



OPEN ACCESS

EDITED BY

Christine Gispen-de Wied,
Gispen4RegulatoryScience, Netherlands

REVIEWED BY

Jackson Burton,
Biogen Idec (United States), United States

*CORRESPONDENCE

Elisabetta Vaudano
✉ elisabetta.vaudano@ihi.europa.eu

RECEIVED 03 January 2025

ACCEPTED 25 February 2025

PUBLISHED 01 April 2025

CITATION

Vaudano E (2025) The Innovative Health Initiative public–private partnership: a catalyst for big data-driven health research and innovation.
Front. Med. 12:1554948.
doi: 10.3389/fmed.2025.1554948

COPYRIGHT

© 2025 Vaudano. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The Innovative Health Initiative public–private partnership: a catalyst for big data-driven health research and innovation

Elisabetta Vaudano*

Innovative Health Initiative Joint Undertaking, Brussels, Belgium

The use of “Big Data” in health research could be a game changer to revolutionize the field and drive innovation. However, the complexity of health data and of the healthcare ecosystem makes big data applications challenging. To overcome these challenges, it is essential that all stakeholders, both data owners and users, feel comfortable with the data use and the results of the analyses. This can only be achieved using a collaborative approach that prioritizes transparency and trust. The European Innovative Health Initiative (IHI) is a prime example of a successful public–private partnership that has achieved this. By fostering a culture of collaboration and trust, the IHI has enabled unprecedented access to large amounts of health data, improved data quality and standards, and created valuable resources and knowledge for progress in areas of unmet public health need. This perspective explores the IHI model and highlights examples from its projects that demonstrate its power as a catalyst for big data-driven health research and innovation. By sharing lessons learned and best practices, we aim to contribute to the development of a robust data-driven evidence base for the translation of research into improved patient care practices and products.

KEYWORDS

collaboration, data access, quality, patients, trust, big data, public–private partnership, health

1 Introduction

The use of “Big Data” in health research has revolutionized the way we approach medical research, diagnosis, and treatment. Big data are extremely large and diverse collections of structured and unstructured data from sources such as electronic health records (EHRs), omics, imaging, and wearables. The progress in advanced analytics and artificial intelligence (AI) (most recently generative AI) has enabled using such rich sources of information to uncover hidden patterns, trends, and correlations within large datasets for a deeper understanding of complex phenomena and unprecedented speeding up of innovation (1).

However, health data are complex and sensitive, and the ecosystem of data owners and users is a mix of public and private stakeholders, making big data applications challenging. For success, it is necessary all actors feel comfortable with the data use and the results of the analyses. This can only be achieved using a collaborative approach based on transparency and trust among all stakeholders. Trust is the cornerstone of public–private partnerships (PPPs), such as those created by the European Innovative Health Initiative (IHI) (2), making these ideal instruments for addressing the challenges of using big data in health research, fostering collaboration, and driving innovation in this critical field. This perspective presents some examples from selected projects (Table 1) of what can be achieved through public–private collaboration.

TABLE 1 List of IHI projects cited in the text with the acronym, title, link to website (or, in the absence of this, to the factsheet on the IHI website), and a short description of project scope.

