

All information regarding future IHI Call topics is indicative and subject to change. Final information about future IHI Calls will be communicated after approval by the IHI Governing Board.

Topic 4: Maximising the potential of synthetic data generation in healthcare applications

Expected impacts to be achieved by this topic

To exploit the full potential of digitalisation and data exchange in health care, this topic is expected to contribute to the following expected impacts:

- wider availability of interoperable, synthetic data generation methodologies and/or datasets facilitating research and development of integrated products and services that will benefit patients;
- improved insight into real-life behaviour and challenges of patients with complex, chronic diseases and co-morbidities thanks to m-health and e-health technologies;
- advanced analytics / artificial intelligence tools supporting health research and innovation resulting in: a) better clinical decision support for increased accuracy of diagnosis and efficacy of treatment; b) faster prototyping and shorter times-to-market of personalised health interventions; and c) better evidence of the added value from new digital health and AI tools, including reduced risk of bias due to improved methodologies.

Expected outcomes

The proposals should contribute to all of the following expected outcomes:

- academic and industrial researchers should have access to relevant, robust, and generalisable synthetic data generation methodologies, including open source when relevant, to create and share pools of synthetic patient data in specific use cases;
- academic and industrial researchers should have access to relevant, high quality synthetic datasets;
- thanks to better availability of robust synthetic datasets for training data models, healthcare providers and industry should have a wider range of performant AI-based and other data-driven tools to support diagnostics, personalised treatment decision-making and prediction of health outcomes.

Scope

Healthcare research using individual patient data is often constrained due to restrictions in data access because of privacy, security, intellectual property (IP) and other concerns. Synthetic health data, i.e., data that is artificially created to mimic individual patient data, can reduce these concerns, leading to more rapid development of reliable data-driven methods including diagnostic, precision medicine, decision support and patient monitoring tools. However, while many synthetic data generation (SDG) methods are currently available, it is not always clear which method is best for which use case, and SDG methods for some types of data are still immature. Furthermore, it is still unclear whether highly detailed synthetic data, which are often needed for research, can be categorised as anonymous.

To address these challenges and maximise the opportunity offered by synthetic data, projects funded under this topic should address the following objectives:

- assemble a cross-sectoral public-private consortium including synthetic data experts, public and private data owners, and healthcare solution developers;
- using high-quality public and private datasets, develop / further develop and validate reliable SDG methods for relevant healthcare use cases. The use cases to be explored must be described and justified in the proposal, complement work that is already ongoing, and should:
 - ensure the broad applicability of the SDG methods developed and include data types that are not currently adequately addressed, such as device data, image data, genomic data etc;
 - include methods to generate: a) fully synthetic datasets that do not contain any real data; b) hybrid datasets composed of a combination of data derived from both real and synthetic data; and c) synthetically-augmented datasets.
 - pay particular attention to bias, both in source data and in the SDG methods.
- validate the synthetic data generation methods applied in the project using source data. This should include assessing the risk of re-identification;
- demonstrate the quality and applicability of the synthetic data generated in the project through the development of relevant models;
- encourage the uptake of the results of the project through a strong communication and outreach plan.

Applicants are expected to consider allocating appropriate resources to explore synergies with other relevant initiatives and projects, including the EC proposal for an European Health Data Space (EHDS)¹ when it becomes operational.

Why the expected outcomes can only be achieved by an IHI project

Development and validation of synthetic data generation methods and tools for data-driven applications requires multidisciplinary collaboration across private and public entities, including public and private data owners, healthcare solution developers, and synthetic data experts.

Indicative budget

Applicant consortia will be competing for the maximum financial contribution from IHI up to EUR 20 000 000.

IHI estimates that an IHI financial contribution of around EUR 10 000 000 would allow a proposal to address these outcomes appropriately. Nonetheless, this does not preclude submission and selection of a proposal requesting different amounts.

Applicant consortia should ensure that at least 45% of the action's eligible costs and costs for action-related additional activities are provided by in-kind contributions to operational activities (IKOP), financial contributions (FC), or in-kind contributions to additional activities (IKAA) from private members and/or contributing partners and the constituent or affiliated entities of the private members and/or of the contributing partners. Contributing partners may not contribute IKAA. See call conditions for further information.

¹ https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

Indicative duration of the actions

Applicants should propose a project duration that matches project activities and expected outcomes and impacts.

Dissemination and exploitation obligations

The specific obligations described in the conditions of the calls and call management rules under 'Specific conditions on availability, accessibility and affordability' do not apply.

INDICATIVE TEXT