controllers. A standardised data model (or data access model) and/or standardised metadata models facilitate consolidation of different data sets and significantly increase the findability, the semantic interoperability and, by consequence, the reusability of data, and thus its 'FAIRness' [9, 14].

Workshop participants agreed that a first minimalistic and yet effective approach to data exchange should consist of a small number of initial online repositories containing references (e.g. links to render data sets findable) along with metadata (e.g. type and scale of content, specifications describing in what systems the data set may be stored and processed), and indications on how to gain access (e.g. requirements and point of contact). This could be designed as a metadata repository containing the metadata of data objects and information on how to gain access to them. The data objects as such may, or indeed should, be stored elsewhere.

Beyond their metadata, however, data sets are bound to greatly differ, as research projects vary largely in scope. Not only would it be cumbersome to record a large number of parameters irrelevant to the specific question addressed, but it would also be problematic from an ethical viewpoint, considering patient personal data protection aspects [15]. It therefore seems more promising to drive standardisation within research communities while looking out for opportunities for overall standardisation.

Thus, the workshop envisioned the HRIC as a distributed collection of data repositories, people and services, which together make up a framework to share and operate as a federated data commons with reproducible software, standards and expertise based on joint policies and guidelines to conduct health research as performed successfully in previous initiatives [2, 16-19]. The need for federation is also highlighted in the proposed EU action plan for 'Making

sense of big data in health research' [3]. Creating such a HRIC opens new frontiers for research and healthcare via the opportunities for strong international collaborations.
Beyond Europe the HRIC should collaborate internationally to drive the development and widespread adoption of global standards and connectivity. Ongoing initiatives exist outside Europe that are aiming at developing global standards for the secure exchange of e.g. health and medical records within health information federations while tracking completeness of the supporting data [20, 21], as well as developing guidelines for data analytics and standardised workflows. They should be considered for the basis of the HRIC as this will provide the scientific community with means of reproducibility, version control and documentation, which will be an important vehicle to drive increased standardisation and connectivity.

A European HRIC should be developed and certified to the highest standards of interoperability and data security

Rationale
The workshop participants enthusiastically endorsed the vision of the HRIC as a federated environment. Through work on the standardisation, harmonisation and integration of genomic data with other health-relevant information to optimise hypothesis-driven analyses, the major blockers for cross-border collaborations in translational research on disease prevention, treatment and management will be addressed through a federated HRIC infrastructure in which data sources remain at their location of origin and are made accessible to users through a metadata repository. Moreover, data security has to be integral to the development of the HRIC, using modern cryptology and access control techniques to ensure the protection of the patients' data contained therein.

Standards in interoperability and data security
European health systems have different ways of managing and storing health data, making the exchange of clinical data between the EU member states complex. The challenges are well illustrated by the use of eHealth records and their use as secondary research material. In a recent OECD report [22] ten countries reported comprehensive record sharing within one country-wide system designed to support each patient having only one EHR (Supplementary Table.). These countries are Estonia, Finland, France, Greece, Ireland, Latvia, Luxembourg, Poland, Slovakia and the United Kingdom (England, Northern Ireland, Scotland and Wales). In these countries, plans call for patient records to be shared among physician offices and between physicians and hospitals regarding patient treatment, current medications, and laboratory tests and medical images. Some have already implemented part or all these functionalities, while others are progressing toward it. In other countries key aspects of record sharing are managed at sub-national level only, such as within provinces, states, regions or networks of health care organisations (for example, Austria, Germany, Italy, Netherlands, Sweden, and Spain; Supplementary Table). Among them, all have implemented or are planning the implementation of a national information exchange that enables key elements to be shared country-wide. Based on the recent reports of the European Commission [23], Belgium, Malta, Portugal, Romania and Slovenia are now developing national EHR systems, leading to a total of 16 EU member states providing such services.
Within the framework of the Joint Action on Rare Cancers, an EU initiative that brings together European research centres, policy makers and other stakeholders with the aim to set the

agenda at national level, an analysis was made on the status of eHealth medical records in the EU member states. This work builds on the OECD study and complements this with information provided by the European Commission on the national laws on EHRs in the EU member states [24]. Thus all EU countries are investing in the development of clinical EHRs, but only some countries are moving forward the possibility of data extraction for research, the provision of statistics and the enablement of other uses that serve the public interest (P. Bogaert, personal communication). Countries that develop EHR systems that combine or virtually link data together to capture patient health care histories have the potential to be used also for long-term follow-up of cancer patients. Figure 1 shows how data from various sources can be integrated to provide a full picture of patients' health status over time and carry out research for patterns and anomalies in large populations using the specific combinations of data and analysis resources relevant for each research project, while ensuring compliance with security and data protection regulations.

The H2020-funded project European Open Science Cloud (EOSC)-Life [25] is developing policies, specifications and tools for the management of data for biological and medical research including aspects of eHealth data. The use of common metadata standards, developed in EOSC-Life, as a foundation for remote data discovery and access was emphasised by the workshop participants as a key enabler for the HRIC. For instance, practical and legal considerations for cloud computing of patient data, which include the responsible use of federated and hybrid clouds set up between academic and industrial partners have been put forward by early EOSC pilot projects [26].

The workshop participants emphasised that sustainability aspects are critical and must be considered from the start. To ensure that a HRIC could respond properly to emerging needs, innovation and technology changes, a distributed federated storage solution offering access to FAIR data and services should be built according to modularity principles. In particular, due to the long-term nature of a HRIC, due consideration should be given to deploying generic and modular computational methods and/or data storage management systems, while the information and communications technology (ICT) infrastructure should be flexible, portable and expandable.

The million European genomes initiative [27] is a case in point: eighteen member states have already signed the Genomics Declaration of Cooperation [28] to enable cross-border access to genomic databases and other health information. This federation of national initiatives [28] will provide secure access to such data resources in the member states to enable the discovery of personalised therapies and diagnostics for the benefit of patients. The initiative involves aligning strategies of ongoing national genomic sequencing campaigns with complementary *de novo* genome sequencing to obtain a total cohort of 1 million Europeans, accessible in a transnational framework, by 2022 [29]. The HRIC would form a basis for such large-scale, permanent collaborations.

The workshop participants recognise that ensuring a maximal data security is paramount to build and maintain trust with the European citizens. To address this issue, we suggest using modern cryptology such as blockchain to ensure data security by design, and holding regular data security assessment (e.g. through hackathons and/or commercial security audits). As shown in recent literature, the use of blockchain in biomedical research is still in its infancy [30]. However, blockchain or other advanced cryptology tools that can be used to protect data in a cloud environment could prove useful for the secure and trustworthy implementation of the HRIC [31].

The HRIC must be supported by a robust ethics and legal framework compliant with the GDPR

Rationale

Compliance with GDPR and other data protection laws, as well as enforcing an ethical usage of data is paramount to gain the general public's support and trust in the HRIC.

Existing ethics and legal framework

Unblocking the legal and administrative barriers for sharing human research data across geographical and organisational boundaries will, if the trust of research participants is preserved, pave the way for continent-scale cohorts in life science research. This will represent a significant innovation as sharing and joint analysis of sensitive data has until now been severely limited because of the different restrictions inherent to the different classes of sensitive data. By using a federated database model, with a metadata repository within the HRIC cloud encrypted environment, data security is maintained while innovative data analyses can be performed by bringing the algorithms to the data rather than centralizing the data [32]. Federation, rather than full integration of all available resources, poses an important challenge for implementation and deployment of an effective HRIC. The set-up and functioning of a HRIC requires a robust foundation of legal agreements and ethics rules and procedures, as well as security and data protection compliance protocols. Importantly these elements must be introduced during the conception and design phase as part of its governance. This is essential in order to allow the different HRIC actors providing access to their data sources, managing them within the cloud, and accessing these resources to incorporate policy requirements into the design and manage the complexity involved in implementing simple and intuitive user interfaces and project portals. This may be challenging considering the heterogeneity of health systems and health market accesses across Europe and will need the sharing of a common agreed vision across EU member states.

Ethical, societal and privacy considerations for (re)using health related data have been outlined in the Code of Practice on Secondary Use of Medical Research Data developed in the European Translational Information and Knowledge Management Services (eTRIKS) project funded by the Innovative Medicines Initiative (IMI) [33]. In addition to clear and explicit consent, explicit dissent may need to be considered for the use/re-use of data. Ultimately, each citizen and patient must be able to access her/his own data and know when and where it has been used and for what purpose. In addition, the difficult question of the business model of using those data should be discussed at various levels from ethical, societal and economical standpoints taking into account potential future developments of products and services using personal medical data. Furthermore, the goal of providing the citizens with personalised services requires technical advancements in the collection and analysis of data (e.g. in data analytics and machine learning). For this type of usage, simple consent mechanisms might not be sufficient. For example, how should a clear data-collection purpose statement be defined, if data is collected for multiple usage scenarios across a distributed/federated cloud in which actors from different geographical and legislative environments will need to interact and cooperate? Would an excessive number of consent requests minimise data provision for research or clinical applications? Another level of complexity is introduced by the heterogeneity of data protection and privacy regulations when the data originate from states within federated national health systems (e.g. Germany, Italy). Development of large-scale

European access mechanisms will require open consultation and engagement with national policymakers, patient organisations and the wider society to build the trust and confidence needed for widespread adoption and sustainable operations.

In addition to technical, ethical and legal specifications, a global integrated governance model needs to be established for the HRIC in line with that of the EOSC, regulating roles and responsibilities of all contributing institutions and users, and procedures for authentication and access control to individual resources. Principles, with specific guidelines on implementation into a HRIC environment, will need to be developed in order to manage and regulate aspects such as ownership, access, transparency, sharing, integration, standardisation of data and metadata formats, tools and frameworks, while ensuring confidentiality and sustainability. All these principles need to be developed with the overarching objective of providing benefit to and preserving the trust of patients and the general public.

Health data mostly represents sensitive data, which need to be managed to preserve the trust of patients, research participants and the general public, respect social norms and naturally comply with the rules and regulations of data protection laws, notably the EU GDPR [15]. Although the GDPR directly applies across the EU and its provisions prevail over national laws, EU member states retain the ability to introduce in their own national legislation certain derogations provided for by the GDPR itself. The GDPR also introduces the notions of 'Privacy by Design', which means that any organisation that processes personal data, must ensure that privacy is built into a system during the whole life cycle of the system or process; and 'Privacy by Default', which means that the strictest privacy settings should apply by default, without any manual input from the end user. In addition, any personal data provided by the user to enable the optimal use of a given health dataset should only be kept for the amount of time necessary to provide the intended product or service [15].

Thus, successfully linking and accessing biomedical and health data across Europe will require many different disciplines and specialists working together with a coordination effort that should encompass controlled access mechanisms to ensure compliance with privacy and data protection regulations. Data providers need logging and monitoring functionalities to comply with the GDPR and to enable tracking of data and methods within the system, controlling instances and routines that check for the adherence of predefined standards and formats to guarantee data integrity. Access mechanisms need to be developed that support the researchers, data producers and data analysts to request permissions and fulfil reporting requirements for data use in national and international research projects, which represents a significant regulatory, political and sustainability challenge [34]. These include in particular considerations about the rights of patient donors and research participants, taking into account the data protection aspects of various legal systems and local regulations. Researchers have to face differences in the understanding of the right to data protection in those different regional or national European ecosystems. There is an urgent need for standardised, usable, data-protection-policy-compliant solutions for sensitive data sharing capable of integrating and analysing health data coming from different sources, organisations and potentially from different research disciplines. These aspects are subject to ongoing discussions and debates in the EOSC initiative [35] and for instance, progress has been made in the Human Brain Project through its Ethics and Society sub-project in collaboration with the project platforms [36, 37]. Other examples of data sharing compliant with data protection policies can be found in the recent literature [38-45]. Furthermore, there is the issue of capacity with the amount of data starting to strain the infrastructure of any individual hospital

or research institute. Thus, the interplay between privacy, data security and access control on one hand and access (including cost-recovery models) to storage, computational and analysis resources on the other hand will be a defining element of the policy and technology development of a decentralised digital health infrastructure. The evolution of a cloud model that could be used in European health research will also have to take into account other specific aspects of the GDPR [15]. For instance, the European Commission intends to facilitate the free flow of non-personal data in the European Digital Single Market, and for health-related research participants it codifies the 'right to be forgotten'. It stipulates that patient donors should be able to retain control over their data regardless of technological developments. A European HRIC could be an important enabler for researchers to comply with these requirements. For example, once certain conditions are met between European and international partners, including those pertaining to data protection and use, federated and hybrid clouds could facilitate the deletion of data sets once a donor exercises her/his 'right to be forgotten', which thus could minimise the necessary transfer of large raw data sets across borders, as the deletion can be performed in the original dataset and easily propagated to the relevant federated data sources.

A proper training environment for HRIC developers and users should be established

Rationale

The workshop identified the lack of trained personnel, with solid skills in both medical and data analytical fields as one of the major bottlenecks when dealing with 'medical big data' [3].

Need for training and ideas

Indeed, effectively developing, operating and maintaining the HRIC will pose serious challenges and require the training of a new generation of data scientists able to navigate smoothly and efficiently between computational, security and medical disciplines. This includes clinical researchers, bioinformaticians, data analysts, data managers, software engineers, cloud engineers, other IT-specialists, ethics officers, and data protection specialists, which represent an essential new field of expertise. Finding professionals that cover more than one or two of the above-described disciplines is nearly impossible. Furthermore, communication between this comprehensive mix of clinical researchers, data managers and IT/bioinformatics specialists needs to be improved, requiring a governance structure well beyond that of a standard research setting. The EU should take inspiration from existing large and successful infrastructures that foster multidisciplinary teams, such as CERN [46, 47]. Thus, it is necessary to rethink the training and education of health professionals and to update them with the HRIC in mind, considering both international standards and practices for data sharing as well as national environments and regulations.

Multiple funding mechanisms are required to drive the development of the HRIC and support its broad use in research projects

Rationale

Delivering the HRIC requires an ambitious reshaping of the European landscape for health data and research through appropriate funding schemes that enable transformation of

fragmented ICT resources and project-centric solutions for data access and governance into a long-term, coherent service ecosystem that can be accessed by users transnationally.

Need for innovative public-private funding initiatives

Indeed, the HRIC needs a trusted and transparent innovation approach that recognises the importance of a clear, long-term ambition in the programme to support the participation of industry and small to medium enterprises (SME) in joint projects with a broad set of societal actions. In particular, there is a need to support the EU ICT industrial/SME innovation ecosystem in order to demonstrate the benefits in advancing data sharing, integration and analysis across Europe, for the benefit of all citizens, thus creating a foundation for attractive private investments.

For this purpose, targeted EU funding mechanisms also involving private investors need to support the development of HRIC-compliant services for data sharing and analysis in health-related research projects (i.e. through reimbursement for storage and computing costs) with incentives to reuse and extend existing infrastructures that favour national HRIC participation rather than rebuilding and fragmenting solutions. Moreover, the EU ICT industrial ecosystem must be supported in order to alleviate the risks associated with storing and sharing data across cloud systems operated by non-EU companies. The EU has established a strict privacy and ethics policy through the implementation of the GDPR legislation, binding for all operators active on the EU territory [48-51].