Acronym, Title, Website	Project scope
AIMS-2-TRIALS Autism Innovative Medicine Studies – 2 – Trials https://www.aims-2-trials.eu/	A large-scale European research project focused on autism spectrum disorders (ASDs). It aims to improve understanding, identify, and validate biomarkers, foster the development of new treatments, and create a clinical trial network for autism across Europe. The project seeks to advance personalized medicine approaches for individuals with ASDs, exploring autism's development from prenatal stages to adulthood.
BD4BO Big Data for Better Outcomes https://bd4bo.eu/	A program aimed to improve the health outcomes of patients in Europe by maximizing the potential of big data through collaborative research projects. It created big data platforms focused on specific disease areas, including Alzheimer's, blood cancers, cardiovascular diseases, and prostate cancer.
BIGDATA@HEART Big Data at Heart https://www.bigdata-heart.eu/	A project that developed a big data-driven platform for cardiovascular research, focusing on improving outcomes in heart diseases.
BIGPICTURE Central repository for digital pathology https://bigpicture.eu/	A project creating a European repository of 3 million digital pathology slides from both humans and laboratory animals, providing a comprehensive dataset for AI tool development across a broad spectrum of diseases and medical conditions where pathology is critical for diagnosis and research.
CONCEPTION Building an ecosystem for better monitoring and communicating of medication safety in pregnancy and breastfeeding: validated and regulatory endorsed workflows for fast, optimized evidence generation https://www.imi-conception.eu/	A project that created a pan-European network that systematically generates and disseminates reliable evidence-based information on medication safety during pregnancy and breastfeeding.
DIRECT Diabetes research on patient stratification https://www.ih.europa.eu/projects-results/project-factsheets/direct	A project that identified biomarkers for the prediction of disease progression and treatment response to develop personalized medicine approaches for type 2 diabetes.
EHDEN European Health Data and Evidence Network https://www.ehdn.eu/	A project aimed to harmonize European health data at a large scale (100 million) to a common model, facilitating large-scale analysis and research to improve patient care and outcomes. It created a unique large-scale federated European health data network for rapid, reproducible research.
EMIF European Medical Information Framework https://www.emif.eu/	A project that created a unified platform to discover, assess, and reuse healthcare data across Europe, focusing on Alzheimer's disease and metabolic complications of obesity. It developed tools for data harmonization, ethical governance, and collaborative research.
EPAD European prevention of Alzheimer's dementia consortium https://ep-ad.org/	A project aimed at facilitating the secondary prevention of Alzheimer's disease (AD). It recruited over 2,000 participants for a longitudinal cohort study across Europe and developed a master protocol for AD secondary prevention-adapted trials.
EPND European Platform for Neurodegenerative Diseases https://epnd.org/	A project that is creating a sustainable European platform to accelerate biomarker discovery for neurodegenerative diseases by facilitating access to clinical samples and data across Europe.
EQIPD European Quality In Preclinical Data https://go-eqipd.org/	A project aimed at improving the quality and reproducibility of preclinical research data through standardized quality systems and educational resources. It developed guidelines, tools, and a certification process to enhance scientific rigor and data reliability.
eTOX Integrating bioinformatics and cheminformatics approaches for the development of Expert systems allowing the <i>in silico</i> prediction of toxicities https://www.ih.europa.eu/projects-results/project-factsheets/etox	A project that developed a comprehensive toxicology database and predictive models to improve drug safety assessment in the early stages of pharmaceutical development, reducing animal testing, and accelerating drug discovery. It built the basis for the eTRANSafe project.
eTRANSafe Enhancing TRANslational SAFEty Assessment through Integrative Knowledge Management https://www.ih.europa.eu/projects-results/project-factsheets/etransafe	A project aimed to enhance translational safety analysis and prediction. It developed an integrated database and computational tools to improve drug safety assessment using preclinical and clinical data from pharmaceutical companies and public sources.
EU-AIMS European Autism Interventions - a Multicentre Study for Developing New Medications https://www.ih.europa.eu/projects-results/project-factsheets/eu-aims	A project aimed at accelerating the development of new treatments for autism spectrum disorders (ASDs) through comprehensive research and collaboration. It created one of the world's largest autism databases, containing extensive medical information on individuals with ASDs. It built the basis for the AIMS-2-TRIALS project.

(Continued)

TABLE 1 (Continued)

Acronym, Title, Website	Project scope
EUROPAIN Understanding chronic pain and improving its treatment https://www.ih.europa.eu/projects-results/project-factsheets/europain	A project aimed to improve understanding of pain mechanisms and accelerate analgesic development. It transformed the neuropathic pain field through collaborative research, enhancing preclinical and clinical studies via novel animal models and objective pain measures as well as translationally validated novel drug targets.
FAIRPlus FAIRPlus https://fairplus-project.eu/	A project aimed to improve data sharing in life sciences by developing tools and guidelines to make research data findable, accessible, interoperable, and reusable (FAIR). It established a robust methodology for prioritizing and selecting datasets for FAIRification, providing practical guidance to improve FAIRness for existing and new datasets.
HARMONY ALLIANCE Healthcare alliance for resourceful medicines offensive against neoplasms in hematology https://www.harmony-alliance.eu/	Alliance of the two projects HARMONY and HARMONY plus that created a big data analytics platform to accelerate the development of more effective treatments for blood cancer patients.
HIPPOCRATES Health initiatives in psoriasis and psoriatic arthritis consortium European states https://www.hippocrates-imi.eu/	A project aimed at improving early diagnosis, prevention, and personalized treatment of psoriatic arthritis (PsA) through advanced research on the molecular basis of PsA, data integration, and patient involvement.
HYPORESOLVE Hypoglycaemia - REdefining SOLutions for better liVEs https://hypo-resolve.eu/	A project aimed at redefining hypoglycemia in diabetes, investigating its causes, consequences, and economic impact. It developed a unique database, identified links between hypoglycemia and cardiovascular events, revealed inflammation's role, and developed a quality-of-life measure for hypoglycemia in diabetes patients.
IDERHA Integration of heterogeneous data and evidence toward regulatory and HTA acceptance https://www.iderha.org/	A project aiming to create a scalable platform integrating diverse health data to improve patient outcomes, focusing initially on lung cancer. It seeks to enhance personalized care and support healthcare professionals, patients, and researchers.
MELLODDY Machine learning ledger orchestration for drug discovery https://www.ih.europa.eu/projects-results/project-factsheets/melloddy	A project that developed a secure (blockchain-based) federated learning platform for drug discovery, enabling pharmaceutical companies to collaborate on machine learning models without sharing proprietary data. It successfully deployed the world's first secure multi-task federated learning platform for drug discovery.
NEWMEDS Novel methods leading to new medications for depression and schizophrenia https://www.ih.europa.eu/projects-results/project-factsheets/newmeds	A project focused on developing new methods for drug discovery in schizophrenia and depression. It created standardized animal models, imaging tools, and one of the largest databases of schizophrenia clinical trials, for improving drug development efficiency. It also made significant contributions to understanding genetic risk factors in schizophrenia:
NEURONET Efficiently networking European neurodegeneration research https://imi-neuronet.org/	A coordination and support initiative that developed different tools to enhance collaboration, synergy, and visibility across IMI's neurodegenerative disorder research projects, aiming to maximize their collective impact.
OPTIMA Optimal treatment for patients with solid tumors in Europe through artificial intelligence https://www.optima-oncology.eu/	A project using AI to improve care for prostate, breast, and lung cancer patients. It aims to develop a GDPR-compliant, real-world oncology data platform, and AI-based decision support tools.
PIONEER Prostate cancer diagnosis and treatment enhancement through the power of big data in Europe https://prostate-pioneer.eu/	A project aimed to improve prostate cancer care by leveraging big data. It developed a comprehensive data platform to address key knowledge gaps in screening, diagnosis, and treatment and in integrating real-world datasets. It conducted analyses to enhance patient outcomes and healthcare efficiency.
Screen4Care Shortening the path to rare disease diagnosis by using newborn genetic screening and digital technologies https://screen4care.eu	A project aimed at accelerating rare disease diagnosis through genetic newborn screening and AI-based tools. It develops a digital infrastructure for patient engagement, AI algorithms for early detection, and a meta-symptom checker to support faster diagnosis.
VICT3R Developing and implementing virtual control groups to reduce animal use in toxicology research https://www.vict3r.eu/	A project aimed at reducing animal use in toxicology research by developing virtual control groups (VCGs) to replace concurrent control groups in non-clinical drug and chemical safety evaluations.

