synthia

Synthetic Data Generation framework for integrated validation of use cases and AI healthcare applications

Navigating the Policy-Privacy Tension for Synthetic Data in Personalised Medicine (EU and Member States)

SYNTHIA is a multi-stakeholder Public-Private Partnership focused on developing practical guidance for the responsible generation and use of synthetic data in health and personalised medicine. The consortium aims to bridge legal, technical, and governance perspectives to support privacy-preserving data use that is demonstrably fit for purpose and trustworthy.

This paper launches the SYNTHIA Policy Engagement Materials Series, a structured set of outputs to support evidence-based policymaking on synthetic data in health and personalised medicine.
The series is designed to translate technical, legal and governance developments into practical materials for policymakers, regulators, health systems, innovators and other stakeholders. It will combine strategic analysis, trust frameworks, governance models, comparative mapping and targeted recommendations at EU and Member State level.

Policy 01 / Strategic Framework

Why this matters now: Europe needs practical governance pathways

European initiatives are simultaneously accelerating AI-enabled innovation in health and biotechnology and reinforcing stringent protection of personal data and fundamental rights. Personalised medicine sits at the centre of these objectives, resulting in tension: progress depends on access to high-quality, representative data, yet lawful use and public trust depend on demonstrable privacy safeguards.

Synthetic data is increasingly positioned as a supplementary bridge—enabling model development, testing and collaboration while reducing reliance on direct sharing of patient-level records for many early-stage, collaborative and external-validation tasks.

At EU level, competitiveness and simplification initiatives (including the Digital Omnibus and the proposed European Biotech Act) explicitly seek to reduce administrative burden while preserving high standards for rights protection. At Member State level, divergent interpretations and enforcement practices around anonymisation, pseudonymisation, and re-identification risk still generate uncertainty and uneven innovation capacity.

The policy–privacy tension in practice

The friction is operational, not theoretical. Stakeholders (patients, hospitals, researchers, industry and regulators) face recurring questions when wanting to use and share their datasets:

- When is a dataset "anonymous enough" to fall outside the GDPR, and for whom?
- How should organisations evidence that re-identification risk is "insignificant" given evolving auxiliary data and attack methods?
- How can cross-border projects be designed so that a synthetic dataset treated as anonymous in one jurisdiction is not treated as personal data in another?
- How can compliance efforts be sustainably reduced without weakening trust?

Synthetic data does not remove these questions; it changes them. Privacy risk shifts from direct disclosure of patient records to:

- Leakage/overfitting from the generative process;
- Membership/attribute/model inference;
- Linkage with auxiliary datasets;
- Governance failures (e.g., access to re-identification keys, weak audit, unclear accountability, misplaced confidence in data anonymity).

EU-level landscape: convergence, but not yet a single playbook

Synthetic data in health are shaped by multiple EU policy and legal regimes that increasingly converge in objectives, but still do not provide a single, coherent playbook for practice.
The table below provides an overview of the main EU regulatory regimes that shape the use of synthetic data in health. The sections below examine each of these regimes in more detail.

Regulation	Focus	Implications for synthetic data
GDPR	Identifiability, re-identification risk	Context- and actor-dependent classification
Data Governance Act	Data reuse and safeguards	Encourages privacy-preserving techniques
AI Act	AI risk classification, governance, accountability, and data quality (Art. 10)	Direct requirements on training/validation datasets (quality, representativeness, bias control), including when synth data are used
European Health Data Space (EHDS)	Secondary use of health data, access governance	Creates structured access pathways; synth data may support privacy-preserving re-use and cross-border research
Proposed European Biotech Act	Innovation, data quality, trials	Rising demand for trustworthy datasets

Unpacking the regulatory landscape: four developments to watch

While the table on the previous page maps the regulatory landscape at a glance, four developments deserve closer attention because they directly shape how synthetic data can be classified, governed and deployed in practice. Together, they illustrate both the direction of travel, toward more contextual, risk-based approaches, and the gaps that still need to be resolved before organisations can act with legal certainty.

1.

Identifiability: actor- and context-dependent

Within the data-protection framework, recent EU case-law reinforces that "identifiability" must be assessed in context and in relation to the actor processing the data. identification is considered not "reasonably likely" where identification is prohibited by law or practically impossible because it would require a disproportionate effort (time, cost and labour). The Court has also clarified that pseudonymisation may, depending on circumstances, prevent persons other than the controller from identifying the data subject, while remaining personal data for the controller (or any party with access to a re-identification key or equivalent means).