European funders, science policy makers and other actors need to develop mechanisms that bring together experience gained and lessons learned from a large portfolio of pathfinding projects and build on current investments, thus leveraging existing project outcomes. This requires an inclusive and integrative approach with programmes that bring together the many different actors into the HRIC, as its construction needs interdisciplinary collaboration with expertise from many disciplines e.g. economics, ICT, biomedical and health, social sciences and policy. In particular, frameworks for public-private partnerships such as IMI have shown a way to include industry in open transparent projects that also includes patients and other public bodies, SMEs and European researchers. Many of the mechanisms proposed in the Lamy report (prioritise research and innovation in EU and national budgets, build a true EU innovation policy that creates future markets, rationalise the EU funding landscape and achieve synergy with structural funds…)[52] and in the development of the Horizon Europe "Missions" [53] would also be well suited to the development of the HRIC and help bring together initiatives from the many funders, national and regional stakeholders. Other opportunities such as those proposed to be included in the Horizon Europe strategic workplan (e.g. the European Information Cloud, the European Institute of Technology and the European Council for Health Research [54, 55]) should be of interest to deploy innovations together with industry at the European level [56].

Furthermore, future programmes need to create incentives such that developed solutions are transformed into long term reusable resources and make sure that this infrastructure is deployed across the EU with development informed by ongoing research projects. The European Joint Programme on Rare Diseases (EJP-RD) [56] provides a good example of how infrastructure development can be linked with research projects at national and international levels. As in the EJP-RD, the European Strategy Forum on Research Infrastructures (ESFRI) can play a role in the development of HRIC. Two further aspects of EJP-RD are worth noting: the programme has a strong emphasis on the importance on the diverse workforce required with a training programme that stretches beyond academia and research networks to reach a

broad set of individuals in health systems and the education sector. Secondly, the research portfolio should fund broadly and recognise that successfully addressing many of the identified challenges will require a diverse portfolio of projects that avoid any artificial boundary between biomedical and health research. Horizon Europe should allow for linkages between the HRIC, ESFRI and other themes of Horizon Europe and, importantly, between HRIC and other funding sources such as the European Structural and Innovation Funds (ESIF) and the cutting-edge basic research successfully supported by the European Research Council (ERC) during the past decade [57, 58].

The HRIC should enable investments in product development for future health-care solutions as well as allow health care providers to procure such solutions. People need to be an integral part of the innovation vision where the HRIC supports a highly skilled future workforce that makes Europe attractive for locating R&D investments. The experience gained in IMI public-private partnership infrastructure projects such as eTRIKS and the European Medical Information Framework (EMIF) should be leveraged. Finally, the current and future EU Framework Programmes for Research and Innovation (Horizon 2020 and Horizon Europe) should consider mobilising funds to support novel pilot actions, pooling of data and resources across the EU, and demonstrate the benefits in advancing data sharing, integration and analysis across Europe, for the benefit of all citizens.

Conclusions, recommendations and action points

Clouds are increasingly becoming a key venue for enabling and hosting European and international collaborations, benefitting from the ability to hold data securely in a single location (or in few locations) and enable collaborative research on computational infrastructure used for analysis. In conclusion, a cloud-based federated data storage solution with interoperable data access services (see Table 1) to local repositories and modular environments that can be configured for a given use case seems to match the data needs of the EU research and medical institutions and of all other stakeholders. The choice of cloud technology provides the ability to manage rapidly growing datasets and provides users with access to the massive computational infrastructure needed for analysis. The federated, cloud based research environment - HRIC- described in this paper, would represent an added value to the entire biomedical and bioinformatics community, because single research institutes and medical institutions lack sufficient infrastructure capacity. The establishment of a transnational HRIC will allow the European research community at large to contribute to the global international leadership required to address societal and scientific challenges through transnational collaborations. In order to ensure the effective and efficient implementation of the European HRIC, the workshop participants endorsed the five recommendations and actions points presented in Table 2 to the EU and all stakeholders.

Application	A set of programs running on one or more computers allowing a user to perform a set of tasks
Cloud Application	An application running in the cloud
Cloud Computing	The delivery of IT services over a network (e.g. the Internet) by means of a combination of infrastructure, software and data hosted by one or more cloud providers using a service model

	similar to that used by traditional utility companies (e. g. water or electricity)
Cloud Federation	The combination of infrastructure, software and services from separate networks and providers, having shared access mechanisms, to perform common actions, achieve load-balancing or optimize availability or costs
Cloud Instance	A virtual server or container of resources running on a physical host computer possibly hosting several independent instances (see virtualization)
Cloud Marketplace	An online marketplace of cloud services and applications operated by a Cloud Service Provider (CSP)
Cloud Service Provider (CSP)	A company or public entity that offers cloud services to individual users or other entities
Cloud Storage	A model of data storage in which data is hosted across one or more facilities by a hosting entity or CSP and remotely accessed by users over the Internet
Container	A type of virtualized instance running on a physical host server in isolated user spaces and possibly preloaded with applications
Hybrid Cloud	A cloud computing infrastructure comprised of a mix of public and private cloud and on-premise instances and resources
Infrastructure	The combination of hardware resources (network, computing, storage, etc.) and virtualized instances supporting an IT environment
Infrastructure as a Service (IaaS)	A model of cloud computing in which a CSP provides an infrastructure of virtualized resources to users as a service over a network (e.g. the Internet)
On-Premise	It refers to infrastructure or software run on computing resources physically hosted by the entity using them
Platform	A computer system where applications can run on or can be built with
Platform as a Service (PaaS)	A model of cloud computing in which a CSP provides the infrastructure and the platforms where users can run and manage their own applications as a service over a network (e.g. the Internet)
Private Cloud	A cloud infrastructure used by a single organization either on-premise or hosted by a third-party CSP over the Internet or dedicated private networks
Public Cloud	A cloud infrastructure hosted by a CSP (or a Federation) and used by the public or multiple organizations across the Internet
Software as a Service (SaaS)	A model of cloud computing in which a CSP hosts and provides applications (software) to users as a service over a network (e.g. the Internet)
Virtualization	A technology that allows to run a software simulation of a physical computer where a full operating system and applications can be installed

Table 1 : Glossary of cloud computing terms.

Recommendation	Rationale	Action points
Provide and foster standards, good practices and guidelines necessary to establish the HRIC.	The HRIC should be supported by predefined standards, data formats, protocols and templates. The data standards and guidelines applied in the HRIC should be designed as to facilitate interoperability between the diverse health systems and policies in Europe and globally.	Suggest the adoption of data formats and architectures in policies, grant applications and projects calls, throughout the EU and its member states.
Develop and certify the infrastructure and services required for operation of the HRIC.	The HRIC should provide computational infrastructures and services, analytical and visualisation tools to all users as a platform to share knowledge, data and guidelines. Services for data sharing, security and analysis should be compliant with an EU certification system.	- In future grant and project calls related to the development of the HRIC, the EU and the applicants should commit to comply with the highest standards of security, interoperability and reproducibility. - The EU should develop a certification system to validate compliance with the standards mentioned.
Enable the HRIC to operate within an ethical and legal framework adequate for health systems.	A robust ethical and legal framework has to be developed that defines rules for privacy, security, ownership, access and usage of data within the HRIC. A federated system architecture should be preferred as it allows for comparison of data and results, while complying with EU (GDPR) and international data protection and sharing rules.	- In grants and project calls, federated and GDPR-compliant data architectures should be preferred.
Establish a proper environment for the training of a new generation of data and medical scientists.	Education and training of health professionals need to be updated with the HRIC in mind, considering both international standards and practices for data sharing as well as national environments and regulations. The EU should take inspiration in existing large and successful infrastructures that foster multidisciplinary teams, such as CERN.	- Scientists and health professional in training should be made aware of the possibilities of the HRIC. - Communication in relevant professional channels should be strengthened.

Fund public and private initiatives for the development of the HRIC through EU Framework Programmes (Horizon 2010 and Horizon Europe).	The EU and its member states should together with private investors develop a coherent, ambitious and long-term action plan supported by innovative funding mechanisms that consolidate the outcomes from the existing project portfolio into a long-term operational infrastructure.	- The EU should invest through calls and grants in order to build and consolidate the HRIC. - The existing industrial ecosystem should be supported, to remain competitive against the other actors in the world.

Table 2: Summary of the recommendations, details on the rationale and suggestions for actions points directed to the funding agencies and the actors of the field.

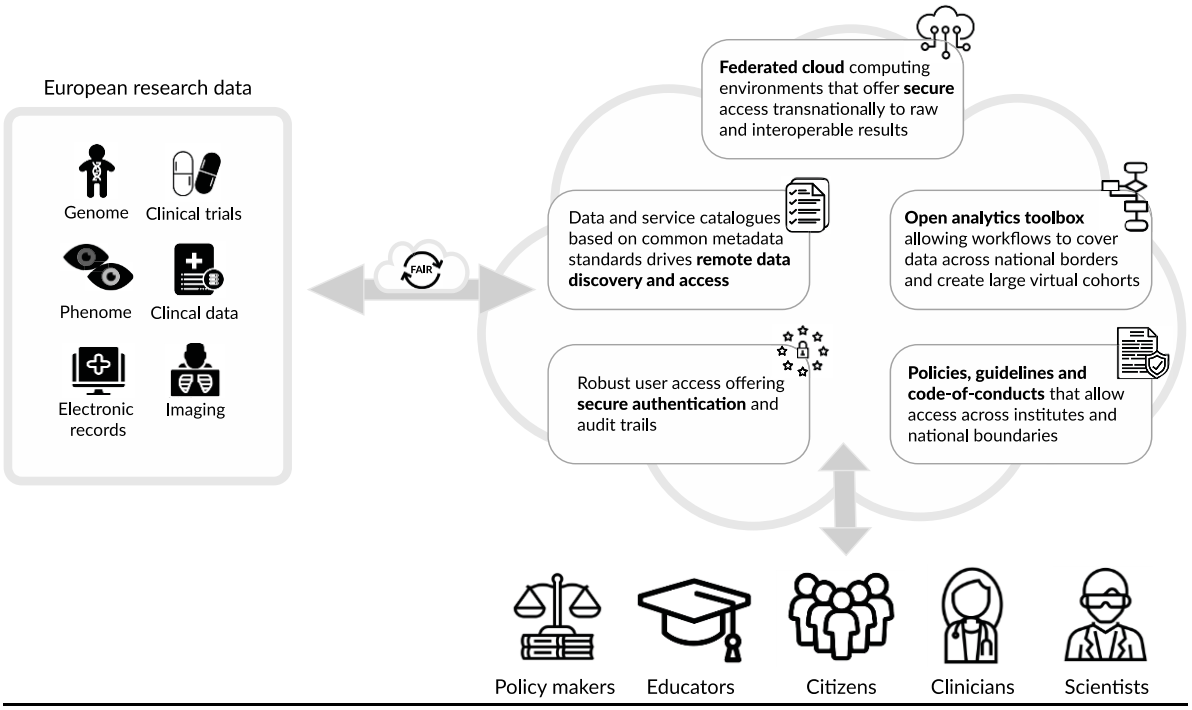

Figure 1 : Proposed general architecture of the HRIC European (inter)national databases, with varying data formats and data types are referenced in a metadata repository, following formatting rules of the Federated data commons as agreed at the HRIC governance level. The different users, after access control to the cloud, use the HRIC interface to access the repository, which gathers the relevant data and performs analysis, with outputs such as mathematical models, data visualisations, statistics and patient's profiles, according to the users' needs.

Table 3: Relevant initiatives for the European Health Research and Innovation cloud

PROJECT	AIMS/SUMMARY	Cloud model used	References
CORBEL project	Creating a platform for harmonised user access to biological and medical technologies, biological samples and data. The project has developed the data harmonisation, ethics guidance and user access protocols necessary for transnational access to both pre-clinical and clinical research infrastructures and is piloting access to participant level data from clinical trials[43].	Scalable cloud-based provision of data access and compute across infrastructures.	[59]
ELIXIR	EUROPEAN RESEARCH INFRASTRUCTURE with 21 members and over 180 research organisations. ELIXIR is creating a	Hybrid cloud ecosystem : i/ Local, private clouds (e.g. EMBL-EBI Embassy);	[60]

	network of local instances of the European Genome-Phenome Archive that give users controlled and secure access to raw data and precomputed results[44,45].	ii/ National community clouds (e.g. cPouta, MetaCentrum cloud, de.NBI); iii/ European research and innovation oriented clouds (e.g. EOSC) iv/ Public/commercial compliant clouds (e.g. Google, Azure, AWS).	
European Translational Information and Knowledge Management Services (eTRIKS)	IMI-funded highly scalable cloud-based platform for translational research, information and knowledge management providing open-source applications that can securely host heterogeneous data types, including multi-omics data, preclinical laboratory data and clinical information, including longitudinal data sets. The platform is a robust translational research knowledge management system that is able to host other data mining applications and support the development of new analytical tools[46].	Scalable cloud-based platform for translational research and applications development. The Openstack technology is used to run a private cloud for eTRIKS.	[61]
European Medical Information Framework (EMIF)	IMI-funded project that has successfully improved access to human health data via providing tools and workflows to discover, assess, access and (re)use human health data. The efforts of this IMI project are being extended through the EHDEN one.	Research analytical service approaches via HER and cohort data platforms.	[62, 63]
Human Brain Project (HBP) Medical Informatics Platform	The HBP Medical Informatics Platform allows researchers around the world to exploit medical data to create machine-learning tools that can analyse these data for new insights into brain-related diseases. The Medical Informatics web-portal[47] provides a software framework based on federated and distributed computing that allows researchers to mine clinical data stored on hospital and laboratory servers, without moving the data from the servers where they reside and without compromising patient privacy.	The HBP Joint Platform plans to adopt in cloud technology and provide those services through its computer centers JSC-Jülich and CSCS-Lugano together with BSC-Barcelona, CINECA-Bologna and CEA-Saclay. Software infrastructure : i/ The (base) infrastructure layer is accessible through an "Infrastructure as a Service" (IaaS) interface; ii/ Tools to enable simulation and modeling as well as data analytics workflows for neuroscience that the HBP operates as a "Platform as a Service" (PaaS);	[64]

