SYNTHEMA + ERN-EuroBloodNet

Joint Training Programme on Synthetic Data Generation in SCD and AML

Funded by the European Union

Introduction to SYNTHEMA & synthetic data in healthcare

Nathan Lea, The European Institute for Innovation through Health Data (i~HD)

Session 2: Regulatory on Medical Devices, AI Act, Privacy-Preserving and Ethical Considerations

#1

Overview for this Session

Understanding the Evolving Regulatory Context

Challenges for Synthetic Data Development and Generation

How is SYNTHEMA Addressing Evolving Compliance Needs?

Understanding the Evolving Regulatory Context

The Situation in 2022

Regulatory bearing on SYNTHEMA

- Protecting Data – GDPR
- Research Governance
  - Independent ethical oversight
  - Participant (data subject) protections, including consent and transparency
- AI Trustworthiness
  - Expert working group in AI recommendations – Assessment List for Trustworthy AI
- Medical Device and In-Vitro Diagnostic development and production were not foreseen…
- Rare disease cohort needs and expectations were, however

The Situation Today

Understanding the Evolving Regulatory Context - 2026

Protecting Data – GDPR

Research Governance

- Independent ethical oversight
- Participant (data subject) protections, including consent and transparency

AI Trustworthiness

- Expert working group in AI recommendations – Assessment List for Trustworthy AI

NEW!! Enactment of the AI Act

- Link to Medical Device and In-Vitro Diagnostic Regulations

NEW!! Enactment of the European Health Data Space Regulations

NEW(ish)!! Data and Data Governance Acts

EVEN NEWER!! Digital Omnibus Proposals

Understanding the Evolving Regulatory Context

- More established regulations need to be viewed and interoperate with the newer regulations
- This is shifting continuously – AI Act and EHDS are gradually being implemented
- Interpretations will be updated – just as we saw with GDPR
- Regulatory compliance is just part of the story – trustworthiness is key
- Understanding the precise requirements of the new legislation is not fully clear yet

Challenges for Synthetic Data Development and Generation

Known Key Challenges

New Challenges

NEW!! Use of Synthetic Data to train Software as Medical or In-Vitro Diagnostic Devices will need to be quality assured

Does this make Synthetic Data Generation a High Risk AI System…?

NEW(ish)!! Greater emphasis on bias detection and source data quality

Particularly challenging in rare disease cohrots

NEW!! Greater emphasis placed on explainability and transparency around function and output, logging and decision making

Adds more pressure on training record keeping and results from model development

Scoping our discussion today

DGA and DA focus on not just personal data, but (mostly) anonymous data

- Data related to machinery, goods, services etc. and not just individuals
- Digital Omnibus is looking to merge DA and DGA (in broad terms)

EHDS has a deployment schedule over the next 5 – 6 years

- Will mostly impact health data holders, care providers etc.
- Has provisions around primary and secondary uses of health data, storage and access management
- Outlines the operational playing field for data handling – implications for opportunities around synthetic data use

The AI Act and ALTAI

AI Act – Regulation

Outlines the risk management approach for developing AI driven systems

Indicates certain kinds of AI are "High Risk" - including those that are medical devices or components of medical devices (as per Medical Device and In-Vitro Diagnostic Regulations)

Hi Risk AI Systems have stricter requirements

ALTAI – Self Assessment

EC Commissioned
First published in 2019, provided a framework for Trustworthy AI
7 Criteria Sets for a self assessment:

- Human Agency and Oversight
- Technical Robustness and Safety
- Privacy and Data Governance
- Transparency
- Diversity, Non-discrimination and Fairness
- Societal and Environmental Well-being
- Accountability

High Risk AI Systems

Adequate risk assessment and mitigation systems;	High quality of the datasets feeding the system to minimise risks and discriminatory outcomes;	Logging of activity to ensure traceability of results;	Detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance;
Clear and adequate information to the user;	Appropriate human oversight measures to minimise risk;	High level of robustness, security and accuracy.	Independent review by a notification authority, competent authority oversight (for bringing to market)

N.B. Interplay with MDR / IVDR

- It is still not clear how Medical Devices and IVDs powered by AI Systems will be assessed.
- Recommendations and the Digital Omnibus proposals are pointing to alignment with the MDR and IVDR – where the AI Act requirements and certifications will be joined with that regulatory pathway
- These are important points for bringing Synthetic Data Generation tools to the market
  - (or the devices they might help develop and power)
- Even if the AI System is not brought to the market (or is lower risk), regulation expects that you aim for high risk criteria

High Risk AI System Governance and Notification - Conformity Assessments, Certifications, Declarations and Notifications

Each Member State will identify its own Notification Body

High Risk systems will apply to the bodies for certifications

They will provide evidence of assessments and declarations

This relies on Quality Management Systems and internal assessments

An AI Board will be established to oversee implementation of the regulation

Competent Authorities will oversee this at Member State Level

How is SYNTHEMA Addressing the Evolving Compliance Needs?