For synthetic data, the practical implication is that classification (personal vs. effectively anonymous) is not only a question of dataset properties; it is also a question of who receives it, what auxiliary data they have, and what controls prevent identification.

2.

Privacy-preserving techniques are an explicit policy objective

The Data Governance Act encourages the development and use of privacy-enhancing technologies to enable data re-use, such as differential privacy and the generation of synthetic data. This provides a strong EU-level anchor: legislators explicitly recognise that technical measures, combined with governance, can unlock the use of socially valuable data while limiting privacy risks. While this provides a strong policy signal, it does not yet resolve how such techniques should be evaluated, governed, or accepted as sufficient safeguards in high-stakes health contexts.

3.

AI Act implementation in health and life sciences

The AI Act's practical impact depends heavily on data governance, transparency, and lifecycle monitoring; collated in technical and non-technical documentation. The Commission has announced a programme of implementation guidelines to support consistent application of the AI Act, including guidance on risk classification, transparency, post-market monitoring, and incident reporting. Additional guidance on obligations for providers and deployers, responsibilities along the AI value chain, and simplified quality management system elements for SMEs and small mid-caps is also expected. Clarification is further anticipated on the AI Act's interaction with other EU legislation, including data-protection law, notably through joint guidance with the European Data Protection Board.

4.

Biotech and clinical trials policy: demand for trustworthy data is increasing

The proposed European Biotech Act aims to accelerate time-to-market while safeguarding high standards for health, ethics, quality and biosecurity. It places emphasis on facilitating AI in health biotechnology and on improving datasets, provenance, interoperability and annotation through mechanisms such as a "Biotechnology Data Quality Accelerator," designed to operate in compliance with key regulations including the GDPR and the European Health Data Space. It also proposes regulatory sandboxes and strengthened approaches to clinical trial innovation, while requiring safeguards for participant rights and data robustness. These initiatives will increase demand for practical, scalable privacy-preserving data strategies, precisely the niche where governed synthetic data can be positioned.



Member State realities: where the policy-privacy tension becomes operational

Even with EU-level principles, implementation is mediated through national systems and practices:

- Divergent standards for anonymisation and re-identification risk assessment across DPAs, ethics committees and health data access bodies;
- Variable maturity of trusted research environments, secure processing facilities and independent validation capacity;
- Inconsistent contracting, IP practices and procurement requirements that shape what synthetic data can be shared and under what conditions;
- Differences in clinical data management (consent models, secondary use permissions, oversight expectations), complicating multinational personalised medicine studies.

The result is a practical barrier: organisations either design to the most restrictive interpretation (reducing utility and slowing innovation) or proceed under uncertainty (increasing legal, ethical and reputational risk).

Strategic Priorities for SYNTHIA and Policymakers: a risk-based, evidence-driven synthetic data playbook

In response to this governance gap, SYNTHIA seeks to articulate a practical, risk-based playbook for the responsible use of synthetic data in personalised medicine.

The recommendations below distinguish between actions that can be taken at EU level and those that require coordination and implementation at Member State level.

EU-level recommendations

1.

Clarify classification and evidence requirements Use the Commission's AI Act guidance programme (and the planned AI Act–data protection interplay guidance) to develop practical criteria for

- when synthetic data is likely to remain personal data vs. be effectively anonymous for a given context;minimum evidence packages to support "reasonable likelihood" assessments (threat modelling with auxiliary data assumptions, linkage analysis, and leakage tests);
- reassessment triggers as technology and auxiliary datasets evolve.

2.

Standardise artefacts for regulated and near-regulated health use cases.

Develop standardized template sets usable by SMEs and public hospitals, aligned with the AI Act's documentation ecosystem, such as

- a privacy and confidentiality evaluation plan (attack simulation and risk scoring);
- representativeness and bias assessment (who is under-represented and why);
- a provenance and "synthetic dataset card" (generation method, constraints, intended use, known limitations);
- a decision record explaining why synthetic data is proportionate for the stated purpose (and when real-data access remains necessary).

3.

Embed synthetic data into EU support infrastructure.

Link guidance explicitly to concrete support instruments (AI Act Service Desk, sandboxes, trusted environments under EHDS, and biotech data quality initiatives) so that compliance is paired with implementation capacity.

Member State recommendations

1.

Create interoperable national approval routes for synthetic data.

Define a standard pathway for approving and using synthetic data for research and innovation, with

- clear interpretation guidelines on when synthetic data is treated as personal or effectively anonymous in specific contexts, and which authorities should be involved accordingly, or not, depending on residual risk (e.g. health data access bodies, data protection authorities, ethics committees);
- default contractual clauses (prohibition of re-identification, audit rights, incident reporting, penalties);
- and secure processing and access controls proportional to residual risk.