2 The Innovative Health Initiative (IHI) collaboration model

The Innovative Health Initiative (IHI)¹ builds on the Innovative Medicines Initiative (IMI). The IHI expands the IMI's focus on pharmaceutical research to a broader, cross-sectoral approach to health innovation, addressing evolving scientific and societal needs. It is a unique framework that brings together private health industry sectors (pharma, biotech, vaccines, and medical technologies) and public partners (e.g., academia, healthcare professionals and providers, patients and caregivers, regulators, health technology assessment bodies, and payers) to conduct research in areas of unmet public health need. The IHI, with an overall budget of 2.4 billion EUR, is the largest initiative of its kind in the health space. Funded projects are precompetitive collaborative “consortia,” leveraging at scale the multidisciplinary skills of all members, public European funding, and in-kind industry resources (experts' time, setting and running of assays and models, clinical studies, etc.) toward the achievement of agreed research objectives. The terms of collaboration and governance in the IHI consortia are established from the start by the Grant Agreement (GA) that all partners must sign with the IHI and by the Consortium Agreement (CA) that all members of the consortium must sign among themselves.² The GA provisions are common for all IHI contracts, while those of the CA are agreed by the consortium members including details for data sharing and access and for intellectual property rights (IPRs). Thus, all partners enter the collaboration with an upfront understanding of their own knowledge/data/assets to be shared with the other collaborators for the implementation of the agreed research program, as well as on rules for access to results and data within and beyond the consortium, guaranteeing respect for the rights of all collaborators and that the collaboration will be rewarding for all involved. Working in the IHI precompetitive trusted framework, collaborating, and sharing data (among companies and with public actors) allows us to learn together to overcome key research and innovation bottlenecks, instead of working in silos. This gives IHI projects the unique ability to access, collate, and make further available an increasing volume and diversity of high-quality data-generating impacts beyond the single projects.

3 How the IHI public–private partnerships have enabled unprecedented (big) data access and connectivity

If the large amount of high-quality data existing in the health research ecosystem could be curated, connected, and made accessible for research use, it would unlock unprecedented possibilities for a comprehensive understanding of biological phenomena, patient journeys, treatment paths, and health outcomes. This is not a trivial endeavor due to legal, technical, data protection, ethical, and IPR

challenges and requires an enabling legal framework and strong data and knowledge management (DKM).

Under the IHI framework, industries work together in a “safe” precompetitive space, which enables large-scale access to high-quality data generated and collected by industry that is normally siloed within individual companies due to their IPR sensitivity. In addition, DKM has been a central theme from the start in the IMI (3) and even more under the IHI, where one of the overall specific objectives is to “*exploit the full potential of digitalisation and data exchange in healthcare*” (4). In all IHI projects, the data governance and stewardship principles agreed in the consortium agreement are further developed in the *data management plan*. This robust data policy has uniquely enabled data access and connectivity of large amounts of both public and corporate data (including from clinical trials) to create highly curated and standardized databases (e.g., projects DIRECT, EPAD, EUROPAIN, HYPORESOLVE, HIPPOCRATES, NEWMEDS, and projects highlighted later) to address key research questions and deliver high-quality research outputs (5) with impacts in important areas such as Alzheimer's disease (6) and diabetes (7).

The impact of IHI partnerships goes beyond a single project, creating lasting collaborative networks to sustain and accelerate the progress of data-driven research and deliver new models, solutions, and practices to make the most of the big data potential (8–10).

