SYNTHEMA Newsletter Article

February 2024

Partial Automation of Data Protection and Privacy Risk Assessment

In this newsletter article, Dr Laura Carmichael and Dr Stephen Phillips from the IT Innovation Centre, University of Southampton, discuss their recent work on data protection and privacy risk classification as part of the SYNTHEMA project.

> "A common misconception with synthetic data is that it is inherently private. This is not the case. […] Significant care is required to produce synthetic data that is useful and comes with privacy guarantees." — Jordan et al. (2022)

Motivation

A wide variety of opportunities for generating and using synthetic data in health and social care exist – such as "to estimate the benefit of screening and healthcare policies, treatments, or clinical interventions" and to "augment machine learning algorithms". Organisations responsible for managing the generation, access and use of synthetic data must ensure legal and ethical compliance whilst also demonstrating trustworthiness and fostering social acceptability. Ensuring effective and appropriate tools, methods, and processes are in place for data protection and privacy risk assessment is essential. As part of this, threat modelling provides one "systematic" approach to eliciting and mitigating data protection, privacy, and security threats. Essentially, according to the Threat Modelling Manifesto, threat modelling gets us to answer four key questions: "1. What are we working on?", "2. What could go wrong?", "3. What are we going to do about it?", and "4. Did we do a good enough job?". From another angle, we can also view threat modelling as a way to achieve certain data protection and privacy goals, such as those defined by the Standard Data Protection Model, which are: "Data minimisation", "Availability", "Integrity", "Confidentiality", "Unlinkability", "Transparency", and "Intervenability" (note: these goals align with the core data protection principles as set out in Article 5 of the General Data Protection Regulation).

Our aim

The central idea is to explore how and to what extent threat modelling can support organisations to make consistent and transparent data governance decisions, in terms of generation, access and use of synthetic data and other related artefacts (e.g., machine learning models for generating synthetic data). In SYNTHEMA, we are looking at how Spyderisk — an existing semi-automated risk assessment tool based on systematic cause-and-effect modelling of threats – can help to automate various aspects of data protection impact assessment (DPIA) reporting, such as for generating data flows and risk treatment plans to be used as supporting evidence in high-quality DPIAs. To achieve this, we need to extend the Spyderisk modelling software and knowledgebase so that risk analysts can use this tool to assess data protection and privacy risks, specifically for federated research networks involving the anonymisation and synthesisation of data.

What we have done so far

Our first task has been focused on classifying data protection and privacy risks in this context. To do this, we have extended an existing data protection and privacy risk classification framework, developed in previous work, taking into consideration specific requirements relating to the generation, access, and use of synthetic data through a literature review. This framework provides standard data protection and privacy risk classification categories in accordance with the Five Safes — 'Safe Projects', 'Safe People', 'Safe Settings', 'Safe Data' and 'Safe Outputs' – which are well-known, best practice principles for safe research and used as dimensions to guide discussions and decision-making concerning access to data and other related artefacts.

As part of data protection and privacy risk modelling, we also need to understand the different types of individual, collective, and societal harms that may arise from potentially harmful situations and activities related to federated research networks involving cross-border synthetic data generation and use. We have therefore examined some different ways in which data protection and privacy harms have been categorised — such as, by the Information Commissioner's Office (ICO), the Nuffield Council on Bioethics, and the National Institute for Standards and Technology (NIST). Consideration has also been given to how synthetic data may amplify "collective data harms".

Our next steps

As an immediate next step, we are organising a workshop on partial automation of data protection and privacy risk assessment to explore the challenges of conducting and reviewing data protection impact assessments (DPIAs) – and how emerging semi-automated risk assessment tools and methods can help to support organisations with carrying out the process.

The data protection and privacy risk factors identified as part of this initial work will help to inform ongoing development of Spyderisk data protection and privacy risk modelling software and knowledgebase as part of the SYNTHEMA project.

* * *

Please note that all views and opinions expressed in this newsletter article are those of the authors.

* * *

About the authors:

Dr Stephen Phillips is a Principal Enterprise Fellow at the IT Innovation Centre, part of the Digital Health and Biomedical Engineering Group, School of Electronics and Computer Science, University of Southampton.

Dr Laura Carmichael is a Research Fellow at the IT Innovation Centre, part of the Digital Health and Biomedical Engineering Group, School of Electronics and Computer Science, University of Southampton.

In the SYNTHEMA project, both authors work as part of Work Package 5: Data Protection and Privacy Assessment.

For more information about Spyderisk, see:

- Spyderisk Open Project on GitHub. Available at: https://github.com/Spyderisk.
- Spyderisk System Modeller Documentation for a reference guide and tutorials on how to use Spyderisk for system modelling.
- Phillips et al. (2023) for a detailed overview of the Spyderisk approach to automated risk assessment.