		iii/ Several software services for data-driven brain simulations and for virtual neurorobot design and operation offered in form of "Software as a Service" (SaaS). HBP operates the following SaaS: a/ Model driven brain simulations; b/ Neurorobotics simulation and development tools across the whole workflow the of neurorobotics life cycle.	
Human Brain Project (HBP) Knowledge Graph Data Platform	The HPB Knowledge Graph (KG) is an online graph database accepting submissions of anonymized human data, animal data, and models from the brain. All data made discoverable and accessible through KG search have been curated. Data are also made available with integrated multilevel HBP Atlases, holding information about the brain in standard reference spaces.		[65]
Helix Nebula	The project is a pan-European public private partnership initiative led by EIROforum and leading commercial cloud computing partners. Since 2011, it has been piloting the use of cloud computing to enable complex data analyses and large-scale data sharing, with life science-oriented projects ranging from complex genome assembly to assessing somatic variation in the context of different types of cancer.	Helix Nebula Science Cloud (HNSciCloud): Hybrid cloud platform linking together commercial cloud service providers and publicly funded research organisations' in-house IT resources via the GEANT network to provide innovative solutions supporting data intensive science. These services support the connection of the research infrastructures identified in the ESFRI Roadmap to the nascent European Open Science Cloud (EOSC) intended to create a single digital research space for Europe's 1.8 million researchers.	[66]
European Open Science Cloud (EOSC)	The pilot project has recently investigated the benefits of data and cloud computer sharing at a pan-European level with life science-oriented projects on pan-cancer analyses[31] and imaging. This has been pursued across	Cloud based services for open sciences - integration and consolidation of e-infrastructure platforms, federation of existing European research infrastructures and scientific clouds.	[25]

	academic cloud computing environments located in Western and Eastern Europe as well as Canada. Similar initiatives have been launched in the USA[8].		
PCAWG	PCAWG - PANCANCER ANALYSIS OF WHOLE GENOMES The study is an international collaboration to identify common patterns of mutation in more than 2,800 cancer whole genomes from the International Cancer Genome Consortium. This project is exploring the nature and consequences of somatic and germline variations in both coding and non-coding regions, with specific emphasis on cis-regulatory sites, non-coding RNAs, and large-scale structural alterations.	Hybrid cloud model The data coordinating center lists collaborative agreements with cloud providers AWS and an academic compute cloud resource maintained at the cancer collaboratory, by the Ontario Institute for Cancer Research and hosted at the Compute Canada facility.	[67, 68]
RD-Connect	RD-Connect is an integrated platform connecting databases, patient registry data, biobanks and clinical bioinformatics for rare disease research. The RD-Connect allow the integration of different data types (e.g. omics, clinical information, patient registries and biobanks). Those integrated data can be accessed and analysed by the scientific community to speed-up research, diagnosis and therapy development for patients with rare diseases to contribute for better health.	Online secured platform connecting different types of patients related rare disease data enabling genome-phenome analysis.	[69]
COMPARE	COMPARE, a network of collaborators of the Global Microbial Identifier initiative (GMI), aims to improve identification and mitigation of emerging infectious diseases and foodborne outbreaks.	One serve all analytical framework and data exchange platform with various data integration for real time analysis and interpretation of sequence-based pathogen.	[70] [71]

Abbreviations

CERN (the European Organisation for nuclear research)
EMR (Electronic Health Record)
EOSC (European Open Science Cloud)
EU (European Union)
FAIR (Findable, Accessible, Interoperable, Retrievable)

FP7 (The European Union's seventh framework programme for Research, Technological Development and Demonstration)
H2020 (The European Union's Horizon Research and Innovation programme)
HBP (Human Brain project)
HRIC (European Health Research and Innovation Cloud)
OECD (Organisation of Economic Collaboration and Development)
IaaS (Infrastructure as a Service)
PaaS (Platform as a Service)
SaaS (Software as a Service)

Declarations

Availability of data and materials

"Not applicable"

Funding

The workshop participants received funding from the European Union Seventh Programme for Research, Technological Development and Demonstration (FP7) and Horizon Research and Innovation Programme (H2020) and other European Union programmes under the following grant agreements: AETIONOMY (Developing an Aetiology-based Taxonomy of Human Disease-Approaches to Develop a New Taxonomy for Neurological Disorders, IMI-nº115568), ANTI-SUPERBUG PCP (ANTISUPERBUG Precommercial Procurement, H2020-nº688878), B-CAST (Breast CAncer Stratification understanding the determinants of risk and prognosis of molecular sub-types, H2020-nº633784), BRIDGE Health (Bridging information and data generation for evidence-based health policy and research, H2020-nº664691), CASyM (Coordinating Action Systems Medicine – Implementation of Systems Medicine across Europe, FP7-n°305033), CENTER-TBI (Collaborative European NeuroTrauma Effectiveness Research in TBI, FP7-nº602150), CECM (Centre for New Methods in Computational Diagnostics and Personalized Therapy, H2020-nº763734), COLOSSUS (Advancing a Precision Medicine Paradigm in Metastatic Colorectal Cancer: Systems based patient stratification solutions, H2020-nº754923), COMPARE (COllaborative Management Platform for detection and Analyses of (Re-)emerging and foodborne outbreaks in Europe, H2020-nº643476), CONNECARE (Personalised Connected Care for Complex Chronic Patients, H2020-nº689802), CREATIVE (Collaborative REsearch on ACute Traumatic brain Injury in intensiVe care medicine in Europe, FP7-nº602714), CHAFEA grant-nº709723), DEFORM (Define the global and financial impact of research misconduct H2020-nº710246), ECCTR (European Cornea and Cell Transplant Registry, ECRIN-IA (European Clinical Research Infrastructures Network-Integrating Activity, FP7-nº284395), EHR4CR (Electronic Health Records Systems for Clinical Research, IMI-nº115189) eInfraCentral (European E-infrastructures Services Gateway, H2020-nº731049), ELIXIR (European Life-science Infrastructure for Biological Information, FP7-n°211601), ELIXIR-EXCELERATE (Fast track ELIXIR implementation and drive early user exploitation across the life sciences, H2020-nº676559), eMEN (e-mental health innovation and transnational implementation platform North West Europe, H2020), EMIF (European Medical Information Framework, IMI-nº115372), ERA PerMed (ERA-net Cofund in Personalised Medicine, H2020-nº779282), eTRIKS (Delivering European Translational Information and Knowledge Management Services, IMI-1-nº115446), E-COMPARED (European COMPARative

Effectiveness Research on online Depression, FP7- nº603098), EuroPOND (Data-driven models for progression of neurological diseases, H2020-n°666992), EurValve (Personalised Decision Support for Heart Valve disease, H2020-nº689617), HBP SGA1/SGA2 (Human Brain Project specific grant agreements, H2020-n°720270/785907), MASTERMIND (Management of Mental Disorders through Advanced Technologies, CIP-nº621000), MedBioinformatics (Creating medically-driven integrative bioinformatics applications focused on oncology, CNS disorders and their comorbidities, H2020-n°634143), ICT4DEPRESSION (User-friendly ICT tools to enhance self-management and effective treatment of depression in the EU, FP7- n°248778), ImpleMentAll (Towards evidence-based tailored implementation strategies for eHealth, H2020-nº733025), INSTRUCT-ULTRA (Releasing the full potential of instruct to expand and consolidate infrastructure services for integrated structural life sciences research, H2020-nº731005), MeDALL (Mechanisms of the Development of ALLergy, FP7-n°261357) MIDAS (Meaningful Integration of Data, Analytics and Services, H2020-nº 727721), MultipleMS (Multiple manifestations of genetic and non-genetic factors in Multiple Sclerosis disentangled with a multi-omics approach to accelerate personalised medicine, H2020-nº733161), myPEBS (Randomized Comparison Of Risk-Stratified versus Standard Breast Cancer Screening European Women Aged 40-74, H2020-nº755394), OpenAIRE-Advance (Advancing Open Scholarship, H2020-nº777541), OpenMedicine (OpenMedicine, H2020-no643796), PanCareSurFup (PanCare Childhood and Adolescent Cancer Survival Care and Follow-up Studies, FP7-no257505), PIONEER (Prostate Cancer DIagnOsis and TreatmeNt Enhancement through the Power of Big Data in EuRope, H2020-IMI-2-n°777492), PREPARE (Platform for European Preparedness Against (Re-)emerging Epidemics, FP7-n°602525), Regions4PerMed (Interregional coordination for a deep and fast uptake of personalised health, H2020-nº825812), RD-CONNECT (An integrated platform connecting registries, biobanks and clinical bioinformatics for rare disease research, FP7-nº305344), Solve-RD (Solving the unsolved Rare Diseases, H2020-nº779257), SPIDIA4P (SPIDIA for Personalised Medicine-Standardization of generic Pre-analytical procedures for In-vitro DIAgnostics for Personalised Medicine, H2020-nº733112), SYSCID (A Systems medicine approach to chronic inflammatory disease, H2020-nº733100), SysCLAD (Systems prediction of Chronic Allograft Dysfunction, FP7-n°305457), SYSCOL (Systems Biology of Colorectal Cancer, FP7-nº258236), SysMedPD (Systems Medicine of Mitochondrial Parkinson's Disease, H2020-n°668738), U-BIOPRED (Unbiased BIOmarkers for the PREDiction of respiratory disease outcomes, IMI-n°115010), VPH-share (Virtual Physiological Human: Sharing for Healthcare - A Research Environment, FP7-n°269978).
Neither the European Commission nor any other person acting on behalf of the Commission is responsible for the use, which might be made of the following information. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Competing interests

RB is a founder and holder of shares of MEGENO S.A. and holds shares of ITTM S.A. The other authors have declared no competing interests.

Authors' contributions

This publication is inspired by a workshop organized by the Health Directorate at the European Commission. The chairs ADM, CGC, EMF, FS, IP, LC, MB, MBu, ML, MVDB, NBl, RB contributed to the concept of the manuscript. GHS produced the draft workshop report, which was used as a basis for the manuscript. The core editing group CA, NB, BDM, JK, NB, MVDB reviewed

and finalized the manuscript. All authors contributed to the content of the manuscript and approved its final version.

Acknowledgements

We would like to thank the scientific policy officers at the European Commission Health Directorate, who developed the workshop concept, organised it and provided input to this publication: Sasa Jenko, Katarina Krepelkova, Christina Kyriakopoulou, Jana Makedonska, Joana Namorado, Elsa Papadopoulou, Jan Van de Loo and Gregor Schaffrath (National Expert in professional training). The workshop was organised by the Health Directorate of the Directorate-General for Research and Innovation at the European Commission with contributions of the Innovative Medicines Initiative, the Directorate General for Communications Networks, Content and Technology and the Directorate General for Health and Food Safety.

References

1. Birney E, Vamathevan J, Goodhand P: **Genomics in healthcare: GA4GH looks to 2022**. In: *Cold Spring Harbor Laboratory*. bioRxiv; 2017.
2. Langmead B, Nellore A: **Cloud computing for genomic data analysis and collaboration**. *Nature reviews Genetics* 2018, **19**(5):325.
3. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, Bernal-Delgado E, Blomberg N, Bock C, Conesa A *et al*: **Making sense of big data in health research: Towards an EU action plan**. *Genome medicine* 2016, **8**(1):71.
4. Mazein A, Ostaszewski M, Kuperstein I, Watterson S, Le Novere N, Lefaudeux D, De Meulder B, Pellet J, Balaur I, Saqi M *et al*: **Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms**. *NPJ systems biology and applications* 2018, **4**:21.
5. Ostaszewski M, Gebel S, Kuperstein I, Mazein A, Zinovyev A, Dogrusoz U, Hasenauer J, Fleming RMT, Le Novere N, Gawron P *et al*: **Community-driven roadmap for integrated disease maps**. *Briefings in bioinformatics* 2019, **20**(2):659-670.
6. Phillips M, Molnar-Gabor F, Korbel JO, Thorogood A, Joly Y, Chalmers D, Townend D, Knoppers BM, Consortium P: **Of Clouds and Genomic Data Protection**. *Nature* 2019.
7. PCAWG C: **Pan-cancer analysis of whole genome**. *Nature* 2019.
8. Lochmuller H, Badowska DM, Thompson R, Knoers NV, Aartsma-Rus A, Gut I, Wood L, Harmuth T, Durudas A, Graessner H *et al*: **RD-Connect, NeurOmics and EURenOmics: collaborative European initiative for rare diseases**. *European journal of human genetics : EJHG* 2018, **26**(6):778-785.
9. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE *et al*: **The FAIR Guiding Principles for scientific data management and stewardship**. *Scientific data* 2016, **3**:160018.
10. Lamy P, Brudermüller M, Ferguson M, Friis L, Garmendia C, Gray I, Gulliksen J, Kulmala H, Maher N, Plentz Fagundes M *et al*: **LAB - FAB - APP Investing in the European future we want**. In. Brussels: Independent High Level Group on maximising the impact of EU research & Innovation Programmes; 2017.
11. Hauser H, Bergman N, Bruncko M, Cosgrave P, Dwyer G, Helder M, Hinrikus T, Hoerr I, Karia B, Kolar J *et al*: **Funding - Awareness - Scale - Talent (FAST) Europen is**

back: Accelerating breakthrough innovation. In. Luxemburg: European Commission; 2018.
12. **Implementation Roadmap for the European Open Science Cloud** [https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none]
13. Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD: **Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud**. *Information Services & Use* 2017, **37**:49-56.
14. **FAIRsharing homepage** [https://fairsharing.org/]
15. **REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL** [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679]
16. Bouzaglo D, Chasida I, Ezra Tsur E: **Distributed retrieval engine for the development of cloud-deployed biological databases**. *BioData mining* 2018, **11**:26.
17. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, Knoetze N, Breden FMW, Christley S, Scott JK *et al*: **iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories**. *Immunological reviews* 2018, **284**(1):24-41.
18. Sreenivasaiah PK, Kim DH: **Current trends and new challenges of databases and web applications for systems driven biological research**. *Frontiers in physiology* 2010, **1**:147.
19. Wade TD: **Traits and types of health data repositories**. *Health information science and systems* 2014, **2**:4.
20. Estiri H, Klann JG, Weiler SR, Alema-Mensah E, Joseph Applegate R, Lozinski G, Patibandla N, Wei K, Adams WG, Natter MD *et al*: **A federated EHR network data completeness tracking system**. *J Am Med Inform Assoc* 2019, **26**(7):637-645.
21. Shenai S, Aramudhan M: **Cloud computing framework to securely share health & medical records among federation of healthcare information systems**. *Biomedical research* 2018; Special Issue: S133-S136.
22. Oderkirk J: **Readiness of electronic health record systems to contribute to national health information and research**. In. Paris: OECD Health Working Papers; 217: 80.
23. **eHealth: Digital health and care** [https://ec.europa.eu/health/ehealth/projects/nationallaws_electronichealthrecords_en]
24. **Overview of the national laws on electronic health records in the EU Member States and their interaction with the provision of cross-border eHealth services - Final report and recommendations** [https://ec.europa.eu/health/sites/health/files/ehealth/docs/laws_report_recommendations_en.pdf]
25. **EOSC homepage** [https://www.eoscpilot.eu/]
26. Molnar-Gabor F, Lueck R, Yakneen S, Korbel JO: **Computing patient data in the cloud: practical and legal considerations for genetics and genomics research in Europe and internationally**. *Genome medicine* 2017, **9**(1):58.
27. **EU countries will cooperate in linking genimc databases across borders** [https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders]
28. **Declaration of cooperation towards access to at least 1 million sequenced genomes in the European Union by 2022** [https://www.euapm.eu/pdf/EAPM_Declaration_Genome.pdf]