A pivotal example is from the IHI projects in drug safety. The eTOX project pioneered the “honest broker” approach (11) to integrate historical toxicological data from 13 pharma companies with publicly available data and share them among 30 public and private stakeholders. The approach was so successful that it was extended to the sharing of clinical data in the large-scale follow-up eTRANSFAE (12). eTRANSFAE overall outputs are a reusable framework and provide best practices for sustainable and reproducible findable, accessible, interoperable, and reusable (FAIR) data and knowledge sharing to drive integrative drug safety research at scale. This framework can be adopted by other research communities to integrate public and private data in the biomedical research space, expanding its impact far beyond drug safety (13).

eTRANSFAE also developed the concept of “virtual control groups,” harnessing the wealth of data on control groups in toxicity tests. This will be developed up to regulatory acceptance in the new IHI project VICT3R (14), showing how the IHI framework generates long-lasting collaborative networks to drive sustainable progress in data-driven research.

The MELLODDY project took a different but even more powerful approach to enable the sharing and combining of proprietary datasets without actually sharing them, exposing them, or even moving them from where they are housed, via blockchain and federated learning (FL). MELLODDY created the largest FL-based study across data from 10 pharmaceutical companies and demonstrated the power of combining multiparty privacy-protected machine learning (ML) techniques for cross-end point federated learning in drug discovery, analyzing a dataset of 2.6+ billion confidential experimental activity data points, documenting 21+ million physical small molecules, and 40+ thousand assays in on-target and secondary pharmacodynamics and pharmacokinetics (15). MELLODDY FL model offers a powerful tool to accelerate drug discovery, improve predictive modeling, and potentially reduce time and cost associated with bringing new treatments to market, all while maintaining the highest standards of

1 Innovative Health Initiative (IHI). <https://www.ihf.europa.eu/>; for simplicity, we will use IHI also to cover IMI.

2 <https://www.ihf.europa.eu/apply-funding/call-documents>

data security and privacy. The project software outputs and methodologies are available openly via GitHub³ for reuse by the research community.

The EMIF project has addressed the challenge of collecting, integrating, and making accessible sensitive patient data: Its platform enabled data access through a federated network of data sources, allowing researchers to find, assess, and reuse health data from a network of 18 population-based and 60 cohort-based data sources across Europe. The platform provided a single point of access for searching aggregated data across different sources and countries. Researchers could browse metadata, assess data suitability, and initiate new research studies with the data sources in a secure environment (16). The EMIF catalog, a key component of the platform, integrated biomedical metadata from multiple institutions using a common metadata schema, allowing researchers to evaluate data source suitability and request access to specific data sources. Data custodians had full control over data visibility levels, ensuring fine-grained access control to the metadata, and a role-based access control system was deployed to enforce EMIF access policies (16). The EMIF project established a sustainable model for sharing healthcare data across Europe, promoting a wider adoption of data-sharing principles (16).

These principles were the basis of the EHDEN project, which was part of the BD4BO program (17). EHDEN built a federated data network for access to the health data of 100 million EU citizens standardized to a common data model. EHDEN has been successful beyond expectations in creating a highly collaborative network of data partners across Europe, supporting them in mapping their data to a common standard and in participating in open science collaborative study-a-thons that have delivered highly significant results (18). These range from demystifying the relation between COVID-19 vaccines and the development of neurological disorders (19) to demonstrating the utility of routinely collected health data to support the validation and prioritization of pharmacovigilance signals (20). Building on EHDEN's successes, a new legal entity, the EHDEN Foundation,⁴ was created to continue fostering a strong and growing open science community.

A novel area that is spearheaded by the IHI is that of “big” pathology imaging. The BIGPICTURE project is creating the first European ethical and regulatory compliant community-based platform for storing, sharing, and processing millions of quality-controlled annotated pathology images and artificial intelligence algorithms. This required the development of a tailor-made data-sharing agreement to align all 44 public and private partners. The BIGPICTURE platform for digital pathology will drastically accelerate “AI pathology” and accelerate the progress from drug discovery to clinical diagnostics (21).

The learnings on data sharing across the IHI projects are gathered in a Playbook (22) to help future participants in precompetitive consortia in tackling internal processes and decisions and accelerate the provision of, and access to, data from the partners. For example, the Playbook recommends that to navigate the GDPR⁵ consortia should determine data flow and GDPR roles early, consider impact

assessments, ensure proper consent, use standardized templates, and address data protection systematically for efficient compliance and sharing. This will create a positive impact on resources, help projects make progress faster, and contribute to a stronger spirit of collaboration.

4 How the IHI PPPs are contributing to enhancing data quality and standards

The success of big data-driven research depends critically on the quality of the data. Issues of data quality and standardization could significantly decrease value and lead to significantly biased results. In the IHI partnerships, the robustness of the data generation methodologies and use of standards is a priority to ensure replicability of results [e.g., EUROPAIN (23)], and important resources have been developed to help the broad scientific community to enhance both.

Two examples are the EQIPD project, which delivered a framework for rigor in preclinical animal experiments (24) and training modules available globally via the Partnership for Assessment and Accreditation of Scientific Practice (PAASP),⁶ and the BIGDATA@HEART project, which developed the CODE-EHR, a minimum standards framework to be used by researchers and clinicians to improve the design of studies and enhance the transparency of study methods for robust and effective utilization of healthcare data in research (25).

The FAIR data principles are the basis for the successful sharing and reuse of data in research. The IHI fostered the implementation of these key principles via the FAIRplus project, which has produced a FAIR Cookbook with practical “recipes” for researchers to make their data FAIR (26).