SYNTHEMA Regulatory Compliance Foundations

Approach steeped in GDPR's Data Protection by Design and Default Approach

This involved putting data protection requirements front and centre from day 0 of the project.

Looped in Independent Ethics Committee Approvals and high standards of ethical practice as regulatory requirements that had to be met

Developed a Data Protection Impact Assessment to cover the project

Used this to inform partner obligations, develop a common view of our data handing and processing as well as Joint Controller Agreement specification - **"Measure twice, cut once" is true of contract execution!**

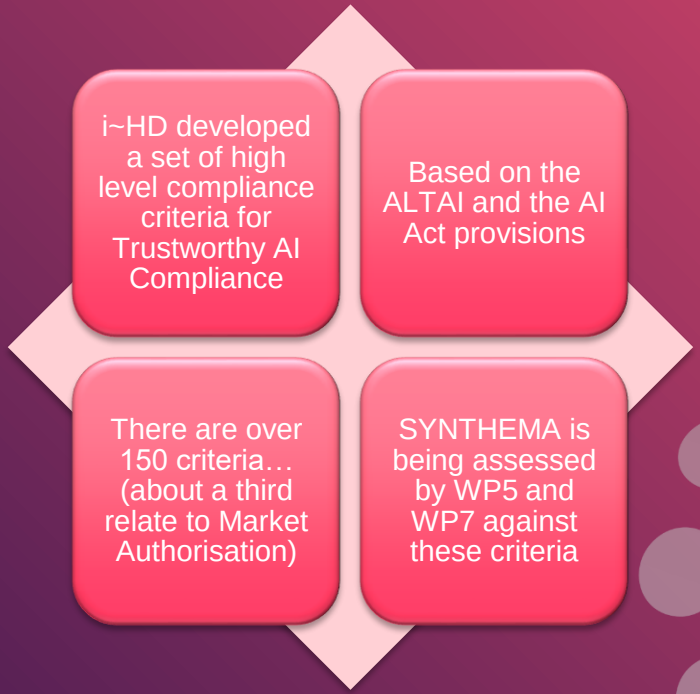
This is routinely updated as the project continues – and the regulations change!