29. Saunders G, Baudis M, Becker R, Beltran S, Béroud C, Birney E, Brooksbank C, Brunak S, Van den Bulcke M, Drysdale R *et al*: **Leveraging European infrastructures to access 1 million human genomes by 2022**. *Nature Reviews Genetics* 2019.
30. Drosatos G, Kaldoudi E: **Blockchain Applications in the Biomedical Domain: A Scoping Review**. *Computational and structural biotechnology journal* 2019, **17**:229-240.
31. Esposito C, De Santis A, Tortora G, Chang H, Choo K-KR: **Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?** *IEEE Cloud Computing* 2018, **5**.
32. Miller JB: **Big data and biomedical informatics: Preparing for the modernization of clinical neuropsychology**. *The Clinical neuropsychologist* 2019, **33**(2):287-304.
33. **The eTRIKS code of practice on secondary use of medical research data** [https://www.etriks.org/code-of-practice/]
34. De Hert P, Papakonstantinou V: **Three Scenarios for International Governance of Data Privacy: Towards an International Data Privacy Organization, Preferably a UN Agency?** *I/S: A Journal of Law and Policy for the Information Society* 2013, **9**(2):271-324.
35. **EU picks team of 11 to run giant European Open Science Cloud** [https://sciencebusiness.net/science-cloud/news/eu-picks-team-11-run-giant-european-open-science-cloud]
36. **Human brain project ethics and society sub-projects** [https://www.humanbrainproject.eu/en/social-ethical-reflective/]
37. Salles A, Bjaalie JG, Evers K, Farisco M, Fothergill BT, Guerrero M, Maslen H, Muller J, Prescott T, Stahl BC *et al*: **The Human Brain Project: Responsible Brain Research for the Benefit of Society**. *Neuron* 2019, **101**(3):380-384.
38. **8 Benefits and Risks of Cloud Computing in Healthcare** [https://solutionsreview.com/cloud-platforms/8-benefits-and-risks-of-cloud-computing-in-healthcare/]
39. Ali O, Shrestha A, Soar J, Wamba F: **Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review**. *International Journal of Information Management* 2018, **43**:146-158.
40. Lian JW: **Establishing a Cloud Computing Success Model for Hospitals in Taiwan**. *Inquiry : a journal of medical care organization, provision and financing* 2017, **54**:46958016685836.
41. de Bruin B, Floridi L: **The Ethics of Cloud Computing**. *Science and engineering ethics* 2017, **23**(1):21-39.
42. Council CSC: **Impact of Cloud Computing on Healthcare**. In.; 2017.
43. Sultan N: **Making use of cloud computing for healthcare provision: Opportunities and challenges**. *International Journal of Information Management* 2014, **34**(2):177-184.
44. Sobeslav V, Maresova P, Krejcar O, Franca TC, Kuca K: **Use of cloud computing in biomedicine**. *Journal of biomolecular structure & dynamics* 2016, **34**(12):2688-2697.
45. Rodrigues JJPC, Sendra Compte S, de la Torra Diez I: **Cloud Computing on e-Health**. In: *e-Health Systems, Theory, Advances and Technical Applications.* Edited by Elsevier IP-; 2016: 191-207.
46. **CERN Openlab homepage** [https://openlab.cern/]
47. **Worldwide LHC Computing Grid homepage** [http://wlcg.web.cern.ch/]
48. Chen J, Qian F, Yan W, Shen B: **Translational biomedical informatics in the cloud: present and future**. *BioMed research international* 2013, **2013**:658925.

49. Khan N, Yaqoob I, Hashem IA, Inayat Z, Ali WK, Alam M, Shiraz M, Gani A: **Big data: survey, technologies, opportunities, and challenges**. *TheScientificWorldJournal* 2014, **2014**:712826.
50. Kuo AM: **Opportunities and challenges of cloud computing to improve health care services**. *Journal of medical Internet research* 2011, **13**(3):e67.
51. Navale V, Bourne PE: **Cloud computing applications for biomedical science: A perspective**. *PLoS computational biology* 2018, **14**(6):e1006144.
52. LAB, FAB, APP: **Investing in the European future we want.** In. Luxemburg: Publication office of the European Union; 2017.
53. Mazzucato M: **Mission-Oriented Research & Innovation in the European Union - a problem-solving approach to fuel innovation-led growth**. In. Luxemburg: Publication office of the European Union; 2018.
54. **EIT in Horizon Europe (2021-2027) - complementarities and synergies with the EIC** [https://eit.europa.eu/sites/default/files/eit_position_paper_horizon_europe_eit_eic_0.pdf]
55. H2020 Scientiic Panel for Health: **Building the future of health research - Proposal for a European Council for Health Research** In.; 2018.
56. **European Joint Programme of Rare Diseases homepage** [https://www.ejprarediseases.org/index.php/about/]
57. **European Research Council homepage** [https://erc.europa.eu]
58. **European Structural and Investment Funds** [https://ec.europa.eu/info/funding-tenders/funding-opportunities/funding-programmes/overview-funding-programmes/european-structural-and-investment-funds_en]
59. **CORBEL project homepage** [http://www.corbel-project.eu]
60. **ELIXIR infrastructure homepage** [https://www.elixir-europe.org/]
61. **The eTRIKS project homepage** [https://www.etriks.org/]
62. **EMIF project homepage** [http://www.emif.eu/]
63. **EHDEN project homepage** [https://www.ehden.eu/]
64. **Human Brain Project homepage** [https://www.humanbrainproject.eu/en/medicine/]
65. **Human Brain Project Knowledge Graph** [https://www.humanbrainproject.eu/en/explore-the-brain/]
66. **HELIX nebula homepage** [http://www.helix-nebula.eu/]
67. [https://dcc.icgc.org/icgc-in-the-cloud/collaboratory]
68. **PCAWG project homepage** [https://dcc.icgc.org/pcawg]
69. **RD Connect homepage** [https://rd-connect.eu/]
70. **COMPARE project homepage** [http://www.compare-europe.eu]
71. [https://www.globalmicrobialidentifier.org]

Towards a European Health Research and Innovation Cloud (HRIC)

Albeyatti A.[1], Armitage W.J.[2], Augello L.[3], Auffray C.[4], Balling R.[5], Benhabiles N.[6*], Bertolini G.[7], Bjaalie J.G.[8], Black M.[9], Blomberg N.[10*], Bogaert P.[11], Bubak M.[12], Claerhout B.[13], Clarke L.[14], De Meulder B.[15], D'Errico G.[16], Di Meglio A.[17], Forgo N.[18], Gans-Combe C.[19], Gray A.E.[20], Gut I.[21], Gyllenberg A.[22], Hemmrich-Stanisak G.[23], Hjorth L.[24], Ioannidis Y.[25], Jarmalaite S.[26], Kel A.[27], Kherif F.[28]., Korbel J.O.[29*], Larue C.[30], Laszlo M.[31], Maas A.[32], Magalhaes L.[33], Manneh-Vangramberen I.[34], Aarestrup F.M.[35], Morley-Fletcher E.[36], Ohmann C.[37], Oksvold P.[38], Oxtoby N.P.[39], Perseil I.[40], Pezoulas V.[41], Riess O.[42], Riper H.[43], Roca J.[44], Rosenstiel P.[45], Sabatier P.[46], Sanz F.[47], Tayeb M.[48], Thomassen G.[49], Van Bussel J.[50], Van den Bulcke M.[51], Van Oyen H.[52]*

Affiliations:

[1]Medicalchain, National Health Service, London United Kingdom, 'abdullah@medicalchain.com'
[2]Bristol Medical School, Translation Health Sciences, Bristol, United Kingdom 'W.J.Armitage@bristol.ac.uk'
[3]Regional Agency for Innovation & Procurement (ARIA), Welfare Services Division, Milano, Lombardy, Italy luca.augello@ariaspa.it
[4]European Institute for Systems Biology and Medicine (EISBM), Vourles, France 'cauffray@eisbm.org'
[5]Luxembourg Centre for Systems Biomedicine, Campus Belval, University of Luxembourg 'rudi.balling@uni.lu'
[6]CEA, French Atomic Energy and alternative energy Commission, Direction de la Recherche Fondamentale, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France 'nora.benhabiles@wanadoo.fr'
[7] Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy 'guido.bertolini@marionegri.it'
[8]Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway 'j.g.bjaalie@medisin.uio.no'
[9]Ulster University, Belfast, United Kingdom 'mm.black@ulster.ac.uk'
[10]ELIXIR, Welcome Genome Campus, Hinxton, CB10 1SD, United Kingdom 'niklas.blomberg@elixir-europe.org'
[11]Sciensano, Brussels, Belgium and Tilburg University, Tilburg, The Netherlands 'Petronille.bogaert@sciensano.be'
[12]Akademia Gornizco Hutnizca University of Science and Technology, Department of Computer Science and Academic Computing Center Cyfronet, Krakow, Poland 'bubak@agh.edu.pl'
[13]Ghent University, Ghent, Belgium 'brecht.claerhout@trinetx.com"
[14]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK 'laura@ebi.ac.uk'
[15]European Institute for Systems Biology and Medicine (EISBM), Vourles, France 'bdemeulder@eisbm.org'
[16]Fondazione Toscana Life Sciences 'gianni.derrico@toscanalifesciences.org'
[17]CERN, European Organization for Nuclear Research, Meyrin, Switzerland 'alberto.di.meglio@cern.ch'
[18]University of Vienna, Vienna, Austria 'nikolaus.forgo@univie.ac.at'
[19]INSEEC School of Business & Economics, Paris, France 'cganscombe@inseec.com'
[20]PwC, Oslo, Norway 'Alexander.Edward.Gray2@ehelse.no'
[21]Center for Genomic Regulations, Barcelona, Spain 'Ivo.gut@cnag.crg.eu'
[22]Neuroimmunology Unit, The Karolinska Neuroimmunology & Multiple Sclerosis Centre, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden 'Alexandra.Gyllenberg@ki.se'
[23]Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany 'g.hemmrich-stanisak@ikmb.uni-kiel.de
[24]Lund University, Skåne University Hospital, Department of Clinical Sciences, Pediatrics, Lund, Sweden 'Lars.Hjorth@skane.se'
[25]Athena Research & Innovation Center and University of Athens, Greece 'yannis@di.uoa.gr'
[26]National Cancer Institute, Vilnius, Lithuania 'sonata.jarmalaite@nvi.lt'
[27]geneXplain GmbH, Wolfenbüttel, Germany 'alexander.kel@genexplain.com'
[28]Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland 'Ferath.Kherif@chuv.ch'
[29]European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany 'korbel@embl.de'
[30]Integrated Biobank of Luxembourg, Luxembourg, 'Catherine.Larue@lih.lu'
[31]'email@mitzilaszlo.org'
[32]Antwerp University Hospital and University of Antwerp, Belgium 'Andrew.Maas@uza.be'
[33]Clinerion Ltd, Basel, Switzerland 'luis.magalhaes@clinerion.com'
[34]European Cancer Patient Coalition, Belgium 'isabelle.manneh-vgb@ecpc.org'

Formatted: Tab stops: 2.85", Left

Formatted: French (France)

[35]Technical University of Denmark, Kongens Lyngby, Denmark 'fmaa@food.dtu.dk'
[36]Lynkeus and Public Policy Consultant, Rome, Italy 'emf@lynkeus.com'
[37]European Clinical Research Infrastructure Network, Duesseldorf, Germany 'christian.ohmann@med.uni-duesseldorf.de'
[38]Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden 'per.oksvold@scilifelab.se'
[39]Centre for Medical Image Computing, Department of Computer Science, University College London, United Kingdom 'n.oxtoby@ucl.ac.uk'
[40]Institut National de la Santé et de la Recherche Médicale, Information Technology Department, Paris, France 'isabelle.perseil@inserm.fr'
[41]Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, Greece 'bpezoulas@gmail.com'
[42]Institute of Medical Genetics and Applied Genomics, Rare Disease Center, Tuebingen, Germany 'Olaf.Riess@med.uni-tuebingen.de'
[43]Vrije Universiteit, Dept. of Behavioural and Movement Sciences, Section Clinical, Neuro and Developmental Psychology, Amsterdam, The Netherlands 'h.riper@vu.nl'
[44]Hospital Clínic de Barcelona, IDIBAPS, University of Barcelona, Barcelona, Spain 'JROCA@clinic.cat'
[45]Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany 'p.rosenstiel@mucosa.de'
[46]French National Centre for Scientific Research, Grenoble, France 'Philippe.Sabatier@imag.fr'
[47]Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain 'ferran.sanz@upf.edu'
[48]Medicalchain, London, United Kingdom 'Mo@medicalchain.com'
[49]University of Oslo, Oslo, Norway 'gardot@usit.uio.no'
[50]Scientific Institute of Public Health, Brussels, Belgium 'Johan.VanBussel@wiv-isp.be'
[51]Scientific Institute of Public Health, Brussels, Belgium 'marc.VanDenBulcke@wiv-isp.be'
[52]Ghent University, Ghent, Belgium and Sciensano, Brussels, Belgium 'Herman.VanOyen@sciensano.be'
*** corresponding authors**

Abstract

The European Union (EU) initiative on the Digital Transformation of Health and Care (Digicare) aims at providing conditions for building a secure, flexible and decentralised digital health infrastructure. Creating a European Health Research and Innovation Cloud (HRIC) within this environment should enable data sharing and analysis for health research across the EU in compliance with data protection legislation while preserving the full trust of the participants. Such a HRIC should learn from and build on existing data infrastructures, integrate best practices, and focus on the concrete needs of the community in terms of technologies, governance, management, regulation and ethics requirements. Here we describe the vision and expected benefits of digital data sharing in health research activities and a roadmap fostering the opportunities while answering the challenges of implementing a HRIC. For this, we put forward five specific recommendations and action points to ensure that a European HRIC: 1) is built on established standards and guidelines, providing cloud technologies through an open and decentralised infrastructure; 2) is developed and certified to the highest standards of interoperability and data security that can be trusted by all stakeholders; 3) is supported by a robust ethical and legal framework compliant with the EU General Data Protection Regulation (GDPR); 4) establishes a proper environment for the training of new generations of data and medical scientists; 5) stimulates research and innovation in transnational collaborations through public and private initiatives and partnerships funded by the EU through Horizon 2020 and Horizon Europe.

Background

Genomics has brought life sciences into the realm of data sciences - large scale DNA and RNA sequencing is now routine in life-science and biomedical research with an estimate of up to 60 million human genomes being available in the coming years [1, 2]. Recent innovations in medical research and healthcare, such as high-throughput genome sequencing, transcriptomics, proteomics, metabolomics, single-cell omics techniques, high-resolution imaging, electronic medical records (EMRs), big-data analytics and a plethora of internet-connected health devices fundamentally change the infrastructure requirements for health research.