The use of standards and common data models has been key to the success of the IHI projects in making a large amount of data shareable (see previous section) and allowed delivering reliable evidence-based information on important areas such as risks and benefits of medication use in pregnancy (27) and in rare diseases (28). It has also enabled synergies with projects beyond IHI, like in the case of cancer projects, to further boost impact in this critical area (29).

The IHI's focus on data quality enabled its projects to contribute to the progress of regulatory science (30) and to impactful collaborations with regulatory authorities. A prime instance is again EHDEN, an exemplar open science collaborative consortium whose high level of trustworthiness led to long-term collaboration with the European Medicines Agency (EMA). EHDEN contributed to the infrastructure for real-world monitoring of treatments and vaccines during the COVID-19 pandemic⁷ and now its work is leveraged for the establishment of the Coordination Centre for the Data Analysis and Real-World Interrogation Network (DARWIN EU®).⁸

6 <https://paasp.net/>

7 <https://www.ema.europa.eu/en/news/covid-19-ema-sets-infrastructure-real-world-monitoring-treatments-and-vaccines>

8 <https://www.ehden.eu/darwin-eu-emc/>

3 <https://github.com/melloddy>

4 <https://www.ehden.eu/vision-and-mission/ehden-legal-entity/>

5 GDPR: General Data Protection Regulation.

5 IHI PPPs boost progress in big data research in areas of high public health need: the example of neurodegeneration, autism, and cancer

A research area where globally there has been massive investment in both public funding and industry resources is that of neurodegeneration. This has generated a huge body of data, both preclinical and patient-derived. However, such data resources are highly fragmented, difficult to access, poorly standardized, and of variable quality. Federated data-sharing approaches can address at least some of these issues, as shown by the IHI and others (31, 32). However, a more fit-for-purpose and collaborative framework is needed to enable access across public and private data collections to significantly boost the availability of data for federated analysis. The NEURONET project gathered important learnings across the IHI portfolio of neurodegenerative projects, identifying as enablers the principles of transparency, standardization, and co-design—from open, accessible metadata catalogs that enhance findability of data, to measures that increase visibility and trust in data reuse (10). The EPND project (33) is going a step further in developing a scalable and sustainable platform for high-quality sample and data sharing in neurodegeneration. Importantly, EPND leverages the existing data platform ADDI⁹ AD Workbench to connect to an existing global network of data scientists and datasets. This highly collaborative approach further promotes open innovation and the generation of additional resources for the research community and, in doing so, speeds up research progress in neurodegeneration.

Despite the potential promise of big data in psychiatry, the sharing of data in this area has long posed ethical and logistical challenges due to research data capturing human characteristics and experiences, often including sensitive information, such as demographic factors and genetic data. These issues are further amplified in child and adolescent psychiatry, where key decisions regarding research participation and data use may be made by a legal guardian on behalf of a young person. The EU-AIMS and follow-up AIMS-2-TRIALS projects have succeeded in tackling these challenges using a powerful combination of striving for data quality and using mechanisms for shared control of data curation between researchers and participants such as dynamic consent (34). They created the largest collection of multimodal longitudinal data in autism (35), the EAGER registry (36), including 1,500 re-contactable participants with a diagnosis of autism or associated rare genetic condition who have undergone whole-genome sequencing, and aim to make both data networks globally shareable.¹⁰

Cancer is another key area where big data applications can help improve patients' quality of life if patient data can be gathered with sufficient high quality and respect for regulations. To this end, the IHI has created several best-practice data platforms for different malignancies. The HARMONY Alliance (blood cancer) and PIONEER (prostate cancer) platforms include large amounts of

harmonized and anonymized data, plus data processing pipelines and analytics ensuring data safety and quality, as well as patient privacy, allowing scientists to answer research questions that cannot be addressed with other methods (37–40). The OPTIMA project (prostate, breast, and lung cancer) is establishing a secure, interoperable, large-scale evidence, GDPR-compliant data platform that includes RWD from >200 million people, data analysis tools, FL tools, and AI algorithms to identify factors supporting individualized care decisions for oncologic patients (41). The recent IDERHA (lung cancer) infrastructure connects existing data providers' resources (low and high-dimensional data) so users can access and analyze the data compliant with data standards, GDPR, and regulatory requirements, enabling secondary use of data while ensuring full control by citizens of their personal health data. IDERHA is building one of the first pan-European health data spaces, aligned with the implementation of the upcoming European Health Data Space (EHDS) (42).

6 Discussion

The IHI collaboration model has demonstrated significant success in enabling large-scale data sharing and connectivity in health research. By creating a trusted precompetitive framework, the IHI has overcome key bottlenecks in this challenging domain and has generated a significant body of high-quality (big) data, as well as infrastructures and learnings for fostering progress in various health areas and enabling faster, more robust innovation while respecting the rights of all stakeholders, particularly patients. Having the patients as active partners in all data activities and decisions on data sharing/access has been essential to justify their transparency and legitimacy (39, 43, 44), ensuring research answers questions that really match the need of the affected people, ultimately boosting the quality of its outputs. This approach is giving fruit as high-quality results and impacts, including showing how big data analytics can dramatically accelerate and make more robust evidence generation to inform clinical decision-making for better patient health outcomes (45, 46). The high scientific quality of the teams, public–private cross-learning, and careful design and monitoring of the activities ensure the project objectives keep and often go beyond the innovation state-of-the-art.