2.

Invest in capability, not only compliance. Fund and operationalise

- trusted research environments that support privacy-preserving generation and testing;
- independent validation capacity (including accredited evaluators) for synthetic datasets used in regulated contexts;
- training for regulators and review bodies to evaluate technical evidence consistently.

Conclusion

From principle to practice: the path forward for synthetic data in Europe

Europe is moving toward a more operational, actor-specific understanding of identifiability, while pursuing simplification that preserves fundamental rights. This shift creates a genuine opening: the tools and legal reasoning needed to treat synthetic data as a credible, privacy-preserving alternative to direct data sharing already exist at EU level. What is missing is not ambition, but consistency, in how identifiability is assessed, how privacy-preserving techniques are evaluated, and how Member States translate shared principles into comparable practice.

Synthetic data cannot resolve this gap on its own, and it does not remove the underlying tension between innovation and privacy — it reshapes it, introducing new risks (from model leakage to governance failures) that require their own safeguards. Treated as a governed capability, however, rather than a compliance shortcut, synthetic data can help reconcile innovation imperatives with privacy expectations: evaluated rigorously, documented transparently, and embedded into both EU guidance and Member State implementation pathways.

This is precisely the space SYNTHIA is designed to occupy. By turning legal principles into consistent, evidence-driven practice, SYNTHIA aims to provide the practical bridge between regulatory intent and operational reality, one that withstands regulatory scrutiny, supports Member State implementation, and earns the trust of patients, clinicians and researchers alike. Through this broader series of policy engagement materials, SYNTHIA will translate emerging synthetic data debates into practical governance capacity for Europe, offering policymakers and practitioners a concrete starting points rather than another restatement of the problem.

The SYNTHIA Policy Engagement Materials Series will consist of:

- **Policy Paper "Tackling Public Trust for Synthetic Data in Health** - A paper exploring the conditions under which synthetic data can earn public trust in healthcare. It will cover governance, assurance, regulation, skills, citizen engagement and practical policy actions for trustworthy deployment.
- **White Paper "Risk-Based Governance Pathways for Synthetic Data in Public Health Decision-Making" -** This paper develops a proportionate governance model for synthetic data used in public health policy contexts. Classifies use cases by decision risk and sets corresponding expectations for validation, accountability and oversight across EU, national and regional levels.
- **European Landscape Paper "From Fragmentation to Functionality: Health Data Access Pathways for Synthetic Data in Europe" -** This paper will map how current European health data access models enable or constrain synthetic data generation and validation. It will combine system-level bottlenecks, good practice examples, and SYTNHIA disease-area specific pathways to support practical implementation under EHDS.
- **Policy Insight Paper "Synthetic Control Arms as a Cross-Cutting Validation Case for Synthetic Data in Europe" -** This paper assesses how synthetic data could contribute to comparator evidence in clinical and HTA decision-making. Uses synthetic control arms to surface broader policy questions around validation standards, proportional use, transparency and trust.
- **Comparative Policy Mapping Paper "Member State SWOT Analysis for Synthetic Data in Europe" -** This paper compares selected Member States through a SWOT lens, examining governance readiness, access pathways, infrastructure maturity and future opportunities for synthetic data under EHDS.

This publication was prepared by:

Authors

- Denis Horgan, EAPM
- Sofia Pazzagli, EAPM

Contributors

- Serena Elizabeth Marshall, DNV
- Davide Cirillo, BSC
- Eleonora Iascone, Humanitas
- Bogdan Kulynych, CHUV
- Stephen Duffield, NICE
- Christophe Maes, i-HD
- Matthias Müllenborn, Novo Nordisk
- Boris Kandov, UNIVIE
- Olga Startseva, UNIVIE
- Ellen de Waal, MATICAL
- Rea Tambic, Science Affairs

Direct link to the SYNTHIA Policy Engagement Materials:
www.ihi-synthia.eu/stakeholders/policy-makers

September 2026

www.ihi-synthia.eu

IHI SYNTHIA

contact@ihi-synthia.eu

IHI-SYNTHIA

@IHI SYNTHIA

ihi_synthia

@ihi-synthia.bsky.social

SYNTHIA, Synthetic Data Generation framework for integrated validation of use cases and AI healthcare applications, is supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 101172872. The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bio, MedTech Europe, and Vaccines Europe and DNV. The UK consortium partner, The National Institute for Health and Care Excellence (NICE) is supported by UKRI Grant 10132181.