Translating these new data together with clinical information into scientific insights and actionable outcomes for improving clinical care is a major challenge. As life-science and health research datasets rapidly grow larger, with an ever-increasing number of study participants required to detect meaningful but weak ~~-~~signals ~~among~~ that may be blurred by a myriad of confounding biological, experimental or environmental factors, the computational resources required to process and analyse this big data increasingly outgrows the capabilities of even large research institutes. The various c~~C~~loud technologies and services listed in Table 1 are~~,~~ based on shared commercial and private computer and storage resources ~~–~~that can be provided on demand to users from a large number of different institutions conducting or participating in joint projects ~~(Table 1)~~. They~~,~~ have emerged as powerful solutions to the challenges of collaborating in research on genomic, biomedical and health data. Biomedical and health research has yet to fully enter the b~~B~~ig d~~D~~ata and c~~C~~loud computing era. The HRIC, as described in this manuscript, would help facilitate this transition, providing access to larger datasets, cutting-edge tools and knowledge, as envisioned in [3]. For example, the HRIC should ease the incorporation of domain expert knowledge into systems disease maps in a format that can be both understood by all stakeholders (patients and clinicians, scientists and drug developers), and processed by a high-performance computers, thus supporting the development of innovative medicines and diagnostics [4, 5]. Cloud technologies (through e.g. Hadoop applications) makes it possible to collaborate, access and reuse data also in situations when privacy concerns or regulation prohibits remote users from downloading data – an important benefit in Europe where national regulations can differ significantly. Clouds allow bringing algorithms to the data and as such can enable data sharing and joint processing without generating unnecessary copies of the data, which comes with potential benefits for data protection~~Cloud technologies (through e.g. Hadoop applications) makes it possible to collaborate, access and reuse data also in situations when privacy concerns or regulation prohibits remote users from downloading data – an important benefit in Europe where national regulations can differ significantly. Clouds allow bringing algorithms to the data and as such can enable data sharing and joint processing without generating unnecessary copies of the data, which comes with potential benefits for data protection~~ [6, 7]~~.~~ Clouds, additionally, enable performing computational analyses at a scale that individual institutions would struggle to manage ~~Clouds, additionally, enable performing computational analyses at a scale where individual institutions would struggle~~ [7]~~.~~ Consequently, in the last few years the large international cancer and other genomics consortia have created specialized genomics and biomedical cloud environments, each supporting individual projects [2]. These projects have made important advances in connecting health research data across disciplines, organisations and national boundaries. For instance, in research on rare diseases,

Field Code Changed

Field Code Changed

Commented [CA1]: RA1: 'The title of this paper invokes "Cloud" but other than using the term "federated cloud" in the article nothing much is discussed how cloud computing will fit into the HRIC?'

Please consider briefly clarifying this in your Background section.

Commented [CA2]: RA2: 'A prefix of "innovation" is used with Cloud, no clear discussion is provided on what kind of innovations are conceived by using Cloud Computing to tackle biomedical research challenges.'

Please consider briefly clarifying this – we assume that the proposed HRIC aims to facilitate research and innovation, but consider briefly clarifying how and what kind of innovation is envisioned.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

international collaborations that integrate genomic, phenotypic, and clinical data have introduced new paradigms in diagnosis and care [8]. However, a project-based, fragmented landscape will not enable access and construction of large data cohorts that are required to address novel or broader biomedical questions that were not anticipated in individual projects when collecting inform consent from the participants, nor will it provide adequate data governance and containable cost models.

Scaling and sustainably managing such solutions to support all European life scientists thus requires a coordinated action from science policy makers, funders and other actors in this complex ecosystem. Connecting Europe's health data to advance the understanding of life and disease requires that research data and analysis tools, standards and computational services are made FAIR – i.e. findable, accessible, interoperable, and reusable for researchers across scientific disciplines and national boundaries [9]. Truly enabling personalised and digital medicine across Europe and beyond will require a connected digital infrastructure for Europe's health data that supports systematic openness and integration of research data with real-world datasets (e.g. environmental monitoring data)~~.~~ generated inside all healthcare systems, government agencies, foundations and private organisations that will adopt it.~~(e.g. environmental monitoring data).~~

On the 13th of March 2018, the Health directorate of the Directorate-General for Research and Innovation of the European Commission of the EU organised a workshop to explore the possibility and challenges involved in establishing a cloud for health research and innovation, accessible by researchers and health professionals throughout Europe, in line with recommendations for a European Innovation Council and the Horizon Europe 2021-2027 framework programme [10, 11]. The cloud computing environment proposed in this manuscript builds upon the European Open Science Cloud (EOSC) initiative developed in the last few years by the European Commission [12], with a focus in the life science and medicine field. The EOSC aims at developing a trusted, open environment for the scientific community for storing, sharing and re-using scientific data and results. Overall, the authors feel that the cloud described in this manuscript, providing the biomedical and health research community with the technical infrastructure and services listed in Table 1 to support the development of innovative diagnostics methods and medical treatments, should become an integral part of the EOSC. The workshop gathered a broad range of experts from multiple biomedical research disciplines, health care, informatics, ethics and legislation, including representatives of more than 45 FP7 and Horizon 2020 (H2020) EU-funded collaborative projects. The participants explored requirements and developed a set of recommendations for a European "*Health Research and Innovation Cloud*" (HRIC) to connect researchers and health data sources in Europe [13] such that clinical data, software, computational resources, methods, clinical protocols, and publications, can be more~~are~~ widely and securely accessed and reused following the FAIR principles [9] than is currently possible with existing European research infrastrucures such as ELIXIR that form a network of heterogenous national nodes. ~~–~~For example, t~~T~~he HRIC infrastructure would benefit from the aforementioned advantages of cloud computing for the archiving and dissemination of health data ~~via compute clouds~~. This paper summarises the main conclusions from the workshop and highlights five recommendations and action points to the EU and national stakeholders (Table 2). The recommendations are key issues that need to be addressed in order to link biological, clinical, environmental and lifestyle information (from single individuals to large cohorts) to the health and wellbeing status of patients and citizens over time, while making this wealth of data and information available for European health research and innovation in clinical care.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Commented [BDM3]: RA4: 'Which basic services will be provided by the HRIC? What types of applications will be used for tackling data integration?'
'What types of innovation can one anticipate by development of HRIC?'

Please briefly answer those points.

Commented [BDM4R3]: Niklas, Alberto

Field Code Changed

Field Code Changed

The HRIC should be built on established standards and guidelines, to foster European-wide medical research

Rationale

Sharing of data, information and knowledge represents the most important functionality in the context of a HRIC. High-level standardisation, common exchange mechanisms, interfaces and protocols, and semantic interoperability form the foundation for widespread adoption of the FAIR principles [9] in health research. Data shared collaboratively in such health-related cloud projects are now largely standardised for the processing of genomic DNA read and genomic variant calling files. By comparison, the sharing of highly sensitive clinical and health data has thus far been much less developed hence representing a key future focus area, and numerous challenges remain with regard to sharing these data in a meaningful way.

Existing standards and guidelines

Many individual projects, in Europe and globally, have demonstrated the opportunities and the added value presented by connecting and exchanging data across countries via standardised protocols. Table 3 lists recent European projects that have developed towards clinical and health data exchange via cloud-based solutions. All those projects have developed and implemented worthy ideas that should be included in the HRIC. However, we envision the HRIC as a disease-agnostic environment, and on a larger scale than the platforms mentioned in Table 3. Moreover, the HRIC should not be tied to a single project or consortium, but should rather be under the governance of an independent body. International exchange of health research data holds tremendous potential in disease research by facilitating better investigation of disease causality and linking of genotypes and phenotypes, such as demonstrated in the Pan-Cancer Analysis of Whole Genomes (PCAWG) project, for example. An important aspect of the cloud-based data governance is that it allows sharing of data outside a consortium through a data request mechanism and a governance infrastructure that track participant consent, and data access. The cloud-based audit trail capabilities of the cloud-based research analyses that can be implemented at both data and infrastructure levels is a direct benefit to the data controllers. A standardised data model (or data access model) and/or standardised metadata models facilitate consolidation of different data sets and significantly increase the findability, the semantic interoperability and, by consequence, the reusability of data, and thus its 'FAIRness' [9, 14].

Workshop participants agreed that a first minimalistic and yet effective approach to data exchange should consist of a small number of initial online repositories containing references (e.g. links to render data sets findable) along with metadata (e.g. type and scale of content, specifications describing in what systems the data set may be stored and processed), and indications on how to gain access (e.g. requirements and point of contact). This could be designed as a metadata repository containing the metadata of data objects and information on how to gain access to them. The data objects as such may, or indeed should, be stored elsewhere.

Beyond their metadata, however, data sets are bound to greatly differ, as research projects vary largely in scope. Not only would it be cumbersome to record a large number of parameters irrelevant to the specific question addressed, but it would also be problematic from an ethical viewpoint, considering patient personal data protection aspects [15]. It

Field Code Changed

Commented [CA5]:
RA5: 'On page 4, line 46-47, reference to Table 3; what is it that these platforms do not provide? Why is there a need for HRIC? What gaps will HRIC address?'

Please consider briefly clarifying this point.

Field Code Changed

Field Code Changed

Field Code Changed

therefore seems more promising to drive standardisation within research communities while looking out for opportunities for overall standardisation.

Thus, the workshop envisioned the HRIC as a distributed collection of data repositories, people and services, which together make up a framework to share and operate as a federated data commons with reproducible software ~~[2, 16-19]~~, standards and expertise based on joint policies and guidelines to conduct health research as performed successfully in previous initiatives [2, 16-19]. ~~.~~ The need for federation is also highlighted in the proposed EU action plan for 'Making sense of big data in health research' [3]. Creating such a HRIC opens new frontiers for research and healthcare via the opportunities for strong international collaborations.

Beyond Europe the HRIC should collaborate internationally to drive the development and widespread adoption of global standards and connectivity. Ongo~~Exist~~ing initiatives exist ~~in the USA, Canada, Africa and the United Kingdom~~outside Europe ~~[20, 21] and~~that are aiming at developing global standards for the secure ~~data~~ exchange ~~and security~~of e.g. health and medical records within health information federations while tracking completeness of the supporting data [20, 21], as well as developing guidelines for data analytics and standardised workflows. T~~, t~~h~~a~~e~~y~~t should be considered for the basis of the HRIC as this will~~. This~~ provide~~s~~ the scientific community with means of reproducibility, version control and documentation, which will be an important vehicle to drive increased standardisation and connectivity.

A European HRIC should be developed and certified to the highest standards of interoperability and data security

Rationale

The workshop participants enthusiastically endorsed the vision of the HRIC as a federated environment. Through work on the standardisation, harmonisation and integration of genomic data with other health-relevant information to optimise hypothesis-driven analyses, the major blockers for cross-border collaborations in translational research on disease prevention, treatment and management will be addressed through a federated HRIC infrastructure in which data sources remain at their location of origin and are made accessible to users through a metadata repository. Moreover, data security has to be integral to the development of the HRIC, using modern cryptology and access control techniques to ensure the protection of the patients' data contained therein.

Standards in interoperability and data security

European health systems have different ways of managing and storing health data, making the exchange of clinical data between the EU member states complex. The challenges are well illustrated by the use of eHealth records and their use as secondary research material. In a recent OECD report [22] ten countries reported comprehensive record sharing within one country-wide system designed to support each patient having only one EHR (Supplementary Table.). These countries are Estonia, Finland, France, Greece, Ireland, Latvia, Luxembourg, Poland, Slovakia and the United Kingdom (England, Northern Ireland, Scotland and Wales). In these countries, plans call for patient records to be shared among physician offices and between physicians and hospitals regarding patient treatment, current medications, and laboratory tests and medical images. Some have already implemented part or all these functionalities, while others are progressing toward it. In other countries key aspects of record

Commented [CA6]: RA6: 'On page 5, lines 18-22 indicate that HRIC is envisioned as a "federated commons". But there is no explanation what biomedical research challenges, a federated data-commons is going to address that are not currently being met by existing infrastructures? There is no reference citation for Data Commons, and advantages of a federated Data Commons over non-federated Data Commons.'

Please consider briefly clarifying/addressing these points.

Answer:
We added four references related to data commons.

Field Code Changed

Field Code Changed

Field Code Changed

Commented [BDM7]: RA7: 'What are those initiatives? There is no context and no references provided for those initiatives'.

Please briefly clarify what those initiatives are.

Commented [BDM8R7]: We added two references related to the initiatives mentioned.

Field Code Changed

Field Code Changed

Field Code Changed

sharing are managed at sub-national level only, such as within provinces, states, regions or networks of health care organisations (for example, Austria, Germany, Italy, Netherlands, Sweden, and Spain; Supplementary Table). Among them, all have implemented or are planning the implementation of a national information exchange that enables key elements to be shared country-wide. Based on the recent reports of the European Commission [23], Belgium, Malta, Portugal, Romania and Slovenia are now developing national EHR systems, leading to a total of 16 EU member states providing such services.

Within the framework of the Joint Action on Rare Cancers, an EU initiative that brings together European research centres, policy makers and other stakeholders with the aim to set the agenda at national level, an analysis was made on the status of eHealth medical records in the EU member states. This work builds on the OECD study and complements this with information provided by the European Commission on the national laws on EHRs in the EU member states [24]. Thus all EU countries are investing in the development of clinical EHRs, but only some countries are moving forward the possibility of data extraction for research, the provision of statistics and the enablement of other uses that serve the public interest (P. Bogaert, personal communication). Countries that develop EHR systems that combine or virtually link data together to capture patient health care histories have the potential to be used also for long-term follow-up of cancer patients. Figure 1 shows how data from various sources can be integrated to provide a full picture of patients' health status over time and carry out research for patterns and anomalies in large populations using the specific combinations of data and analysis resources relevant for each research project, while ensuring compliance with security and data protection regulations.

The H2020-funded project European Open Science Cloud (EOSC)-Life [25] is developing policies, specifications and tools for the management of data for biological and medical research including aspects of eHealth data. The use of common metadata standards, developed in EOSC-Life, as a foundation for remote data discovery and access was emphasised by the workshop participants as a key enabler for the HRIC. For instance, practical and legal considerations for cloud computing of patient data, which include the responsible use of federated and hybrid clouds set up between academic and industrial partners have been put forward by early EOSC pilot projects [26].

The workshop participants emphasised that sustainability aspects are critical and must be considered from the start. To ensure that a HRIC could respond properly to emerging needs, innovation and technology changes, a distributed federated storage solution offering access to FAIR data and services should be built according to modularity principles. In particular, due to the long-term nature of a HRIC, due consideration should be given to deploying generic and modular computational methods and/or data storage management systems, while the information and communications technology (ICT) infrastructure should be flexible, portable and expandable.