Many other funders and projects are following this path, but the IHI has done it at scale, and the results of the projects are fostering sustainable ecosystems to carry on learning and innovating, as flagged in the most recent evaluation report by the European Commission (47).

IHI partnerships prioritize data quality, standards, and the use of common data models for interoperability. These practices enable the delivery of trustworthy results and a robust evidence base for translating research into improved patient care practices and products, including contributions to progress in regulatory science.

The IHI journey in the big data area is only at the beginning, and many challenges remain. For example, federated learning (FL) has shown promising results in projects such as EHDEN and MELLODDY, leveraging privacy-preserving technologies and adhering to standards, regulations, and ethical requirements. However, significant work remains to be done to fully realize FL's potential and expand its adoption. A focus is on optimizing the use of real-world data (RWD), including RWD generated by medical devices. A recent multistakeholder workshop at the IHI stressed that growing regulatory

⁹ Alzheimer's Disease Data Initiative ADDI <https://www.alzheimersdata.org/>

¹⁰ <https://www.aims-2-trials.eu/our-research/data-management-analysis/>

science, building on transparency, trust, and legitimacy, improving standardization, boosting data and digital literacy, and building on already existing infrastructures are key to boosting progress in RWD use in healthcare.¹¹ These are ideal areas for future IHI partnerships.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: this perspective describe multiple projects results from IHI projects details can be found via <https://www.ih.europa.eu/projects-results>.

Ethics statement

This is a perspective paper based on published results of IHI projects. The studies involving humans that are described under this perspective and in general all research of ethical concern carried out by projects funded by IHI is previously approved by relevant ethical committees and regulatory authorities as relevant. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from different sources in different projects cited. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements. The animals in this perspective refer to studies from IHI projects that were approved by relevant local committees. It does not include original research. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

EV: Conceptualization, Writing – original draft, Writing – review & editing.

¹¹ Report of the workshop will be available in 2025 on the IHI website.

References

- Damilola Oluwaseun Ogundipe. The impact of big data on healthcare product development: a theoretical and analytical review. *Int Med Sci Res J.* (2024) 4:341–60. doi: 10.51594/imsrj.v4i3.932
- Laverty H, Meulien P, Laverty H. The innovative medicines Initiative –10 years of public-private collaboration. *Front Med.* (2019) 6:1–13. doi: 10.3389/fmed.2019.00275
- Stevens H, Van Overwalle G, Van Looy B, Huys I. Innovative medicines initiative (IMI) case study analysis reveals the true added value of early-phase public-private partnerships (PPPs). *Biotechnol Law Rep.* (2015) 34:153–65. doi: 10.1089/blr.2015.29008.hs
- Innovative Health Initiative Joint Undertaking. Strategic Research and Innovation Agenda [Internet]. Publications Office of the European Union. (2022). Available at: https://www.ih.europa.eu/sites/default/files/uploads/Documents/About/IHI_SRIA_ApprovedJan22.pdf
- Innovative Health Initiative. Bibliometric analysis of ongoing projects 15th report - 2024. Brussels. (2024). Available online at: https://www.ih.europa.eu/sites/default/files/uploads/Documents/About/Reports/IHI%202024%20Bibliometrics%20report_Full_04.11.24%20.pdf
- Vaudano E, Georges J, Hofmann-Apitius M, Lo D. Editorial: impacts of public-private collaborative research on Alzheimer's disease: the case of the innovative medicines initiative. *Front Neurol.* (2023) 14:1–3. doi: 10.3389/fneur.2023.1321225
- Brito M, De F Torre C, Silva-Lima B. Scientific advances in diabetes: the impact of the innovative medicines Initiative. *Front Med.* (2021) 8:1–13. doi: 10.3389/fmed.2021.688438
- Teare HJA, De Masi F, Banasik K, Barnett A, Herrgard S, Jablonka B, et al. The governance structure for data access in the DIRECT consortium: an innovative medicines initiative (IMI) project. *Life Sci Soc Policy.* (2018) 14:1–17. doi: 10.1186/s40504-018-0083-0
- Bell J, Mourby M, Kaye J. Contractual mechanisms for securing the public interest in data sharing in public-private health research partnerships. (2023). Available online at: <https://www.imi.europa.eu/about-imi>
- Bradshaw A, Hughes N, Vallez-Garcia D, Chokoshvili D, Owens A, Hansen C, et al. Data sharing in neurodegenerative disease research: challenges and learnings from the innovative medicines initiative public-private partnership model. *Front Neurol.* (2023) 14:1–16. doi: 10.3389/fneur.2023.1187095

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

I would like to thank Niklas Blomberg and Nathalie Seigneuret for their valuable comments on this perspective piece.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Views and opinions expressed are those of the author only and do not necessarily reflect those of the Innovative Health Initiative Joint Undertaking, nor the European Union, the private members, and those contributing partners of the IHI JU. Neither of the aforementioned parties can be held responsible for them.