- We are looking at how we explore compliance for the future post project

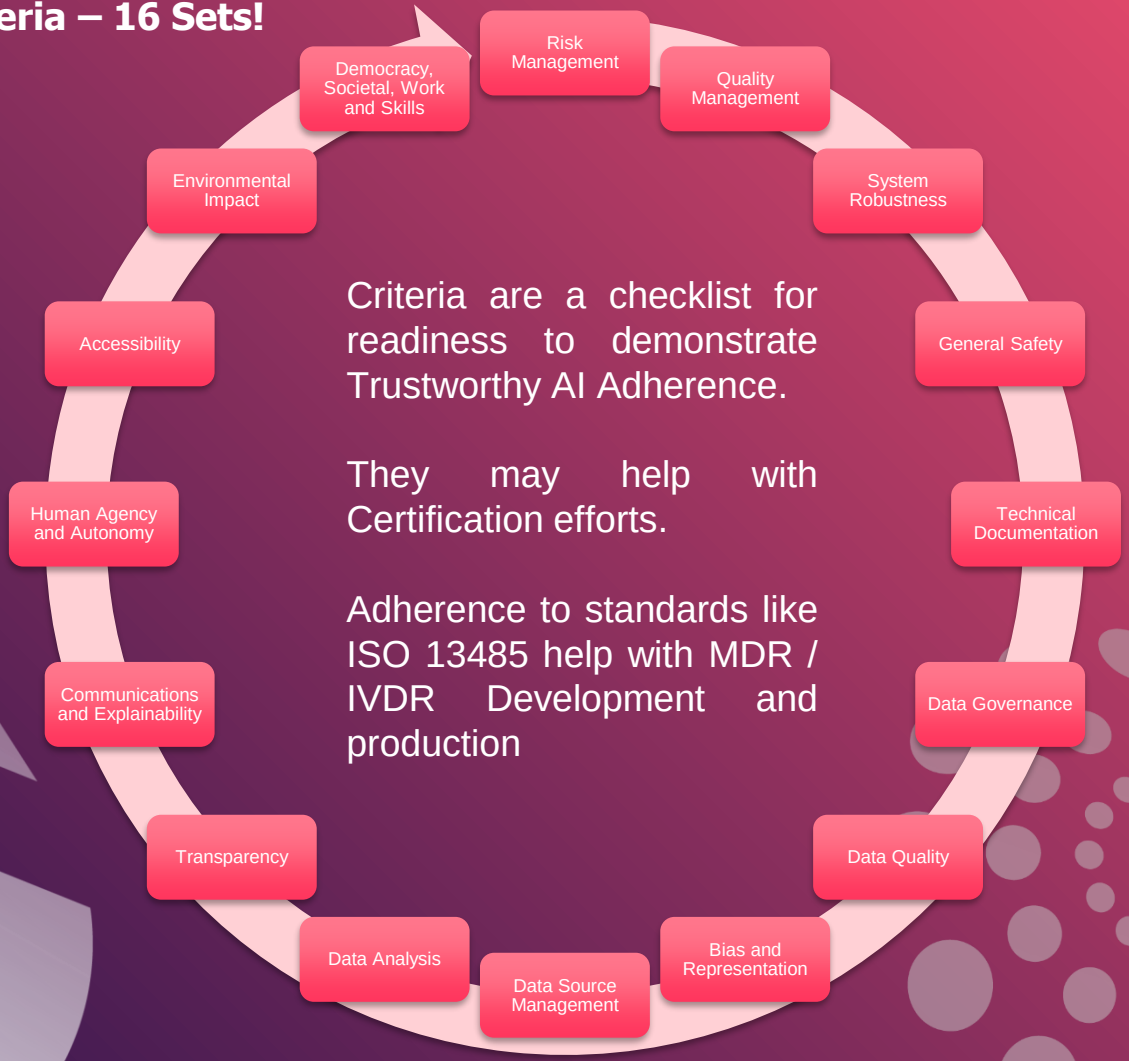
Provided a foundation and framework that allowed the compliance exploration

Allowed a deep dive into what that entailed for SYNTHEMA

Trustworthy AI Compliance



AI Compliance Criteria – 16 Sets!



Criteria are a checklist for readiness to demonstrate Trustworthy AI Adherence.

They may help with Certification efforts.

Adherence to standards like ISO 13485 help with MDR / IVDR Development and production

Assessments are Ongoing

The goal for the assurance is to internally self-assess against the criteria

The assessments need to be validated against evidence

Some of the Criteria are easier to assess against than others

For Data Governance links to the DPIA, Joint Controller Agreements etc. are an example	For risk management and quality assurance, reference to process and activity in line with established best practice are others	How do we assess against societal impacts…? Speculation

You will hear more about how we are addressing technical security and privacy assessments as specific development activity

Privacy assessment metrics are a key area for SYNTHEMA's regulatory compliance needs	Security and technology development decisions are others

Not Just About Compliance

- This isn't a tick box exercise
- You may achieve a high level of protection and certification, but you need to keep it
- Synthetic Data Generation implies data will be shared and there will be an ongoing assessment therein as well for distribution and reuse.
- This needs to be considered in line with the shifting regulatory scene that we are seeing
- It also needs to be founded on evidenced assessments – that you are about to hear more about!

Thanks!

Any questions?

Keep in touch! nathan.lea@i-hd.eu https://www.linkedin.com/in/nathanclea/

eurobloodnet.eu /ERNEuroBloodNet @ERNEuroBloodNet @erneurobloodnet.bsky.social

synthema.eu /synthema @SYNTHEMA_EU @synthemaeu.bsky.social

Funded by
the European Union

Acknowledgements

European Reference Network
for rare or low prevalence complex diseases
Network Hematological Diseases (ERN EuroBloodNet)

Funded by the European Union

This project is supported by the European Reference Network on Rare Haematological Diseases (ERN-EuroBloodNet)-Project ID No 101085717. ERN-EuroBloodNet is partly co-funded by the European Union within the framework of the Fourth EU Health Programme.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

SYNTHEMA

Funded by the European Union

SYNTHEMA is an initiative funded by the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101095530.