The million European genomes initiative [27] is a case in point: eighteen member states have already signed the Genomics Declaration of Cooperation [28] to enable cross-border access to genomic databases and other health information. This federation of national initiatives [28] will provide secure access to such data resources in the member states to enable the discovery of personalised therapies and diagnostics for the benefit of patients. The initiative involves aligning strategies of ongoing national genomic sequencing campaigns with complementary *de novo* genome sequencing to obtain a total cohort of 1 million Europeans, accessible in a transnational framework, by 2022 [29]. The HRIC would form a basis for such large-scale, permanent collaborations.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

The workshop participants recognise that ensuring a maximal data security is paramount to build and maintain trust with the European citizens. To address this issue, we suggest using modern cryptology such as blockchain to ensure data security by design, and holding regular data security assessment (e.g. through hackathons and/or commercial security audits). As shown in recent literature, the use of blockchain in biomedical research is still in its infancy [30]. However, blockchain or other advanced cryptology tools that can be used to protect data in a cloud environment could prove useful for the secure and trustworthy implementation of the HRIC [31].

Commented [CA9]: RA8: 'On page 7, lines 9-10, There is no citation provided to support Block Chain can be successfully deployed for large scale Cloud based biomedical data management and access?'

Field Code Changed

Field Code Changed

The HRIC must be supported by a robust ethics and legal framework compliant with the GDPR

Rationale

Compliance with GDPR and other data protection laws, as well as enforcing an ethical usage of data is paramount to gain the general public's support and trust in the HRIC.

Existing ethics and legal framework

Unblocking the legal and administrative barriers for sharing human research data across geographical and organisational boundaries will, if the trust of research participants is preserved, pave the way for continent-scale cohorts in life science research. This will represent a significant innovation as sharing and joint analysis of sensitive data has until now been severely limited because of the different restrictions inherent to the different classes of sensitive data. By using a federated database model, with a metadata repository within the HRIC cloud encrypted environment, data security is maintained while innovative data analyses can be performed by bringing the algorithms to the data rather than centralizing the data [32]. Federation, rather than full integration of all available resources, poses an important challenge for implementation and deployment of an effective HRIC. The set-up and functioning of a HRIC requires a robust foundation of legal agreements and ethics rules and procedures, as well as security and data protection compliance protocols. Importantly these elements must be introduced during the conception and design phase as part of its governance. This is essential in order to allow the different ~~actors~~HRIC actors providing access to their data sources, managing them within the cloud, and accessing these resources –to incorporate policy requirements into the design and manage the complexity involved in implementing simple and intuitive user interfaces and project portals. This may be challenging considering the heterogeneity of health systems and health market accesses across Europe and will need the sharing of a common agreed vision across EU member states.

Field Code Changed

Ethical, societal and privacy considerations for (re)using health related data have been outlined in the Code of Practice on Secondary Use of Medical Research Data developed in the European Translational Information and Knowledge Management Services (eTRIKS) project funded by the Innovative Medicines Initiative (IMI) [33]. In addition to clear and explicit consent, explicit dissent may need to be considered for the use/re-use of data. Ultimately, each citizen and patient must be able to access her/his own data and know when and where it has been used and for what purpose. In addition, the difficult question of the business model of using those data should be discussed at various levels from ethical, societal and economical standpoints taking into account potential future developments of products and services using personal medical data. Furthermore, the goal of providing the citizens with personalised

Field Code Changed

services requires technical advancements in the collection and analysis of data (e.g. in data analytics and machine learning). For this type of usage, simple consent mechanisms might not be sufficient. For example, how should a clear data-collection purpose statement be defined, if data is collected for multiple usage scenarios across a distributed/federated cloud in which actors from different geographical and legislative environments will need to interact and cooperate? Would an excessive number of consent requests minimise data provision for research or clinical applications? Another level of complexity is introduced by the heterogeneity of data protection and privacy regulations when the data originate from states within federated national health systems (e.g. Germany, Italy). Development of large-scale European access mechanisms will require open consultation and engagement with national policymakers, patient organisations and the wider society to build the trust and confidence needed for widespread adoption and sustainable operations.

In addition to technical, ethical and legal specifications, a global integrated governance model needs to be established for the HRIC in line with that of the EOSC, regulating roles and responsibilities of all contributing institutions and users, and procedures for authentication and access control to individual resources. Principles, with specific guidelines on implementation into a HRIC environment, will need to be developed in order to manage and regulate aspects such as ownership, access, transparency, sharing, integration, standardisation of data and metadata formats, tools and frameworks, while ensuring confidentiality and sustainability. All these principles need to be developed with the overarching objective of providing benefit to and preserving the trust of patients and the general public.

Health data mostly represents sensitive data, which need to be managed to preserve the trust of patients, research participants and the general public, respect social norms and naturally comply with the rules and regulations of data protection laws, notably the EU GDPR [15]. Although the GDPR directly applies across the EU and its provisions prevail over national laws, EU member states retain the ability to introduce in their own national legislation certain derogations provided for by the GDPR itself. The GDPR also introduces the notions of 'Privacy by Design', which means that any organisation that processes personal data, must ensure that privacy is built into a system during the whole life cycle of the system or process; and 'Privacy by Default', which means that the strictest privacy settings should apply by default, without any manual input from the end user. In addition, any personal data provided by the user to enable the optimal use of a given health dataset should only be kept for the amount of time necessary to provide the intended product or service [15].

Thus, successfully linking and accessing biomedical and health data across Europe will require many different disciplines and specialists working together with a coordination effort that should encompass controlled access mechanisms to ensure compliance with privacy and data protection regulations. Data providers need logging and monitoring functionalities to comply with the GDPR and to enable tracking of data and methods within the system, controlling instances and routines that check for the adherence of predefined standards and formats to guarantee data integrity. Access mechanisms need to be developed that support the researchers, data producers and data analysts to request permissions and fulfil reporting requirements for data use in national and international research projects, which represents a significant regulatory, political and sustainability challenge [34]. These include in particular considerations about the rights of patient donors and research participants, taking into account the data protection aspects of various legal systems and local regulations. Researchers have to face differences in the understanding of the right to data protection in

Field Code Changed

Field Code Changed

Field Code Changed

those different regional or national European ecosystems. There is an urgent need for standardised, usable, data-protection-policy-compliant solutions for sensitive data sharing capable of integrating and analysing health data coming from different sources, organisations and potentially from different research disciplines. These aspects are subject to ongoing discussions and debates in the EOSC initiative [35] and for instance, progress has been made in the Human Brain Project through its Ethics and Society sub-project in collaboration with the project platforms [36, 37]. Other examples of data sharing compliant with data protection policies can be found in the recent literature [38-45]. Furthermore, there is the issue of capacity with the amount of data starting to strain the infrastructure of any individual hospital or research institute. Thus, the interplay between privacy, data security and access control on one hand and access (including cost-recovery models) to storage, computational and analysis resources on the other hand will be a defining element of the policy and technology development of a decentralised digital health infrastructure. The evolution of a cloud model that could be used in European health research will also have to take into account other specific aspects of the GDPR [15]. For instance, the European Commission intends to facilitate the free flow of non-personal data in the European Digital Single Market, and for health-related research participants it codifies the 'right to be forgotten'. It stipulates that patient donors should be able to retain control over their data regardless of technological developments. A European HRIC could be an important enabler for researchers to comply with these requirements. For example, once certain conditions are met between European and international partners, including those pertaining to data protection and use, federated and hybrid clouds could facilitate the deletion of data sets once a donor exercises her/his 'right to be forgotten', which thus could minimise the necessary transfer of large raw data sets across borders, as the deletion can be performed in the original dataset and easily propagated to the relevant federated data sources.

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

A proper training environment for HRIC developers and users should be established

Rationale

The workshop identified the lack of trained personnel, with solid skills in both medical and data analytical fields as one of the major bottlenecks when dealing with 'medical big data' [3].

Field Code Changed

Need for training and ideas

Indeed, effectively developing, operating and maintaining the HRIC will pose serious challenges and require the training of a new generation of data scientists able to navigate smoothly and efficiently between computational, security and medical disciplines. This includes clinical researchers, bioinformaticians, data analysts, data managers, software engineers, cloud engineers, other IT-specialists, ethics officers, and data protection specialists, which represent an essential new field of expertise. Finding professionals that cover more than one or two of the above-described disciplines is nearly impossible. Furthermore, communication between this comprehensive mix of clinical researchers, data managers and IT/bioinformatics specialists needs to be improved, requiring a governance structure well beyond that of a standard research setting. The EU should take inspiration from existing large and successful infrastructures that foster multidisciplinary teams, such as CERN [46, 47]. Thus, it is necessary to rethink the training and education of health professionals and to update

Field Code Changed
Field Code Changed

them with the HRIC in mind, considering both international standards and practices for data sharing as well as national environments and regulations.

Multiple funding mechanisms are required to drive the development of the HRIC and support its broad use in research projects

Rationale

Delivering the HRIC requires an ambitious reshaping of the European landscape for health data and research through appropriate funding schemes that enable transformation of fragmented ICT resources and project-centric solutions for data access and governance into a long-term, coherent service ecosystem that can be accessed by users transnationally.

Need for innovative public-private funding initiatives

Indeed, the HRIC needs a trusted and transparent innovation approach that recognises the importance of a clear, long-term ambition in the programme to support the participation of industry and small to medium enterprises (SME) in joint projects with a broad set of societal actions. In particular, there is a need to support the EU ICT industrial/SME innovation ecosystem in order to demonstrate the benefits in advancing data sharing, integration and analysis across Europe, for the benefit of all citizens, thus creating a foundation for attractive private investments.

For this purpose, targeted EU funding mechanisms also involving private investors need to support the development of HRIC-compliant services for data sharing and analysis in health-related research projects (i.e. through reimbursement for storage and computing costs) with incentives to reuse and extend existing infrastructures that favour national HRIC participation rather than rebuilding and fragmenting solutions. Moreover, the EU ICT industrial ecosystem must be supported in order to alleviate the risks associated with storing and sharing data across cloud systems operated by non-EU companies. The EU has established a strict privacy and ethics policy through the implementation of the GDPR legislation, binding for all operators active on the EU territory [48-51].

European funders, science policy makers and other actors need to develop mechanisms that bring together experience gained and lessons learned from a large portfolio of pathfinding projects and build on current investments, thus leveraging existing project outcomes. This requires an inclusive and integrative approach with programmes that bring together the many different actors into the HRIC, as its construction needs interdisciplinary collaboration with expertise from many disciplines e.g. economics, ICT, biomedical and health, social sciences and policy. In particular, frameworks for public-private partnerships such as IMI have shown a way to include industry in open transparent projects that also includes patients and other public bodies, SMEs and European researchers. Many of the mechanisms proposed in the Lamy report (prioritise research and innovation in EU and national budgets, build a true EU innovation policy that creates future markets, rationalise the EU funding landscape and achieve synergy with structural funds…)[52] and in the development of the Horizon Europe "Missions" [53] would also be well suited to the development of the HRIC and help bring together initiatives from the many funders, national and regional stakeholders. Other opportunities such as those proposed to be included in the Horizon Europe strategic workplan (e.g. the European Information Cloud, the European Institute of Technology and the European Council for Health Research [54, 55]) should be of interest to deploy innovations together with industry at the European level [56].

Commented [CA10]: RA9: 'On page 10, lines 28-30, it is not clear how risks in storing data in the Cloud will be lowered by EU ICT industrial ecosystem, when compared to their non-EU counterparts?'

Please consider briefly clarifying or addressing these minor points.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Furthermore, future programmes need to create incentives such that developed solutions are transformed into long term reusable resources and make sure that this infrastructure is deployed across the EU with development informed by ongoing research projects. The European Joint Programme on Rare Diseases (EJP-RD) [56] provides a good example of how infrastructure development can be linked with research projects at national and international levels. As in the EJP-RD, the European Strategy Forum on Research Infrastructures (ESFRI) can play a role in the development of HRIC. Two further aspects of EJP-RD are worth noting: the programme has a strong emphasis on the importance on the diverse workforce required with a training programme that stretches beyond academia and research networks to reach a broad set of individuals in health systems and the education sector. Secondly, the research portfolio should fund broadly and recognise that successfully addressing many of the identified challenges will require a diverse portfolio of projects that avoid any artificial boundary between biomedical and health research. Horizon Europe should allow for linkages between the HRIC, ESFRI and other themes of Horizon Europe and, importantly, between HRIC and other funding sources such as the European Structural and Innovation Funds (ESIF) and the cutting-edge basic research successfully supported by the European Research Council (ERC) during the past decade [57, 58].

Field Code Changed

Field Code Changed

Field Code Changed

The HRIC should enable investments in product development for future health-care solutions as well as allow health care providers to procure such solutions. People need to be an integral part of the innovation vision where the HRIC supports a highly skilled future workforce that makes Europe attractive for locating R&D investments. The experience gained in IMI public-private partnership infrastructure projects such as eTRIKS and the European Medical Information Framework (EMIF) should be leveraged. Finally, the current and future EU Framework Programmes for Research and Innovation (Horizon 2020 and Horizon Europe) should consider mobilising funds to support novel pilot actions, pooling of data and resources across the EU, and demonstrate the benefits in advancing data sharing, integration and analysis across Europe, for the benefit of all citizens.

Conclusions, recommendations and action points

Clouds are increasingly becoming a key venue for enabling and hosting European and international collaborations, benefitting from the ability to hold data securely in a single location (or in few locations) and enable collaborative research on computational infrastructure used for analysis. In conclusion, a cloud-based federated data storage solution with interoperable data access services (see Table 1) to local repositories and modular environments that can be configured for a given use case seems to match the data needs of the EU research and medical institutions and of all other stakeholders. The choice of cloud technology provides the ability to manage rapidly growing datasets and provides users with access to the massive computational infrastructure needed for analysis. The federated, cloud based research environment - HRIC- described in this paper, would represent an added value to the entire biomedical and bioinformatics community, because single research institutes and medical institutions lack sufficient infrastructure capacity. The establishment of a transnational HRIC will allow the European research community at large to contribute to the global international leadership required to address societal and scientific challenges through transnational collaborations. In order to ensure the effective and efficient implementation of the European HRIC, the workshop participants endorsed the five recommendations and actions points presented in Table 2 to the EU and all stakeholders.

Application	A set of programs running on one or more computers allowing a user to perform a set of tasks
Cloud Application	An application running in the cloud
Cloud Computing	The delivery of IT services over a network (e.g. the Internet) by means of a combination of infrastructure, software and data hosted by one or more cloud providers using a service model similar to that used by traditional utility companies (e. g. water or electricity)
Cloud Federation	The combination of infrastructure, software and services from separate networks and providers, having shared access mechanisms, to perform common actions, achieve load-balancing or optimize availability or costs
Cloud Instance	A virtual server or container of resources running on a physical host computer possibly hosting several independent instances (see virtualization)
Cloud Marketplace	An online marketplace of cloud services and applications operated by a Cloud Service Provider (CSP)
Cloud Service Provider (CSP)	A company or public entity that offers cloud services to individual users or other entities
Cloud Storage	A model of data storage in which data is hosted across one or more facilities by a hosting entity or CSP and remotely accessed by users over the Internet
Container	A type of virtualized instance running on a physical host server in isolated user spaces and possibly preloaded with applications
Hybrid Cloud	A cloud computing infrastructure comprised of a mix of public and private cloud and on-premise instances and resources
Infrastructure	The combination of hardware resources (network, computing, storage, etc.) and virtualized instances supporting an IT environment
Infrastructure as a Service (IaaS)	A model of cloud computing in which a CSP provides an infrastructure of virtualized resources to users as a service over a network (e.g. the Internet)
On-Premise	It refers to infrastructure or software run on computing resources physically hosted by the entity using them
Platform	A computer system where applications can run on or can be built with
Platform as a Service (PaaS)	A model of cloud computing in which a CSP provides the infrastructure and the platforms where users can run and manage their own applications as a service over a network (e.g. the Internet)
Private Cloud	A cloud infrastructure used by a single organization either on-premise or hosted by a third-party CSP over the Internet or dedicated private networks
Public Cloud	A cloud infrastructure hosted by a CSP (or a Federation) and used by the public or multiple organizations across the Internet

Software as a Service (SaaS)	A model of cloud computing in which a CSP hosts and provides applications (software) to users as a service over a network (e.g. the Internet)
Virtualization	A technology that allows to run a software simulation of a physical computer where a full operating system and applications can be installed

Table 1 : Glossary of cloud computing terms.