11. Sanz F, Pognan F, Steger-Hartmann T, Díaz C, Cases M, Pastor M, et al. Legacy data sharing to improve drug safety assessment: the eTOX project. *Nat Rev Drug Discov.* (2017) 16:811–2. doi: 10.1038/nrd.2017.177
12. Sanz F, Pognan F, Steger-Hartmann T, Díaz C, Asakura S, Amberg A, et al. eTRANSAFE: data science to empower translational safety assessment. *Nat Rev Drug Discov.* (2023) 22:605–6. doi: 10.1038/d41573-023-00099-5
13. Sarntivijai S, Blomberg N, Lauer KB, Briggs K, Steger-hartmann T, Lei J, et al. eTRANSAFE: building a sustainable framework to share the eTRANSAFE consortium. *F1000Research.* (2024) 11:1–12.
14. Steger-Hartmann T, Sanz F, Bringezu F, Soininen I. IHI VICT3R: developing and implementing virtual control groups to reduce animal use in toxicology research. *Toxicol Pathol.* (2024). doi: 10.1177/01926233241303906
15. Heyndrickx W, Mervin L, Morawietz T, Sturm N, Friedrich L, Zalewski A, et al. MELLODDY: cross-pharma federated learning at unprecedented scale unlocks benefits in QSAR without compromising proprietary information. *J Chem Inf Model.* (2024) 64:2331–44. doi: 10.1021/acs.jcim.3c00799
16. Lovestone S. The European medical information framework: a novel ecosystem for sharing healthcare data across Europe. *Learn Health Syst.* (2019) 4:e10214. doi: 10.1002/lrh2.10214
17. Szócska M, Jayawardana S, Smand C, Salimullah T, Reed C, Hanif S. Big data for better outcomes: supporting healthcare transformation in Europe. *Eurohealth.* (2017) 23:1–5.
18. Hughes N, Rijnbeek PR, Van Bochove K, Vizcaya D, Ryan P. Evaluating a novel approach to stimulate open science collaborations: a case series of “study-a-thon” events within the OHDSI and European IMI communities. *JAMIA Open.* (2022) 5:ooac100. doi: 10.1093/jamiaopen/ooac100
19. Li X, Raventós B, Roel E, Pistillo A, Martínez-hernandez E, Delmestri A, et al. Association between COVID-19 vaccination, SARS-CoV-2 infection, and risk of immune mediated neurological events: population based cohort and self-controlled case series analysis. *BMJ.* (2022) 376:e068373. doi: 10.1136/bmj-2021-068373
20. Gauffin O, Brand JS, Hedfors S, Daniele V, Suvi S, Marti A. Supporting pharmacovigilance signal validation and prioritization with analyses of routinely collected health data: lessons learned from an EHDE network study. *Drug Saf.* (2023) 46:1335–52. doi: 10.1007/s40264-023-01353-w
21. Moulin P, Grünberg K, Barale-Thomas E, der Laak JV. IMI—Bigpicture: a central repository for digital pathology. *Toxicol Pathol.* (2021) 49:711–3. doi: 10.1177/0192623321989644
22. EFPIA. DataSharingPlayBook. EFPIA. European Federation of Pharmaceutical Industries and Associations. (2024). Available online at: https://www.ihieuropea.eu/sites/default/files/uploads/Documents/ProjectResources/IMI_IHI_DataSharingPlayBook_2024.pdf
23. Knopp KL, Stenfors C, Baastrup C, Bannan AW, Calvo M, Caspani O, et al. Experimental design and reporting standards for improving the internal validity of pre-clinical studies in the field of pain: consensus of the IMI-Europain consortium. *Scand J Pain.* (2015) 7:58–70. doi: 10.1016/j.sjpain.2015.01.006
24. Vollert J, Macleod M, Dirnagl U, Kas MJ, Michel MC, Potschka H, et al. The EQUIP framework for rigor in the design, conduct, analysis and documentation of animal experiments. *Nat Methods.* (2022) 19:1334–7. doi: 10.1038/s41592-022-01615-y
25. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, et al. CODE-EHR best-practice framework for the use of structured electronic health-care records in clinical research. *Lancet Digit Health.* (2022) 4:e757–64. doi: 10.1016/S2589-7500(22)00151-0
26. Welter D, Juty N, Philippe R, Xu F, Henderson D, Gu W, et al. FAIR in action - a flexible framework to guide FAIRification. *Sci Data.* (2023) 10:1–9. doi: 10.1038/s41597-023-02167-2
27. Richardson JL, Moore A, Bromley RL, Stellfeld M, Geissbühler Y, Bluett-Duncan M, et al. Core data elements for pregnancy pharmacovigilance studies using primary source data collection methods: recommendations from the IMI ConcePTION project. *Drug Saf.* (2023) 46:479–91. doi: 10.1007/s40264-023-01291-7
28. Hübner M, Nyongui E, Gazzero E, Hassoun L, Zschüntzsch J, Krefting D, et al. Improving clinical documentation of rare neuromuscular diseases: development of a standardised information model. *Stud Health Technol Inform.* (2024) 316:1418–9. doi: 10.3233/SHTI240677
29. Gyrard A, Gribbon P, Hussein R, Abedian S, Bonmati LM, Cabornero GL, et al. Synergies among health data projects with cancer use cases based on health standards. *Stud Health Technol Inform.* (2024) 316:1292–6. doi: 10.3233/SHTI240649
30. Goldman M, Seigneuret N, Eichler HG. The innovative medicines Initiative: an engine for regulatory science. *Nat Rev Drug Discov.* (2014) 14:1–2. doi: 10.1038/nrd4520
31. Toga AW, Phatak M, Pappas I, Thompson S, McHugh CP, Clement MHS, et al. The pursuit of approaches to federate data to accelerate Alzheimer’s disease and related dementia research: GAAIN, DPUK, and ADDI. *Front Neuroinform.* (2023) 17:1175689. doi: 10.3389/fninf.2023.1175689
32. Saunders S, Gregory S, Clement MHS, Birk C, Geyten S, Der V, et al. The European prevention of Alzheimer’s dementia programme: an innovative medicines initiative-funded partnership to facilitate secondary prevention of Alzheimer’s disease dementia. *Front Neurol.* (2022) 13:1051543. doi: 10.3389/fneur.2022.1051543
33. Bose N, Brookes AJ, Scordis P, Visser PJ. Data and sample sharing as an enabler for large-scale biomarker research and development: the EPND perspective. *Front Neurol.* (2022) 13:13. doi: 10.3389/fneur.2022.1031091
34. Oakley B, Lautarescu A, Charman T, Chatham C, Loth E, Beckmann C, et al. Data sharing in child and adolescent psychiatry research: key challenges (and some potential solutions). *Open Europe.* (2025)
35. Charman T, Loth E, Tillmann J, Crawley D, Wooldridge C, Goyard D, et al. The EU-AIMS longitudinal European autism project (LEAP): clinical characterisation. *Mol Autism.* (2017) 8:1–21. doi: 10.1186/s13229-017-0145-9
36. Bloomfield M, Lautarescu A, Heraty S, Douglas S, Violland P, Plas R, et al. European autism GENomics registry (EAGER): protocol for a multicentre cohort study and registry. *BMJ Open.* (2024) 14:e080746. doi: 10.1136/bmjopen-2023-080746
37. D’agostino M, Cairns DA, Lahuerta JJ, Wester R, Bertsch U, Waaga A, et al. Second revision of the international staging system (R2-ISS) for overall survival in multiple myeloma: a European myeloma network (EMN) report within the HARMONY project. *J Clin Oncol.* (2022) 3406–20. doi: 10.1200/JCO.21.02614
38. Omar MI, Roobol MJ, Ribal MJ, Abbott T, Agapow PM, Araujo S, et al. Introducing PIONEER: a project to harness big data in prostate cancer research. *Nat Rev Urol.* (2020) 17:351–62. doi: 10.1038/s41585-020-0324-x
39. Lawlor A, Beyer K, Russell B, Steinbeisser C, Bjartell A, De Meulder B, et al. PIONEER big data platform for prostate cancer: lessons for advancing future real-world evidence research. *Nat Rev Urol.* (2024) 22:116–24. doi: 10.1038/s41585-024-00925-4
40. Gandaglia G, Pellegrino F, Golozar A, De Meulder B, Abbott T, Achtman A, et al. Clinical characterization of patients diagnosed with prostate cancer and undergoing conservative management: a PIONEER analysis based on big data. *Eur Urol.* (2024) 85:457–65. doi: 10.1016/j.eururo.2023.06.012
41. N’Dow J, Smith EJ, Polychronopoulos K, Cannon A, Roobol M, Auweter S, et al. 917P OPTIMA: improve care for patients with prostate, breast, and lung cancer through artificial intelligence. *Ann Oncol.* (2022) 33:S966. doi: 10.1016/j.annonc.2022.07.1042
42. Hussein R, Balaur I, Burmann A, Ćwiek-Kupczyńska H, Gadiya Y, Ghosh S, et al. Getting ready for the European health data space (EHDS): IDERHA’s plan to align with the latest EHDS requirements for the secondary use of health data. *Open Res Europe.* (2024) 4:1–23. doi: 10.12688/openreseurope.18179.1
43. Georges J, Diaz-Ponce A, Lamirel D, Moradi-Bachiller S, Gove D. Keeping track of and recognizing the value of public involvement work in dementia research. *Front Neurol.* (2022) 13:13. doi: 10.3389/fneur.2022.1031831
44. Lautarescu A, Trost B, Kushki A, Oakley B, Heraty S, Belton D, et al. Global autism data sharing: benefits, challenges, and recommendations. *PLoS Digital Health.* (2025)
45. Burn E, Weaver J, Morales D, Prats-Urbe A, Delmestri A, Strauss VY, et al. Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study. *Lancet Rheumatol.* (2019) 1:e229–36. doi: 10.1016/S2665-9913(19)30075-X
46. Suter-Crazzolaro C. Better patient outcomes through mining of biomedical big data. *Front ICT.* (2018) 5:30. doi: 10.3389/fict.2018.00030
47. European Commission. Interim evaluation of the innovative health initiative (IHI) and final evaluation of the innovative medicines initiative (IMI2) evaluation support study on horizon Europe’s contribution to resilient Europe. (2024). Available online at: <https://op.europa.eu/en/publication-detail/-/publication/eb6f1086-52d9-11ef-acbc-01aa75ed71a1/language-en>