Topic ida submitted to IHI - Reference Number: TI_001220

Are you submitting the idea:
☐ **in your personal capacity?**
☒ **on behalf of an organisation?**

Please indicate the name of the group organisation: Pistoia Alliance

Please select from the list below the type of stakeholders your organization represents:
Charity/foundation

1 Title of your idea

Please provide a short title that accurately reflects the objective(s) of your idea:
In Vivo Data Redefined: Empowering Drug Discovery with Advanced Insights Through the Right Data Frameworks and Standards

2 Scope

Explain the specific challenges/problems to be addressed by your idea and how these affect relevant stakeholders, taking into account what is already known and/or available in the field:
In the realm of biopharmaceutical research and development, nothing is as critical as the quality, reproducibility, and exploitation of in-vivo data. Robust and reliable data form the bedrock of scientific rigour, enabling the accelerated pace of knowledge acquisition and the timely development of innovative drugs that benefit patients. The advent of advanced artificial intelligence (AI) tools and large language models presents an unprecedented opportunity to exploit vast datasets for drug development. However, to harness this potential, it is essential to establish comprehensive data frameworks and standards that ensure high-quality FAIR-enabled data production across multiple industry sources, resulting in clear data products that are human and AI actionable.
Our proposed topic seeks to establish best-in-class data production ready standards and frameworks tailored for pre-clinical in-vivo research, enabling seamless interoperability, translatability and integration of diverse datasets. Proven interoperability will facilitate the exploitation of these datasets by modern AI and machine learning tools, significantly reducing the time and resources required for manual data curation. By addressing the limitations of current practice, and leveraging AI capabilities, we envision a future where in-vivo data drives drug development through advanced exploitation by AI and large language models.
Furthermore, once established these data sets, data standards and data frameworks will power new research and analytics that will enable reduction and replacement of animal studies through new approach methodologies (NAMS). These will efficiently leverage legacy data for regulatory decision making, translational research, testing and validation of digital twin approaches, and multiple impacts in drug discovery and allied industries.

Please indicate which IHI specific objective(s) (SO), as described in the IHI Strategic Research and Innovation Agenda (SRIA), your idea addresses:

["SO2: integrate fragmented health research and innovation efforts bringing together health industry sectors and other stakeholders, focusing on unmet public health needs, to enable the development of tools, data, platforms, technologies and processes for improved prediction, prevention, interception, diagnosis, treat- ment and management of diseases, meeting the needs of end-users"
"SO5: enable the development of new and improved evaluation methodologies and models for a comprehensive assessment of the added value of innovative and integrated health care solutions"]

Please select the keywords that are most relevant to your idea:

["Cardiovascular diseases"
"Immune system diseases"
"Metabolic diseases"
"Neurodegenerative diseases"
"Oncology"
"Rare diseases"
"Prediction"
"Prevention"
"Digital health"
"Health technology"]

In alignment with the IHI specific objective(s) selected above, specify the objectives of your idea:

AI is data hungry and requires extensive, freely available datasets to train models. These datasets adhere to standards that represent the current state of scientific understanding and enable the integration of multiple data sources providing the large sets of harmonised data that AI needs to succeed. Currently, the diverse and conflicting standards for in vivo data are blocking AI approaches in this critical domain of drug discovery and healthcare. To resolve this we need to :
build the scientific community collaborative ecosystem of stakeholders, reduction of animal research due to shared data and knowledge
expand data frameworks and standards for in vivo data that comprehensively span the early research to translational research in collaboration with data standards bodies
utilise the data frameworks and standards to provision high quality FAIR machine readable & actionable in vivo data from partners across the industries to power AI and machine learning model development
ensure a sustainable future of this important new cross-sector data standards and framework for extension to other data types and applications

3 Expected impacts to be achieved by your idea

Briefly describe the expected impacts to be achieved by your idea, ensuring that they contribute to IHI general and relevant specific objectives, as described in the IHI SRIA:

__Impacts__ are wider long-term effects on society (including the environment), the economy and science, enabled by the outcomes of R&I investments. Impacts generally occur sometime after the end of the project, e.g. successful implementation of digital solutions supporting people-centred care.

***IHI general objectives:** 1. contribute towards the creation of an EU-wide health research and innovation ecosystem that facilitates translation of scientific knowledge into innovations, notably by launching at least 30 large-scale, cross-sectoral projects, focussing on health innovations; 2. foster the development of safe, effective, people-centred and cost-effective innovations that respond to strategic unmet public health needs, by exhibiting, in at least 5 examples, the feasibility of integrating health care products or services, with demonstrated suitability for uptake by health care systems. The related projects should address the prevention, diagnosis, treatment and/or management of diseases affecting the EU population, including contribution to 'Europe's Beating Cancer Plan'; 3. drive cross-sectoral health innovation for a globally competitive European health industry and contribute to reaching the objectives of the new Industrial Strategy for Europe and the Pharmaceutical Strategy for Europe.*

This idea would create a FAIR data framework and data standards in an area of drug development where these do not currently exist. It would remove a significant obstacle to seamless end to end drug discovery, enabling faster, simpler and cheaper development of drugs. Interactions between researchers, companies, CROs and stakeholders across the drug development process would be simplified.
Critically, once established these data frameworks and standards would power AI and machine learning in a vital part of drug discovery and development where this is not currently possible. The development of new AI modelling approaches with benchmarking and validation will simplify translational research, streamline regulatory interaction and leverage control data for NAMS and other activities.
The impact of such a project will drive cross-sectoral innovation in the drug discovery and development ecosystem, contribute to the objectives of the Industrial and Pharmaceutical Strategies for Europe, and form a key role in driving digital transformation.

4 Why should your idea become an IHI call topic?

Explain why collaboration through a cross-sectoral and multidisciplinary public private partnership is needed in particular:

Why does it require collaboration among several industry sectors (*e.g.* pharma, vaccines, biotech, medical devices, in vitro diagnostics, radiotherapy, medical imaging health ICT)?

Why does it require collaboration between private (industry) and public partners (*e.g.* academia, healthcare practitioners, patients, regulators)?

The data frameworks and standards for in vivo data requires collaboration across many of the industry sectors that participate in IHI as well as the scientific community as a whole, academia and regulators. To fully realise the promise of AI and machine learning, large amounts of high quality data is required. But to do this successfully, there is a need to build the FAIR data frameworks and standards that would structure the data such that AI models could be developed. As with any AI activity, the validation of models and benchmarks is a key scientific community activity.
Beyond these immediate needs, the interaction of researchers with regulatory authorities in biotech, pharma and medical devices requires the submission of data in standard formats. The development of data frameworks and standards would simplify these interactions and over time become a key part of new product development.

Why is the contribution of industry needed to achieve the expected impacts?

***Contribution of industry**: Large companies that are members of the IHI industry partners (i.e. COCIR, EFPIA, EuropaBio, MedTech Europe, Vaccines Europe) contribute to the programme, primarily through 'in-kind' contributions (e.g. their researchers' time, laboratories, data, compounds). At least 45% of each project's total costs have to be in-kind contribution.*

The development of data standards and frameworks is an activity that by necessity needs the whole scientific community to become involved. In terms of in vivo data for AI, the pharma and biotech world has a particular interest and can contribute knowledge and data accordingly. Wider engagement with CROs, academics, SMEs and standards bodies is essential, and potentially also regulatory representatives in the longer term. This work could also impact vaccines and medical device research with the development of specific standards supporting those activities too. This would be a large and significant effort but could change the face of drug development shortening timeframes and reducing costs across the industry. The Pistoia Alliance members (large pharma, biotech, SMEs and CROs) are already working together to develop this idea.