Recommendation	Rationale	Action points
Provide and foster standards, good practices and guidelines necessary to establish the HRIC.	The HRIC should be supported by predefined standards, data formats, protocols and templates. The data standards and guidelines applied in the HRIC should be designed as to facilitate interoperability between the diverse health systems and policies in Europe and globally.	Suggest the adoption of data formats and architectures in policies, grant applications and projects calls, throughout the EU and its member states.
Develop and certify the infrastructure and services required for operation of the HRIC.	The HRIC should provide computational infrastructures and services, analytical and visualisation tools to all users as a platform to share knowledge, data and guidelines. Services for data sharing, security and analysis should be compliant with an EU certification system.	- In future grant and project calls related to the development of the HRIC, the EU and the applicants should commit to comply with the highest standards of security, interoperability and reproducibility. - The EU should develop a certification system to validate compliance with the standards mentioned.
Enable the HRIC to operate within an ethical and legal framework adequate for health systems.	A robust ethical and legal framework has to be developed that defines rules for privacy, security, ownership, access and usage of data within the HRIC. A federated system architecture should be preferred as it allows for comparison of data and results, while complying with EU (GDPR) and international data protection and sharing rules.	- In grants and project calls, federated and GDPR-compliant data architectures should be preferred.
Establish a proper environment for the training of a new generation of data and medical scientists.	Education and training of health professionals need to be updated with the HRIC in mind, considering both international standards and practices for data sharing as well as national environments and regulations. The EU should take inspiration in existing large and successful infrastructures that foster multidisciplinary teams, such as CERN.	- Scientists and health professional in training should be made aware of the possibilities of the HRIC. - Communication in relevant professional channels should be strengthened.

Fund public and private initiatives for the development of the HRIC through EU Framework Programmes (Horizon 2010 and Horizon Europe).	The EU and its member states should together with private investors develop a coherent, ambitious and long-term action plan supported by innovative funding mechanisms that consolidate the outcomes from the existing project portfolio into a long-term operational infrastructure.	- The EU should invest through calls and grants in order to build and consolidate the HRIC. - The existing industrial ecosystem should be supported, to remain competitive against the other actors in the world.

Table 2: Summary of the recommendations, details on the rationale and suggestions for actions points directed to the funding agencies and the actors of the field.

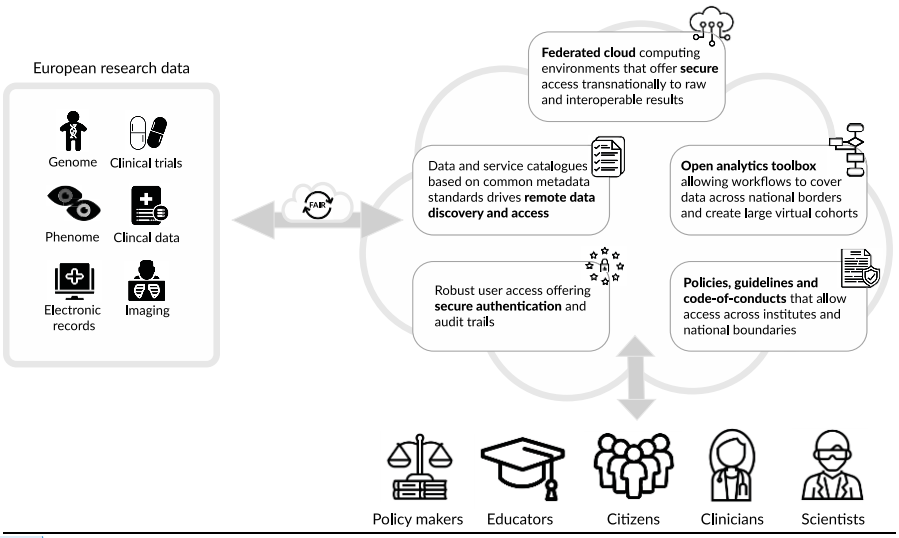

Figure 1 : Proposed general architecture of the HRIC European (inter)national databases, with varying data formats and data types are referenced in a metadata repository, following formatting rules of the Federated data commons as agreed at the HRIC governance level. ~~as agreed at the HRIC governance level.~~ The different users, after access control to the cloud, use the HRIC interface to access the repository, which gathers the relevant data and performs analysis, with outputs such as mathematical models, data visualisations, statistics and patient's profiles, according to the users' needs.

Commented [CA11]: RA3: 'Figure 1 legend has a statement "rules of federated data commons", but does not discuss in the text what those rules are and how a relationship between federated Cloud and the Commons will be established?'

Please consider briefly clarifying this point.

Answer:
[...] following formatting rules of the Federated data commons as agreed at the HRIC governance level.

Table 3: Relevant initiatives for the European Health Research and Innovation cloud

PROJECT	AIMS/SUMMARY	Cloud model used	References
CORBEL project	Creating a platform for harmonised user access to biological and medical technologies, biological samples and data. The project has developed the data harmonisation, ethics guidance and user access protocols necessary for transnational access to both pre-clinical and clinical research infrastructures and is piloting access to participant level data from clinical trials[43].	Scalable cloud-based provision of data access and compute across infrastructures.	[59]
ELIXIR	EUROPEAN RESEARCH INFRASTRUCTURE with 21 members and over 180 research organisations. ELIXIR is creating a	Hybrid cloud ecosystem : i/ Local, private clouds (e.g. EMBL-EBI Embassy);	[60]

	network of local instances of the European Genome-Phenome Archive that give users controlled and secure access to raw data and precomputed results[44,45].	ii/ National community clouds (e.g. cPouta, MetaCentrum cloud, de.NBI); iii/ European research and innovation oriented clouds (e.g. EOSC) iv/ Public/commercial compliant clouds (e.g. Google, Azure, AWS).	
European Translational Information and Knowledge Management Services (eTRIKS)	IMI-funded highly scalable cloud-based platform for translational research, information and knowledge management providing open-source applications that can securely host heterogeneous data types, including multi-omics data, preclinical laboratory data and clinical information, including longitudinal data sets. The platform is a robust translational research knowledge management system that is able to host other data mining applications and support the development of new analytical tools[46].	Scalable cloud-based platform for translational research and applications development. The Openstack technology is used to run a private cloud for eTRIKS.	[61]
European Medical Information Framework (EMIF)	IMI-funded project that has successfully improved access to human health data via providing tools and workflows to discover, assess, access and (re)use human health data. The efforts of this IMI project are being extended through the EHDEN one.	Research analytical service approaches via HER and cohort data platforms.	[62, 63]
Human Brain Project (HBP) Medical Informatics Platform	The HBP Medical Informatics Platform allows researchers around the world to exploit medical data to create machine-learning tools that can analyse these data for new insights into brain-related diseases. The Medical Informatics web-portal[47] provides a software framework based on federated and distributed computing that allows researchers to mine clinical data stored on hospital and laboratory servers, without moving the data from the servers where they reside and without compromising patient privacy.	The HBP Joint Platform plans to adopt in cloud technology and provide those services through its computer centers JSC-Jülich and CSCS-Lugano together with BSC-Barcelona, CINECA-Bologna and CEA-Saclay. Software infrastructure : i/ The (base) infrastructure layer is accessible through an "Infrastructure as a Service" (IaaS) interface; ii/ Tools to enable simulation and modeling as well as data analytics workflows for neuroscience that the HBP operates as a "Platform as a Service" (PaaS);	[64]

		iii/ Several software services for data-driven brain simulations and for virtual neurorobot design and operation offered in form of "Software as a Service" (SaaS). HBP operates the following SaaS: a/ Model driven brain simulations; b/ Neurorobotics simulation and development tools across the whole workflow the of neurorobotics life cycle.	
Human Brain Project (HBP) Knowledge Graph Data Platform	The HPB Knowledge Graph (KG) is an online graph database accepting submissions of anonymized human data, animal data, and models from the brain. All data made discoverable and accessible through KG search have been curated. Data are also made available with integrated multilevel HBP Atlases, holding information about the brain in standard reference spaces.		[65]
Helix Nebula	The project is a pan-European public private partnership initiative led by EIROforum and leading commercial cloud computing partners. Since 2011, it has been piloting the use of cloud computing to enable complex data analyses and large-scale data sharing, with life science-oriented projects ranging from complex genome assembly to assessing somatic variation in the context of different types of cancer.	Helix Nebula Science Cloud (HNSciCloud): Hybrid cloud platform linking together commercial cloud service providers and publicly funded research organisations' in-house IT resources via the GEANT network to provide innovative solutions supporting data intensive science. These services support the connection of the research infrastructures identified in the ESFRI Roadmap to the nascent European Open Science Cloud (EOSC) intended to create a single digital research space for Europe's 1.8 million researchers.	[66]
European Open Science Cloud (EOSC)	The pilot project has recently investigated the benefits of data and cloud computer sharing at a pan-European level with life science-oriented projects on pan-cancer analyses[31] and imaging. This has been pursued across	Cloud based services for open sciences - integration and consolidation of e-infrastructure platforms, federation of existing European research infrastructures and scientific clouds.	[25]

	academic cloud computing environments located in Western and Eastern Europe as well as Canada. Similar initiatives have been launched in the USA[8].		
PCAWG	PCAWG - PANCANCER ANALYSIS OF WHOLE GENOMES The study is an international collaboration to identify common patterns of mutation in more than 2,800 cancer whole genomes from the International Cancer Genome Consortium. This project is exploring the nature and consequences of somatic and germline variations in both coding and non-coding regions, with specific emphasis on cis-regulatory sites, non-coding RNAs, and large-scale structural alterations.	Hybrid cloud model The data coordinating center lists collaborative agreements with cloud providers AWS and an academic compute cloud resource maintained at the cancer collaboratory, by the Ontario Institute for Cancer Research and hosted at the Compute Canada facility.	[67, 68]
RD-Connect	RD-Connect is an integrated platform connecting databases, patient registry data, biobanks and clinical bioinformatics for rare disease research. The RD-Connect allow the integration of different data types (e.g. omics, clinical information, patient registries and biobanks). Those integrated data can be accessed and analysed by the scientific community to speed-up research, diagnosis and therapy development for patients with rare diseases to contribute for better health.	Online secured platform connecting different types of patients related rare disease data enabling genome-phenome analysis.	[69]
COMPARE	COMPARE, a network of collaborators of the Global Microbial Identifier initiative (GMI), aims to improve identification and mitigation of emerging infectious diseases and foodborne outbreaks.	One serve all analytical framework and data exchange platform with various data integration for real time analysis and interpretation of sequence-based pathogen.	[70] [71]

Abbreviations

CERN (the European Organisation for nuclear research)
EMR (Electronic Health Record)
EOSC (European Open Science Cloud)
EU (European Union)
FAIR (Findable, Accessible, Interoperable, Retrievable)

FP7 (The European Union's seventh framework programme for Research, Technological Development and Demonstration)
H2020 (The European Union's Horizon Research and Innovation programme)
HBP (Human Brain project)
HRIC (European Health Research and Innovation Cloud)
OECD (Organisation of Economic Collaboration and Development)
IaaS (Infrastructure as a Service)
PaaS (Platform as a Service)
SaaS (Software as a Service)

Declarations

Availability of data and materials
"Not applicable"

Funding
The workshop participants received funding from the European Union Seventh Programme for Research, Technological Development and Demonstration (FP7) and Horizon Research and Innovation Programme (H2020) and other European Union programmes under the following grant agreements: AETIONOMY (Developing an Aetiology-based Taxonomy of Human Disease-Approaches to Develop a New Taxonomy for Neurological Disorders, IMI-nº115568), ANTI-SUPERBUG PCP (ANTISUPERBUG Precommercial Procurement, H2020-nº688878), B-CAST (Breast CAncer Stratification understanding the determinants of risk and prognosis of molecular sub-types, H2020-nº633784), BRIDGE Health (Bridging information and data generation for evidence-based health policy and research, H2020-nº664691), CASyM (Coordinating Action Systems Medicine – Implementation of Systems Medicine across Europe, FP7-n°305033), CENTER-TBI (Collaborative European NeuroTrauma Effectiveness Research in TBI, FP7-nº602150), CECM (Centre for New Methods in Computational Diagnostics and Personalized Therapy, H2020-nº763734), COLOSSUS (Advancing a Precision Medicine Paradigm in Metastatic Colorectal Cancer: Systems based patient stratification solutions, H2020-nº754923), COMPARE (COllaborative Management Platform for detection and Analyses of (Re-)emerging and foodborne outbreaks in Europe, H2020-nº643476), CONNECARE (Personalised Connected Care for Complex Chronic Patients, H2020-nº689802), CREATIVE (Collaborative REsearch on ACute Traumatic brain Injury in intensiVe care medicine in Europe, FP7-nº602714), CHAFEA grant-nº709723), DEFORM (Define the global and financial impact of research misconduct H2020-nº710246), ECCTR (European Cornea and Cell Transplant Registry, ECRIN-IA (European Clinical Research Infrastructures Network-Integrating Activity, FP7-nº284395), EHR4CR (Electronic Health Records Systems for Clinical Research, IMI-nº115189) eInfraCentral (European E-infrastructures Services Gateway, H2020-nº731049), ELIXIR (European Life-science Infrastructure for Biological Information, FP7-n°211601), ELIXIR-EXCELERATE (Fast track ELIXIR implementation and drive early user exploitation across the life sciences, H2020-nº676559), eMEN (e-mental health innovation and transnational implementation platform North West Europe, H2020), EMIF (European Medical Information Framework, IMI-nº115372), ERA PerMed (ERA-net Cofund in Personalised Medicine, H2020-nº779282), eTRIKS (Delivering European Translational Information and Knowledge Management Services, IMI-1-nº115446), E-COMPARED (European COMPARative

Effectiveness Research on online Depression, FP7- nº603098), EuroPOND (Data-driven models for progression of neurological diseases, H2020-n°666992), EurValve (Personalised Decision Support for Heart Valve disease, H2020-nº689617), HBP SGA1/SGA2 (Human Brain Project specific grant agreements, H2020-n°720270/785907), MASTERMIND (Management of Mental Disorders through Advanced Technologies, CIP-nº621000), MedBioinformatics (Creating medically-driven integrative bioinformatics applications focused on oncology, CNS disorders and their comorbidities, H2020-n°634143), ICT4DEPRESSION (User-friendly ICT tools to enhance self-management and effective treatment of depression in the EU, FP7- n°248778), ImpleMentAll (Towards evidence-based tailored implementation strategies for eHealth, H2020-nº733025), INSTRUCT-ULTRA (Releasing the full potential of instruct to expand and consolidate infrastructure services for integrated structural life sciences research, H2020-nº731005), MeDALL (Mechanisms of the Development of ALLergy, FP7-n°261357) MIDAS (Meaningful Integration of Data, Analytics and Services, H2020-nº 727721), MultipleMS (Multiple manifestations of genetic and non-genetic factors in Multiple Sclerosis disentangled with a multi-omics approach to accelerate personalised medicine, H2020-nº733161), myPEBS (Randomized Comparison Of Risk-Stratified versus Standard Breast Cancer Screening European Women Aged 40-74, H2020-nº755394), OpenAIRE-Advance (Advancing Open Scholarship, H2020-nº777541), OpenMedicine (OpenMedicine, H2020-no643796), PanCareSurFup (PanCare Childhood and Adolescent Cancer Survival Care and Follow-up Studies, FP7-no257505), PIONEER (Prostate Cancer DIagnOsis and TreatmeNt Enhancement through the Power of Big Data in EuRope, H2020-IMI-2-n°777492), PREPARE (Platform for European Preparedness Against (Re-)emerging Epidemics, FP7-n°602525), Regions4PerMed (Interregional coordination for a deep and fast uptake of personalised health, H2020-nº825812), RD-CONNECT (An integrated platform connecting registries, biobanks and clinical bioinformatics for rare disease research, FP7-nº305344), Solve-RD (Solving the unsolved Rare Diseases, H2020-nº779257), SPIDIA4P (SPIDIA for Personalised Medicine-Standardization of generic Pre-analytical procedures for In-vitro DIAgnostics for Personalised Medicine, H2020-nº733112), SYSCID (A Systems medicine approach to chronic inflammatory disease, H2020-nº733100), SysCLAD (Systems prediction of Chronic Allograft Dysfunction, FP7-n°305457), SYSCOL (Systems Biology of Colorectal Cancer, FP7-nº258236), SysMedPD (Systems Medicine of Mitochondrial Parkinson's Disease, H2020-n°668738), U-BIOPRED (Unbiased BIOmarkers for the PREDiction of respiratory disease outcomes, IMI-n°115010), VPH-share (Virtual Physiological Human: Sharing for Healthcare - A Research Environment, FP7-n°269978).
Neither the European Commission nor any other person acting on behalf of the Commission is responsible for the use, which might be made of the following information. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Competing interests

RB is a founder and holder of shares of MEGENO S.A. and holds shares of ITTM S.A. The other authors have declared no competing interests.

Authors' contributions

This publication is inspired by a workshop organized by the Health Directorate at the European Commission. The chairs ADM, CGC, EMF, FS, IP, LC, MB, MBu, ML, MVDB, NBl, RB contributed to the concept of the manuscript. GHS produced the draft workshop report, which was used as a basis for the manuscript. The core editing group CA, NB, BDM, JK, NB, MVDB reviewed

and finalized the manuscript. All authors contributed to the content of the manuscript and approved its final version.

Acknowledgements
We would like to thank the scientific policy officers at the European Commission Health Directorate, who developed the workshop concept, organised it and provided input to this publication: Sasa Jenko, Katarina Krepelkova, Christina Kyriakopoulou, Jana Makedonska, Joana Namorado, Elsa Papadopoulou, Jan Van de Loo and Gregor Schaffrath (National Expert in professional training). The workshop was organised by the Health Directorate of the Directorate-General for Research and Innovation at the European Commission with contributions of the Innovative Medicines Initiative, the Directorate General for Communications Networks, Content and Technology and the Directorate General for Health and Food Safety.

References

1. Birney E, Vamathevan J, Goodhand P: **Genomics in healthcare: GA4GH looks to 2022**. In: *Cold Spring Harbor Laboratory*. bioRxiv; 2017.
2. Langmead B, Nellore A: **Cloud computing for genomic data analysis and collaboration**. *Nature reviews Genetics* 2018, **19**(5):325.
3. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, Bernal-Delgado E, Blomberg N, Bock C, Conesa A *et al*: **Making sense of big data in health research: Towards an EU action plan**. *Genome medicine* 2016, **8**(1):71.
4. Mazein A, Ostaszewski M, Kuperstein I, Watterson S, Le Novere N, Lefaudeux D, De Meulder B, Pellet J, Balaur I, Saqi M *et al*: **Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms**. *NPJ systems biology and applications* 2018, **4**:21.
5. Ostaszewski M, Gebel S, Kuperstein I, Mazein A, Zinovyev A, Dogrusoz U, Hasenauer J, Fleming RMT, Le Novere N, Gawron P *et al*: **Community-driven roadmap for integrated disease maps**. *Briefings in bioinformatics* 2019, **20**(2):659-670.
6. Phillips M, Molnar-Gabor F, Korbel JO, Thorogood A, Joly Y, Chalmers D, Townend D, Knoppers BM, Consortium P: **Of Clouds and Genomic Data Protection**. *Nature* 2019.
7. PCAWG C: **Pan-cancer analysis of whole genome**. *Nature* 2019.
8. Lochmuller H, Badowska DM, Thompson R, Knoers NV, Aartsma-Rus A, Gut I, Wood L, Harmuth T, Durudas A, Graessner H *et al*: **RD-Connect, NeurOmics and EURenOmics: collaborative European initiative for rare diseases**. *European journal of human genetics : EJHG* 2018, **26**(6):778-785.
9. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE *et al*: **The FAIR Guiding Principles for scientific data management and stewardship**. *Scientific data* 2016, **3**:160018.
10. Lamy P, Brudermüller M, Ferguson M, Friis L, Garmendia C, Gray I, Gulliksen J, Kulmala H, Maher N, Plentz Fagundes M *et al*: **LAB - FAB - APP Investing in the European future we want**. In. Brussels: Independent High Level Group on maximising the impact of EU research & Innovation Programmes; 2017.
11. Hauser H, Bergman N, Bruncko M, Cosgrave P, Dwyer G, Helder M, Hinrikus T, Hoerr I, Karia B, Kolar J *et al*: **Funding - Awareness - Scale - Talent (FAST) Europen is**

back: Accelerating breakthrough innovation. In. Luxemburg: European Commission; 2018.
12. **Implementation Roadmap for the European Open Science Cloud** [https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none]
13. Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD: **Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud**. *Information Services & Use* 2017, **37**:49-56.
14. **FAIRsharing homepage** [https://fairsharing.org/]
15. **REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL** [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679]
16. Bouzaglo D, Chasida I, Ezra Tsur E: **Distributed retrieval engine for the development of cloud-deployed biological databases**. *BioData mining* 2018, **11**:26.
17. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, Knoetze N, Breden FMW, Christley S, Scott JK *et al*: **iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories**. *Immunological reviews* 2018, **284**(1):24-41.
18. Sreenivasaiah PK, Kim DH: **Current trends and new challenges of databases and web applications for systems driven biological research**. *Frontiers in physiology* 2010, **1**:147.
19. Wade TD: **Traits and types of health data repositories**. *Health information science and systems* 2014, **2**:4.
20. Estiri H, Klann JG, Weiler SR, Alema-Mensah E, Joseph Applegate R, Lozinski G, Patibandla N, Wei K, Adams WG, Natter MD *et al*: **A federated EHR network data completeness tracking system**. *J Am Med Inform Assoc* 2019, **26**(7):637-645.
21. Shenai S, Aramudhan M: **Cloud computing framework to securely share health & medical records among federation of healthcare information systems**. *Biomedical research* 2018; Special Issue: S133-S136~~7~~.
22. Oderkirk J: **Readiness of electronic health record systems to contribute to national health information and research**. In. Paris: OECD Health Working Papers; 217: 80.
23. **eHealth: Digital health and care** [https://ec.europa.eu/health/ehealth/projects/nationallaws_electronichealthrecords_en]
24. **Overview of the national laws on electronic health records in the EU Member States and their interaction with the provision of cross-border eHealth services - Final report and recommendations** [https://ec.europa.eu/health/sites/health/files/ehealth/docs/laws_report_recommendations_en.pdf]
25. **EOSC homepage** [https://www.eoscpilot.eu/]
26. Molnar-Gabor F, Lueck R, Yakneen S, Korbel JO: **Computing patient data in the cloud: practical and legal considerations for genetics and genomics research in Europe and internationally**. *Genome medicine* 2017, **9**(1):58.
27. **EU countries will cooperate in linking genimc databases across borders** [https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders]
28. **Declaration of cooperation towards access to at least 1 million sequenced genomes in the European Union by 2022** [https://www.euapm.eu/pdf/EAPM_Declaration_Genome.pdf]

Field Code Changed

Field Code Changed

29. Saunders G, Baudis M, Becker R, Beltran S, Béroud C, Birney E, Brooksbank C, Brunak S, Van den Bulcke M, Drysdale R *et al*: **Leveraging European infrastructures to access 1 million human genomes by 2022**. *Nature Reviews Genetics* 2019.
30. Drosatos G, Kaldoudi E: **Blockchain Applications in the Biomedical Domain: A Scoping Review**. *Computational and structural biotechnology journal* 2019, **17**:229-240.
31. Esposito C, De Santis A, Tortora G, Chang H, Choo K-KR: **Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?** *IEEE Cloud Computing* 2018, **5**.
32. Miller JB: **Big data and biomedical informatics: Preparing for the modernization of clinical neuropsychology**. *The Clinical neuropsychologist* 2019, **33**(2):287-304.
33. **The eTRIKS code of practice on secondary use of medical research data** [https://www.etriks.org/code-of-practice/]
34. De Hert P, Papakonstantinou V: **Three Scenarios for International Governance of Data Privacy: Towards an International Data Privacy Organization, Preferably a UN Agency?** *I/S: A Journal of Law and Policy for the Information Society* 2013, **9**(2):271-324.
35. **EU picks team of 11 to run giant European Open Science Cloud** [https://sciencebusiness.net/science-cloud/news/eu-picks-team-11-run-giant-european-open-science-cloud]
36. **Human brain project ethics and society sub-projects** [https://www.humanbrainproject.eu/en/social-ethical-reflective/]
37. Salles A, Bjaalie JG, Evers K, Farisco M, Fothergill BT, Guerrero M, Maslen H, Muller J, Prescott T, Stahl BC *et al*: **The Human Brain Project: Responsible Brain Research for the Benefit of Society**. *Neuron* 2019, **101**(3):380-384.
38. **8 Benefits and Risks of Cloud Computing in Healthcare** [https://solutionsreview.com/cloud-platforms/8-benefits-and-risks-of-cloud-computing-in-healthcare/]
39. Ali O, Shrestha A, Soar J, Wamba F: **Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review**. *International Journal of Information Management* 2018, **43**:146-158.
40. Lian JW: **Establishing a Cloud Computing Success Model for Hospitals in Taiwan**. *Inquiry : a journal of medical care organization, provision and financing* 2017, **54**:46958016685836.
41. de Bruin B, Floridi L: **The Ethics of Cloud Computing**. *Science and engineering ethics* 2017, **23**(1):21-39.
42. Council CSC: **Impact of Cloud Computing on Healthcare**. In.; 2017.
43. Sultan N: **Making use of cloud computing for healthcare provision: Opportunities and challenges**. *International Journal of Information Management* 2014, **34**(2):177-184.
44. Sobeslav V, Maresova P, Krejcar O, Franca TC, Kuca K: **Use of cloud computing in biomedicine**. *Journal of biomolecular structure & dynamics* 2016, **34**(12):2688-2697.
45. Rodrigues JJPC, Sendra Compte S, de la Torra Diez I: **Cloud Computing on e-Health**. In: *e-Health Systems, Theory, Advances and Technical Applications.* Edited by Elsevier IP-; 2016: 191-207.
46. **CERN Openlab homepage** [https://openlab.cern/]
47. **Worldwide LHC Computing Grid homepage** [http://wlcg.web.cern.ch/]
48. Chen J, Qian F, Yan W, Shen B: **Translational biomedical informatics in the cloud: present and future**. *BioMed research international* 2013, **2013**:658925.

Field Code Changed

Field Code Changed

Field Code Changed

49. Khan N, Yaqoob I, Hashem IA, Inayat Z, Ali WK, Alam M, Shiraz M, Gani A: **Big data: survey, technologies, opportunities, and challenges**. *TheScientificWorldJournal* 2014, **2014**:712826.
50. Kuo AM: **Opportunities and challenges of cloud computing to improve health care services**. *Journal of medical Internet research* 2011, **13**(3):e67.
51. Navale V, Bourne PE: **Cloud computing applications for biomedical science: A perspective**. *PLoS computational biology* 2018, **14**(6):e1006144.
52. LAB, FAB, APP: **Investing in the European future we want.** In. Luxemburg: Publication office of the European Union; 2017.
53. Mazzucato M: **Mission-Oriented Research & Innovation in the European Union - a problem-solving approach to fuel innovation-led growth**. In. Luxemburg: Publication office of the European Union; 2018.
54. **EIT in Horizon Europe (2021-2027) - complementarities and synergies with the EIC** [https://eit.europa.eu/sites/default/files/eit_position_paper_horizon_europe_eit_eic_0.pdf]
55. H2020 Scientiic Panel for Health: **Building the future of health research - Proposal for a European Council for Health Research** In.; 2018.
56. **European Joint Programme of Rare Diseases homepage** [https://www.ejprarediseases.org/index.php/about/]
57. **European Research Council homepage** [https://erc.europa.eu]
58. **European Structural and Investment Funds** [https://ec.europa.eu/info/funding-tenders/funding-opportunities/funding-programmes/overview-funding-programmes/european-structural-and-investment-funds_en]
59. **CORBEL project homepage** [http://www.corbel-project.eu]
60. **ELIXIR infrastructure homepage** [https://www.elixir-europe.org/]
61. **The eTRIKS project homepage** [https://www.etriks.org/]
62. **EMIF project homepage** [http://www.emif.eu/]
63. **EHDEN project homepage** [https://www.ehden.eu/]
64. **Human Brain Project homepage** [https://www.humanbrainproject.eu/en/medicine/]
65. **Human Brain Project Knowledge Graph** [https://www.humanbrainproject.eu/en/explore-the-brain/]
66. **HELIX nebula homepage** [http://www.helix-nebula.eu/]
67. [https://dcc.icgc.org/icgc-in-the-cloud/collaboratory]
68. **PCAWG project homepage** [https://dcc.icgc.org/pcawg]
69. **RD Connect homepage** [https://rd-connect.eu/]
70. **COMPARE project homepage** [http://www.compare-europe.eu]
71. [https://www.globalmicrobialidentifier.org]

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed