GDPR Compliance Challenges for Interoperable Health Information Exchanges (HIEs) and Trustworthy Research Environments (TREs)

Ed Conley[1*] and Matthias Pocs[2]

[1] SHiELD Horizon 2020 and Connected Health Cities Projects, AIMES, Liverpool Innovation Park, L7 9NJ, United Kingdom

[2] SHiELD Horizon 2020 Project, Stelar Security Technology Law Research, 21035 Hamburg, Germany

Abstract

Background: We present our current approaches to improving personal data protection in (i) large (regional/national/international) scale health information exchanges (HIEs) and (ii) UK NHS IG toolkit and ISO 27001-compliant trustworthy research environments (TREs) for discovery science communities. In particular we examine impacts of the General Data Protection Regulation (GDPR) on these technology designs and developments and the responses we have made to control complexity.

Methods: The paper discusses multiple requirements to implement the key GDPR principles of "data protection by design" and "data protection by default", each requiring new capabilities to embed multiple security tests and data protection tools in common deployable infrastructures. Methods are presented for consistent implementation of diverse data processing use cases.

Results: We describe how modular compositions of GDPR-compliant data processing software have been used to implement use case(s) and deliver information governance (IG) requirements transparently. Security surveillance analysis is embedded throughout the application lifecycle, namely at design, implementation and operation (run-time) phases. A solution is described to the challenge of integrating coherent research (analytic) environments for authorized researchers to access data and analytic tools without compromising security or privacy.

Conclusion: We recognise the need for wider implementation of rigorous interoperability standards concerning privacy and security management. Standards can be disseminated within low-cost commodity infrastructures that are shared across consortium partners. Comprehensive model-based approaches to information management will be fundamental to guaranteeing security and privacy in challenging areas such as ethical use of artificial intelligence in medicine. The target architecture is still in evolution but needs a number of community-collaborative API developments to couple advanced specifications fulfilling all IG requirements.

Keywords

General Data Protection Regulation (GDPR); Information governance; Secure health information exchange; e-consent; Secure DevOps

Correspondence to:

Dr. Ed Conley Conley, Chief Scientific Officer, AIMES
The Innovation Park, Liverpool L7 9NJ, United Kingdom.
E-mail: ed.conley@aimes.net

EJBI 2018; 14(3):48-61
Received: March 26, 2018
Accepted: May 25, 2018
Published: July 06, 2018

1 Introduction

Our focus within the Horizon 2020 project SHiELD [1] and Connected Health Cities project [2] concerns data protection in health and research information exchange use cases. In particular, we are interested in impacts of the General Data Protection Regulation (EU) 2016/679 [3] also known as GDPR[1] on processing of personal data and on the free movement of such data [4]. This regulation can impose significant penalties for non-compliant data controllers and processors once it comes into force in the spring of 2018. Fundamentally, GDPR aims to provide a set of standardized data protection laws across EU countries. This is intended to make it easier for EU citizens to understand how their data is being used and to raise any complaints. For implementers, it has potential to reduce fragmentation and administrative burdens where business activities flow through local, regional, national and international data exchanges. A full treatment of the data protection principles that drive compliance to the GDPR is beyond the scope of this paper, but an abundance of introductory resources is available on the Internet.

[1] All abbreviations and acronyms used are tabulated at the end of the paper.

The new legal obligation of 'data protection by design' introduced by the GDPR requires data controllers to ensure (and demonstrate) that the traditional data protection principles like data subject rights, lawfulness, fairness and transparency, purpose limitation, data minimization, etc., are supported by technology design as an integral part of the system (GDPR Article 25(1)). GDPR has increased requirements for data controllers[2] to demonstrate compliance: Data controllers must build, implement and be able to demonstrate a comprehensive data privacy compliance programme. They must assess the "likelihood and severity of the risk" of any personal data processing operation relating to any use that "from personal data processing could lead to physical, material or non-material damage". The categories where risks could arise are summarized in Table 1. As a response to these, SHiELD consortium[3] proposes to use an open and extendable architecture with privacy-by-design modelling and embedded risk analysis tools. The aim is to provide systematic protection for storage and interoperable exchange of health data that is scalable across European borders. The exchange use cases are subject to permissions control (electronic consents) by the data subjects, compatible with existing regulatory frameworks. The goal is to ensure privacy, availability and correctness of health data whilst improving trust of patients in the security of their data and its use to address their needs.

This paper is a "current perspective" of challenges when implementing new large-scale infrastructures addressing health care and research domain problems (within the constraints of GDPR and manifold security threats). Data processing architectures are in rapid and continuing evolution, which further challenges for implementers faced with legal constraints. With respect to GDPR compliance, both controllers and processors need to demonstrate status and match their data processing steps to a collaborative IG plan. In this paper we illustrate how controlled reduction of complexity by fitting use cases to a symbolic abstraction set has benefits of increased transparency when applying IG rules across real-world data processing ecosystems. Controlled complexity reduction will become increasingly important as problem-solving data ecosystems scale and federate across the world.

The security and privacy standards landscape relevant to the domains of SHiELD and Connected Health Cities projects is summarised in Table 2. Meeting the challenge of interoperable privacy and security has been described as requiring services and mechanisms that are dynamic, distributed and intelligent [5]. Consistency and cross-compatibility of multiple deployed security and privacy solutions require conformance to international standards that are fit for purpose in complex domains of health and biomedical research. Currently developed international standards that support a path to greater interoperability do exist including standards for privilege management and access control developed by NIST, ISO and HL7 (Table 2). Traditional role-based access control (RBAC) standards are foundational but new specifications e.g. for security and privacy labelling (tagging) of segmented health information can improve interoperability (see also Discussion, Section 4). Comprehensive privilege management and access control (PMAC) principles [6] require explicit, ontology-based, formal (and therefore machine-processable) policies to implement at scale. While considerable theoretical work and a body of standards already exist for PMAC [5, 6, 7, 8], their degree of implementation in real-world solutions is limited (e.g. in large-scale programmes by Kaiser Permanente and

Table 1: Summary of categories where personal data processing could lead to physical, material or non-material damage.

Consequential Risk	Examples
Various losses	Discrimination, identity theft or fraud, financial loss, damage to reputation; loss of confidentiality of personal data protected by professional secrecy
No authorisation	Unauthorised reversal of pseudonymisation
Disadvantage	Any other significant economic or social disadvantage
Deprivation	Where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data
Revelations	Where personal data are processed which reveal racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership
Sensitivites	Processing of genetic data; data concerning health, sex life, criminal convictions and offences or related security measures
Personal evaluations	Where personal aspects are evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles
Vulnerabilities	Where personal data of vulnerable natural persons, in particular of children, are processed
Scaling risks	Where processing involves a large amount of personal data and affects a large number of data subjects

[2] For definition, see later section "Shared responsibilities and roles under the GDPR".

[3] The SHiELD consortium (in alphabetical order) is AIMES, Fondazione Centro San Raffaele (FCSR, Milan), IBM Research (Haifa), IT Innovation (University of Southampton), Metrarc, North West Shared Infrastructure Service (NWSIS, UK NHS), Osakidetza, Stelar Security Technology Law Research, Symphonic Software and Tecnalia. Illustrations of their corresponding interests and expertise are shown in Figure 4.

Table 2: Health informatics standards referenced by SHiELD including security and privacy elements.

Standard (alphabetical)	Sources: HL7 International, ISO TC 215 Health informatics, CEN TC 251 Health informatics, NIST
EN ISO 21549-5	Health informatics - Patient healthcard data - Part 5 Identification data
EN ISO FDIS 17523	Health informatics - Requirements for electronic prescriptions
Generic Component Model (GCM)	The Generic Component Model (GCM) as a system-theoretical, architecture-centric, ontology-driven, and policy-controlled approach to privacy and security [6]
HL7 HCS	Context-sensitive segmentation of health information in HL7 International Healthcare Privacy and Security Classification System (HCS) Release 3 [9]
ISO 13606	Health informatics - Electronic Health Record Communication - Part 4 Security
ISO 17090-5	Health informatics - Public-key infrastructure - Part 1 Overview of digital certificate services - Part 2 Certificate profile - Part 3 Policy management of certification authority - Part 4 Digital signatures for healthcare documents - Part 5 Authentication using healthcare PKI credentials
ISO 21298	Health informatics - Functional and structural roles
ISO 22600	Health informatics - Privilege management and access control (PMAC) [5] - Part 1 Overview and policy management - Part 2 Formal models - Part 3 Implementations
ISO 25237	Health informatics - Pseudonymization
ISO 27005	Information Technology -provides guidelines for information security risk management
ISO 27789	Health informatics - Audit trails for EHRs
ISO 27799	Health informatics - Information security management in health using ISO/IEC 27002
ISO TR 18638	Health informatics - Components of education to ensure healthcare information privacy
ISO TS 11633-1	Health informatics - Information security management for remote maintenance of medical devices and MIS - Part 1 Requirements and risk analysis - Part 2 Implementation of an information security management system (ISMS)
ISO/HL7 10781	Health informatics - Electronic Health Record Sytems Functional Model
ISO/IEC TR 80001	Application of risk management for IT-networks incorporating medical devices - Part 1 Roles, responsibilities and activities - Part 2-2 Guidance for the communication of medical device security needs, risks and controls (security capabilities) - Part 2-8 Application guidance - Guidance on standards for establishing the security capabilities identified in IEC 80001-2-2 - Part 2-9 Application guidance - Guidance for use of security assurance cases to demonstrate confidence in IEC/TR 80001-2-2 security capabilities
NIST Security Labels	Security Labels as described in FIPS PUB 188 [36]

the US Veterans Administration). This is a conundrum, given the strong and increasing societal demand for robust privacy and security systems and the large proportions of budgets often apportioned to these aspects. Ability to adapt to rapidly changing environments and wide use case challenges is also essential. In ISO 22600 [7], security and privacy domains are defined by explicit ontologies and policies that can dynamically adapt to changing contextual and environmental conditions and can represent individual preferences at any level of granularity [6, 8]. The HL7 International Healthcare Privacy and Security Classification System (HCS) Release 3 [9] consists of a system-theoretical approach for context-sensitive segmentation of

health information (enabling security and privacy labelling of data segments for machine processing - for potential benefits see Section 4). Of wide utility is also the Generic Component Model (GCM) [6, 8] as a system-theoretical, architecture-centric, ontology-driven, and policy-controlled approach to privacy and security. Constraining the GCM can systematically and formally model any system or subsystem of actors (persons, organizations, but also devices, applications, or components) in reusable segments bound to context-specific rules.

2 Methods

2.1 Approach to Secure Cross-Border Exchange (SHiELD)

The SHiELD use cases are based on cross-border health information exchange (HIE) via a national contact point to relay source system messages in respective countries. The approach to GDPR-compliance is direct implementation of the key GDPR principles of "data protection by design" and "data protection by default". Data protection tools are embedded in a common HIE infrastructure that is deployable by Secure DevOps technology [10]. The information exchange infrastructure is based on an extended OpenNCP architecture [11] itself implementing components of EpSOS [12]. The secure exchange of health data across borders is driven from a set of use cases (see below). Secure DevOps offers unique advantages for software deployments, semi-automating APIs that work over large geographies for source and receiving system connections. There are also advantages for reduction of infrastructure costs, efficiency of upgrades and security tool co-provision as part of the distributive model.

Detailed technical descriptions of SHiELD's privacy-by-design innovations such as security risk modelling, enhanced digital permissions (consent), and enforcement mechanisms shall appear elsewhere. The consortium is working together to specify procedures for privacy-by-design in eHealth interoperability solutions, refining and deploying infrastructure, preparing legal recommendations (for policymakers, regulators and standards bodies), engaging in threat modelling and designing risk mitigation tools. Other approaches identify security requirements and provide automated analysis of data structures to identify sensitive elements vulnerable to specific threats. The overall objective is to enable systematic protection of health data against threats and cyber-attacks.

2.2 Approach for Analytic/Research Data Processing (Connected Health Cities)

In the Connected Health Cities (CHC) project [2] we have considered scalability of multi-EHR system (i.e. regional-to-national scale) information exchange implementations coupled to analytic or research data processing. A key challenge is standardising information governance (IG) at scale (source systems collectively serving c.5 to 7 million patient population sizes). In these designs, patient permissions concerning data use and access can be created electronically (within any source system connected to the exchange). As part of well-established policy-based access control mechanisms [13, 14, 15] these are consistently enforced, independent of the information requesting system. Without system-wide consistency, complexities and ambiguities of interpretation (e.g. diverse consent models applying slightly different sharing rules) can compromise personal data protection. The CHC project is focused on development of scalable Learning Health Systems. These often require significant data processing to enact analytical and research processes (see example below). A challenge for creating consistent data processing infrastructure partly comes from variability in the technical specifications needed to meet complex information governance requirements. There are also significant differences in the way multiple (independently implemented) electronic consent solutions work. Implementers of consent apps often "begin again" and solve only the immediate (local) problems for information sharing or use (making no reference to existing interoperable standards for setting and enforcing policy-driven electronic consents). As a result, local consent applications frequently do not interoperate and compliance to the GDPR becomes more difficult to achieve in practice.

The core concept of Information Governance (IG) requires some definition. Health information governance is complex and has often been contextualised to a geographic region or legal system. The UK NHS legal framework governing the use of personal confidential data in health care includes the NHS Act 2006, the Health and Social Care Act 2012, the Data Protection Act, the Human Rights Act and the UK Data Protection Bill (equivalent to GDPR). The law allows personal data to be shared between those offering care directly to patients and protects patients' confidentiality when data about them are used for other purposes. GDPR impacts the entire spectrum of NHS uses with a "legal basis for data processing" needing to be established for all data flows. While analyses concerning quality of care, what treatments work best, commissioning of clinical services, public health service planning can use non-identifiable datasets, analysis that need to use personal identifiable data may require consent of the patient with some well-defined exceptions.

2.3 Legal and Standards Compliance as Basics of Security and Privacy

The cross-European scope of the SHiELD project has referenced an important body of standards work alongside the GDPR (Table 2) and reviewed in the wider context of interoperability [16]. In addition, organisational measures need to be taken in response to the legislative requirements from the GDPR. Whereas many elements and principles

already existed according to previous EU legislation (e.g., data minimisation, lawfulness, supervision by data protection authorities, purpose limitation, etc.), some have been introduced by the GDPR such as accountability, data protection impact assessments, data protection by design, data portability, one-stop shop, etc. For example, the GDPR requires certain elements and principles to be included in organisational measures such as binding corporate rules (for controllers [17] and processors [18] - see also Section 4.3, Figure 3). Key topics such as privacy policies, privilege management and access control have been specifically addressed by IMIA and EFMI Security Work Groups, but also at HL7, ISO, CEN standards development organisations. Joint IMIA (Security in Health Information Systems) and EFMI (Security, Safety and Ethics) work groups have recognised challenges in trustworthiness in the security and safety of solutions and infrastructure deployed. A joint workshop "Personal Health Data –Privacy Policy Harmonization and Global Enforcement" highlighted privacy concerns by presenting different cases and approaches to develop a mechanism for a global healthcare information certification framework.

The CHC project requires a vendor-neutral framework based on interoperability standards as a solution for consent. In a foundational project (miConsent, [19]) the implementation standards in the HL7 Consent Directive, IHE BPPC [14] and APPC [20] have been evaluated. Currently, we have not yet fully evaluated for suitability in SHiELD or CHC health information exchanges the HL7 FHIR Consent Directive [15] or Consent2Share [16] frameworks. Further work is underway within SHiELD and CHC (i) critically evaluating whether or not blockchain is a security technology compatible with the GDPR's "right to erasure" (see also Section 4.3, Figure 3c). SHIELD partners are developing OpenNCP-associated API's that will support cross-border interoperable consent statements. Tools also are also being developed in order to simplify use of XACML [21] (a general-purpose access control policy language) in health information exchange and these will be described in future publications. A number of methodological improvements for security and privacy interoperability are discussed in Section 4.

3 Results

3.1 Consistent Matching of Information Governance Requirements to Data Processing

A graphical method was used to map required information flows within a limited number of privacy zones [22]. We nominated privacy zones as Care Zone, Non-care Zone and Research Zone [23] but additionally incorporated a use-case driven Trustworthy Research Environment [24] for data analytics. The Trustworthy Research Environment (commonly abbreviated to TRE) is a fully-implemented system supporting secure, regulated (authenticated researcher) access to datasets and tools. We emphasise its name as "trustworthy" not "trusted" as it is designed according to a set of principles that are deserving of trust or confidence and as such are more dependable or reliable. We recognize a continual process for design improvements. The term "trusted" is an absolute which may not be defensible, for example in the event of a breach. If real-world breaches occurred, "trustworthiness" would mean that an immediate and effective mitigation measure would be put into place (by virtue of the security risk modelling tools, see below). TRE's work with underlying health information exchanges and standalone sources of data that require specific (bespoke) processing. Current generation TRE designs use virtual machine (VM) and secure network technology to implement interoperable interfaces, databases, data processing routines and transformation operators driven by IG requirements. The next-generation approach (also embedded in SHiELD) will also use Secure DevOps technology to deploy modular data processing infrastructure builds within a coherent technical architecture. A key innovation is the coupling of use case data flows (Figure 1a) to reduce complexity privacy zones describing the information governance requirements (Figure 1b) to actual data processing infrastructure needed to deploy the entire end-to-end system with use-case to use-case consistency (Figure 1c).

3.2 Creation of Trustworthy Research (Analytic) Environments

The role of the Trustworthy Research Environment (TRE) implementing compute applications [25] governs legitimate researcher access to data collections and the invocation of permitted analytic services. In the current-generation TRE, this was achieved through:

i. **Data provisioning** where data is stored in accordance with the NHS IG toolkit and ISO 27001:2013 standards; these compliance processes are not a 'one-off' but a matter of continuous improvement, vigilance and organizational awareness. Data provisioning also provides workflow infrastructure enabling production of data pipelines, which automate extract transformation and dataset preparation.

ii. **Analytics provisioning** is secured using two-factor authentication over VPN, providing (for example) university-based analysts with access to an eight core / 32GB RAM data science virtual desktop and access to software packages including common statistical packages and geospatial software. TRE's can be used in many areas of the research enterprise including collaborative drug target prioritisation, medication repurposing and stratification of populations into cohorts for personalized medicine, exposome/ adverse event registration, comparative treatment effectiveness and pharmacovigilance.

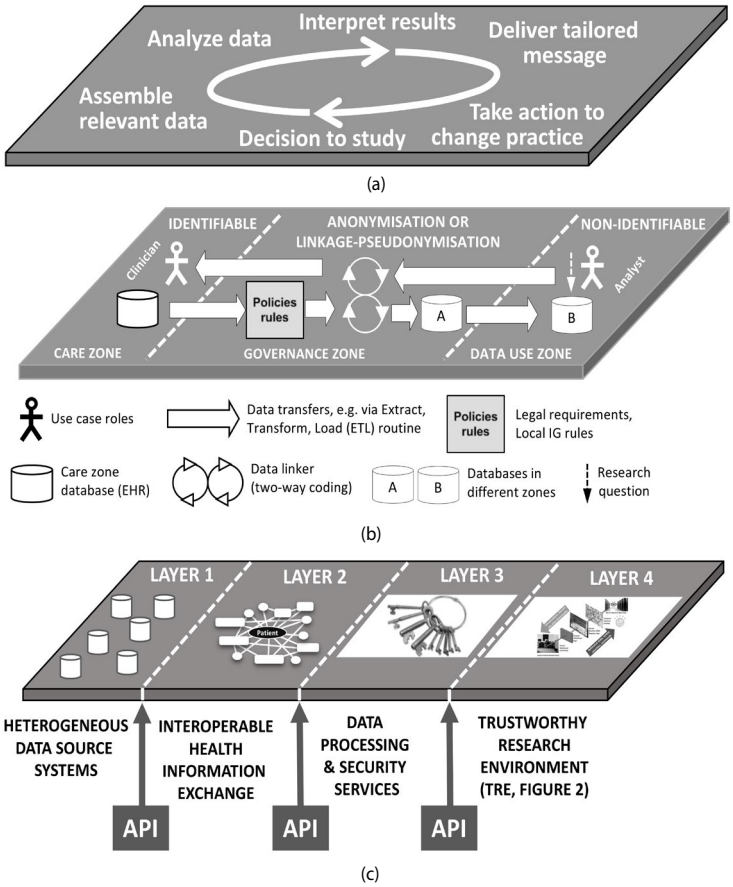


Figure 1: Consistent matching of use case information governance requirements to data processing; (a) An example use case (for learning health) of cross-system information flows; (b) Defined privacy zones for the total (end-to-end) cross-system data paths (IG model); (c) Four-layer architecture enabling end-to-end flows to enact the use case. Modular data processing infrastructure (Layer 3) and research environment (Layer 4) tooling is mapped to the IG model. APIs can permit bidirectional flow if the IG model permits. See text for management of contiguity of security within and between the four layers.

The design of TREs assumes re-use (large-scale hosting) of existing cohort data (for retrospective studies) and admission of electronic health record system data for prospective studies complying with the GDPR. The "Researcher View" of the TRE is illustrated in Figure 2. Security standards (above and Table 2) are maintained across different levels of a four-layer architecture with approved flows and privacy zones managed by an end-to-end IG plan. Currently TRE data processing components are selected and installed manually according to UK NHS IG toolkit and ISO 27001-compliant trustworthy research environment guidelines. Cross-border data exchange components in SHiELD will be selected and hosted using DevOps technology (Section 3.5). Where explicit consent is required for storage, sharing or use of personally identifiable data it is managed by electronic consent documents implemented using the IHE BPPC (Basic Patient Privacy Consents) profile. Permissions for sharing or use specify the access control within the health information exchange.

Common Interfaces (APIs) Coupling Analytic Information Flows: With reference to Figure 1, information flows from source systems exploiting a common API (all source systems need to agree standards to develop an interoperable "web-of-care" [26]). Source systems will likely include patient-identifiable data, within a classic health information exchange (HIE) that is designated as Layer 2 in the 4-Layer architecture (see Section 3.1, Figure. 1c, right-hand side). To facilitate cross-system exchange common APIs – i.e. using agreed standards to which all connected participants conform - are critical to scaling interoperability. In healthcare and patient-facing systems, interfaces use HL7 (V2, V3, FHIR) sometimes employing IHE profiles where they fit (e.g. IHE MHD for mobile clients) or web service interfaces (WS, SOAP). If further data processing is required (data transfers, storage, linkage/coding and security analytics) this constitutes Layer 3 (data processing) services. The specification of these is critical to ensure information processing compliant with legislation and the IG plan (Figure 1b). For example, there are obligations in GDPR for providing "opt-in, informed, free choice" consents, a mechanism to revoke such permissions, enable personal data portability and support statements on data holdings. In order to scale, all such features need to be facilitated as part of modular data processing services. Secure DevOps technology (Section 3.5) will assist information

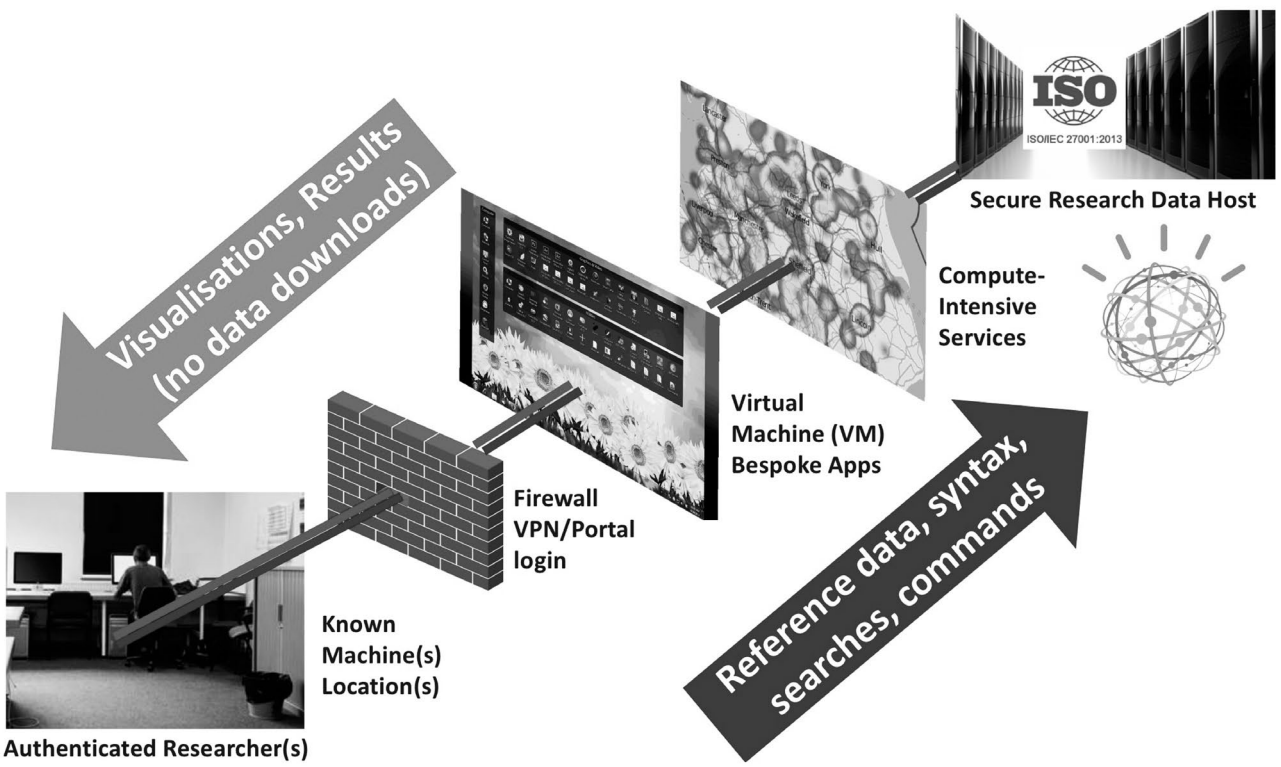


Figure 2: Researcher view of a current generation data-protective Trustworthy Research Environment (TRE).

system designers optimising selection of services and security validation tools in both design-time and run-time environments.

3.3 Comprehensive Security Threats Modelling and Mitigation for Use Cases

The SHiELD consortium implements a wide variety of risk mitigation tools in the context of its cross-border information exchange use case scope. It has introduced comprehensive security threat modelling and testing directly into the development process. The understanding comes from comprehensive intelligence of known descriptions of risks e.g. as described by ISACA [27] plus those in the collective experience of the SHiELD partners.

- Security risk mitigation approaches currently within the project include:
- **Asset inventory** - comprehensive records kept of assets and applications.
- **Configuration management** - Vulnerability modelling activities act as a comprehensive reference. Configuration tools are evaluated for capabilities in log management and additional threat analysis, intrusion detection and network vulnerabilities (for example: Puppet, Salt, Ansible, Chef and API-driven tools).
- **Counteraction measures** - Threat-associated rules that trigger threat counteraction mechanisms; these prevent unauthorized access, loss of data and cyber-attacks.
- **Documentation of policies/procedures** - Policies need to cover all steps of the production release process and need to be available to auditors.
- **Cross-border regulatory management** - Maintaining compatibility with regulations in different countries with data is being exchanged.
- **Logging of access and activity during development** - Timestamped code modifications against each developer, e.g. provided by Cucumber and Jira.
- **Introduction of novel security technologies** - Data hiding/masking and sensitive data analysis; anonymisation/pseudonymisation; provision of data and privacy protection to detect and prevent emerging threats such as inference attacks including cryptographic methods to prevent conventional attacks.
- **Peer review processes** - All code is peer reviewed with explicit rules regulating the independence of code approvers.
- **Performance** - Metrics are created with paths to solve problems.
- **Releases/deployments verification** - Automated releases require consistent deployment architecture for serving repeatable scalable processes described by

a use case (see Figure 1). Deployments use a rule base for consistency, but any design-time security mitigations need to be verified in the operational phase

- **Security experts** - Included as part of the stable development and deployment team.
- **Security training for developers** - Training for correct application of tests and external validation procedures.
- **Software module dependency tracking** - For reuse of fully defined blocks of code (modular computational workflow) to minimize opportunities for insecure code injection.
- **Streamlining processes** - Minimising errors through increased automation and raised quality; i.e. fewer code approvals but more trustworthy, continuous improvement.
- **Test types** - Static, dynamic, interactive and runtime application of security tests (evaluating tools such as Veracode, Waratek, Contrast Security, Fortify).
- **Traceability of lessons learned** - Tracking past software errors and mitigations.
- **Vulnerability points analysis** - Access control-related, device-related, consent-related; security tool assessments will adopt a continuous approach to analysing gaps.

3.4 Privacy-Protecting Legal Compliance Actions

The SHiELD health information exchange implements a number of legal and privacy-enhancement and security actions. These include Article 25 of GDPR for health data exchange, using documents of the Article 29 Working Party on Data Protection (e.g. on EHR) and European technical standardisation of "privacy by design" and obligation to "data protection by design and by default". These actions cover the GDPR data protection principles such as data minimisation, technical privacy measures such as pseudonymisation in response to the potential privacy impacts from automatic health data exchange.

Relevant international and European standardisation (ISO, CEN) is identified and addressed, for example, ISO/AWI 22697 ‚Health informatics - Application of privacy management to personal health information'. The collaboration agreements with standardisation bodies are approved by CEN-CENELEC/JTC 13 Cybersecurity and data protection. HL7 standards (e.g. CDA, V2, V3 and FHIR) are used in the technical implementations for documents and interfaces. Where appropriate, IHE profiles of standards such as Basic Patient Privacy Consents (BPPC) and IHE MHD (Mobile access to Health Documents) are employed. The project is currently developing architectural enhancements to the ePSOS/OpenNCP data exchange architecture (including extensions that address process models to handle incremental privacy threats and inference attacks, see Section 3.3).

Use cases for SHiELD include (i) chronic disease involving European travel with continuous monitoring and linkage of personal health data with secure exchange (ii) an emergency use case (e.g. stroke and loss of consciousness, with a "break glass" scenario to access records). In these use cases, patients are given the opportunity to consult their health data without having to reveal their identity to cloud operators that may be linked to previous consultations. SHiELD implements the only example of a technical and organisational measure that the GDPR [3, 4] offering pseudonymisation designed to achieve data minimisation (a prime example of ‚data protection by design' as cited in GDPR Article 25(1)). Like other legal obligations such as ‚accountability' (GDPR Articles 5(2), 24(1), which suggests to implement a data protection management system) the obligation of ‚data protection by design' is also subject to feasibility and risk-based conditions. Any data controller needs to take into account of:

The state-of-the-art - this may begin with standards such as ISO 25237 (health informatics pseudonymisation) or any future application of ISO/IEC 20889 (privacy-friendly de-identification techniques) in addition to guidance by the data protection authorities (for example [28, 29] are considered for relevance).

The nature, scope, context and purposes of processing - the use case descriptions need to be detailed to inform controllers accordingly.

Risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing - therefore, the European Charter of Fundamental Rights [30] has been analysed in the project.

Overall, a way of implementing ‚data protection by design' including pseudonymisation is the IG model. Both concepts are being used in the SHiELD project with a view to making subsequent proposals into technical standards bodies in the domain of health informatics [31], cybersecurity and data protection [32]. One currently undefined role is that of a scalable trusted third party (TTP) actor for generation and management of pseudonymisation keys. SHiELD needs to conduct external discussions in order to come up with meaningful recommendations for this functionality.

3.5 Impacts of Secure DevOps Technologies for End-to-End System Deployments

The SHiELD project [1] infrastructure development plan cites "Secure DevOps" methodology i.e. semiautomatic compilation of code, including deployment and testing with embedded security surveillance tools. This is a principal approach for raising security standards in health information exchange. Multiple security interventions can be embedded into the design and development phases. Currently the OpenNCP information exchange source code is being

analysed for performance with a range of security tools. Code and infrastructure elements will also be comprehensively tested at run-time with monitoring tools that can detect potential vulnerabilities (reports being generated for the developer/end user). Releasing software that has security vulnerabilities is a retrograde step. The Secure DevOps approach creates fundamental value for enabling reusable deployments meeting security and legal compliance requirements. It would impact every aspect of development, testing, integration, deployment and operations team work. It also represents a move to increasing automation of the agile application development process and deployment on to highly-scalable platforms. Such semi-automated approaches improve ability to deploy modular infrastructure builds specified to fulfil an IG plan (see Figure 1b, c).

4 Discussion

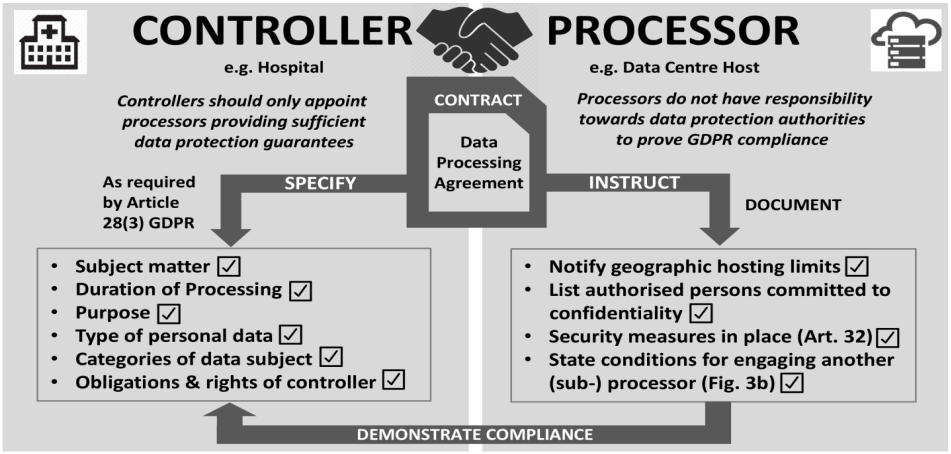
4.1 Raising "Trustworthiness"

Overall, our approaches follow (i) an open and extendable architecture supported by (ii) security mechanisms, (iii) privacy-by-design modelling (iii) risk analysis tools and (iv) Trustworthy Research Environments for research or analytic applications. The aim is to provide systematic protection for the storage, exchange and use of health care data across European borders and in distributed research projects. Within a SHiELD point to point information exchange, data use is controlled by the data subject, compatible with regulatory frameworks and compliance to the GDPR. The consortium members have a common focus on privacy, with improved availability and accuracy of data. This aims to raise the level of trust patients will have in the security of their data and its use to address their needs. This aim directs our focus on solving data security and privacy threats in different phases of the application lifecycle, namely, design, implementation and operation (run-time) using the methods and technologies.

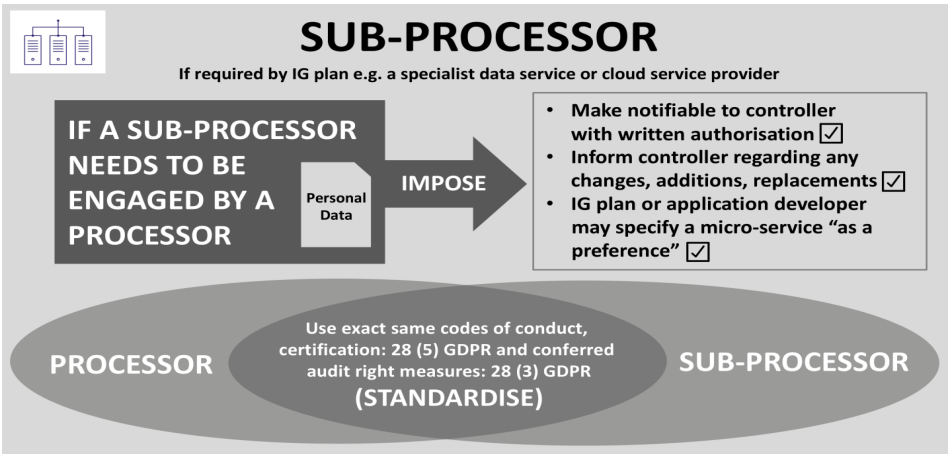
The wide range of test/validation checks exemplifies a key shift in the importance of security and data protection regulation concerns. There is a commitment for "attention to detail" as it is well known that simple mistakes and "weakest links" can easily create security vulnerabilities. Establishing security checklists is only a start point – a rigorous solution requires continuous evaluation and collaboration (for the system to meet fitness-of-purpose). The challenges in this paper can only be met by critical shifts in culture where security and data protection become the responsibility of all members of collaborating organisations using SHiELD outputs. In the future, these approaches are foundational for rational, secure and ethical approaches using artificial intelligence (AI) and personalized medicine [33]. For example, it is widely recognized that AI has untapped potential to improve reliability of diagnoses, higher quality prognostic indicators with applications in medicine [34]. Initiatives such as the 100,000 Genomes Project [35] already show the power of data combination governed by common data models from across multiple settings.

At current status, the projects *per se* are providing the Trust Framework - establishment of trust between the sender and the receiver systems. Trust is static i.e. established prior to any exchange through mutual participation, but in future a dynamic trustworthiness is needed, meaning that the conditions of the exchange and governing policies are negotiated at runtime. In this case, the expression and conveyance of policy includes the security labels applied to shared information and the application of privacy protections, markings and handling instructions bound to the exchange policies. In order to be effective, the Trust Framework must be legally binding and can apply retrospectively to the exchange pattern of publish and subscribe. The reliability of labelling solutions (next section) depends on the trustworthiness of the labelling entity and involved authorities including related accreditation and certification processes. The "cross-border" record sharing of SHiELD Health Information Exchange and "cross-domain" use of Trustworthy Research Environments makes the Trust Framework around core infrastructure critical (see role of the core OpenNCP infrastructure in Section 4.4, Figure 4).

Security/Privacy Labels to Model Use Cases, IG Zones, Data Processing Infrastructure: The "use case to IG zoning to infrastructure build" relationships shown in Figures 1a, 1b and 1c requires a substantially-researched interoperability framework in order to scale. One beneficial approach is security labels. These are markers bound to a resource, which connect an information object to a set of security and privacy attributes. The HL7 HCS specification defines Confidentiality labels, Sensitivity labels, Integrity labels, Compartment and Handling Caveats labels. The four labels (tags) can enable security and privacy rules about specific health information objects. Handling caveat labels convey dissemination controls and information handling caveats such as obligations and refrain policies to which an IT resource custodian or receiver must comply. Overall, Security Policy Information Files define which security labels are valid and how they can be checked against the Clearances – through these innovations, privilege and access control management in health information systems can be automated. The HL7 HCS Security Labels are described in NIST FIPS PUB 188 [36]. Operationalising the HCS is assisted at runtime by a Security Labelling Service [37] and the Privacy and Protective Services. The latter enforces obligations by applying various transforms to the response package including masking, redaction, annotations, anonymization or pseudonymisation based upon rules. If this standard was applied to the scheme illustrated by Figures 1a, 1b and 1c, objects can be reused by an access control system to support access decisions (e.g. matching classification labels



(a)



(b)

(c)

Figure 3: Shared responsibilities and roles under the GDPR; (a) Role of the data processing agreement and other expectations between data controllers and processors under the GDPR; (b) Conditions applying if the data processor needs to invoke a separate data processing service to fulfil the use case and IG requirements; (c) Data subject rights that the GDPR seeks to uphold.

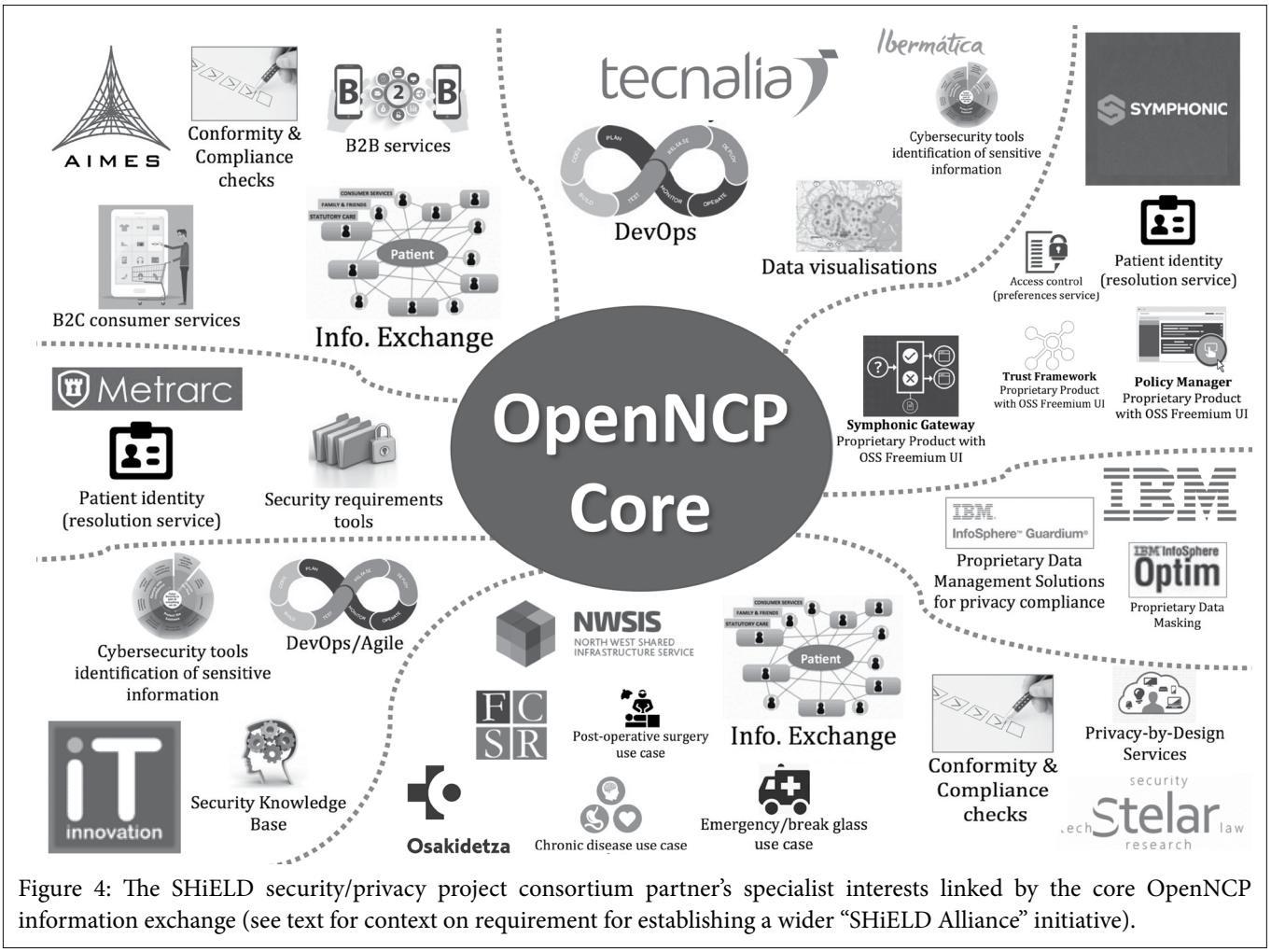


Figure 4: The SHiELD security/privacy project consortium partner's specialist interests linked by the core OpenNCP information exchange (see text for context on requirement for establishing a wider "SHiELD Alliance" initiative).

to clearances or other attributes specified by a security policy). These policies can be dynamic (e.g. in patient preferences) so HCS labels are applied at runtime (rather than being permanently stored with information objects). The runtime approach ensures the most current policy and trust framework (controlling the information exchange between sender and receiver) are enacted. Currently information exchange is bound by conventional IHE BPPC transactions, but their shortcomings are recognised. For example bespoke policy formulation is highly complex and the transactional nature of access control can become fragile as numbers of systems joined to the exchange increases.

4.2 Monitoring of System Privacy/Security Compliance

To date, privacy compliance checklists have been developed for organisations that are considered to be data controllers. Checklists are not as rigorous as the new GDPR obligation of data protection by design. GDPR brings accountability (not just responsibility) which means new requirements to demonstrate compliance. The data protection by design legal obligations address data controllers who may need to ensure the obligations are transferred to the suppliers. In the context of future SHiELD-based service use, data controllers could be hospitals, while data processors could be IT companies. Health organisations may act alone or as a joint buyer consortium creating supply tenders. They would specify data protection requirements and the SHiELD approaches could help meet the specifications. This would need to cover both the data protection side (coinciding with GDPR) as well as cybersecurity (digital security, information security, IT security, ISO27000-series) aspects and the legal basis for models of consent where these influence geographic scalability. The latter frequently depends on whether the GDPR permits national deviations; sometimes there are no extra permissions for the national legislators (e.g. a SHiELD "break glass" use case which concerns the vital interest of the data subject; in this case the national legislators cannot deviate from the GDPR rules according to GPDR Article 6 (1)(d), (2)). The concept of scalability is also tied to "legal interoperability".

4.3 Impacts of Shared (Contractual) Responsibilities under the GDPR

Irrespective of the approaches in this project some generic impacts also need addressing within a discussion of impacts. The applicable scopes of the GDPR is large - the

sum of national populations across the EU itself. Within the UK[4] for example, there will be more than 60 million data subjects (persons who have data stored about them) and approximately 500,000 data controllers (companies or organisations which store data about data subjects). The GDPR was intended to harmonise Europe's data protection laws. However, its flexibility and scope will likely create differences on how it is applied. Whereas a data controller is someone who "determines the purposes and means of the processing of personal data" (GDPR Article 4(7)), a processor is "any person who processes personal data on behalf of the controller (GDPR Article 4(8); other than a person who is an employee of the controller)". One of the major changes in the GDPR is that data processors have specific obligations. For example, if a processor fails to report a data loss to their controller, then the processor can be subject to regulatory action from the data protection authority (e.g. the Information Commissioner), and this is not possible under the Data Protection Act in such a strict way. To clarify these overriding issues, Figure 3a outlines some of the shared responsibilities between data controllers and processors as overriding considerations. Figure 3b summarises the relationship between processors and sub-processors. Finally the objective of guaranteeing data subject rights is annotated in Figure 3c. An organisation is likely to hold a data processor role if it does not decide the goals and means of data processing of the health data itself. It may host and maintain an infomation platform, but unless it is processing data for its own purposes, it is unlikely to be a data controller. A processor has much less responsibility towards data procesing authorities to prove compliance with data processing law. Data processors are not the first line of contact for Data Subject rights (GDPR Articles 12-22). It does, however have responsibilities to keep minimal records of processing it carries out for data controllers.

4.4 Scaling Security/Privacy Standards in Real-World Implementations

We set out to write a "current perspective" of challenges when implementing new large-scale infrastructures addressing health care and research domain problems (within the constraints of GDPR and manifold security threats). We acknowledge it is not a completed set of work but early communication (dissemination) is vital as there is a community-building aspect to the project. For example, a comprehensive list of security risk mitigation approaches (as described in Section 3.3) will require a community-based interoperability approach to sustain and refine. A key difficulty is how multiple interested parties (many of them competitors in the market and from mixed sectors of expertise) can move forward coherently and in control producing high quality pre-competitive guidelines that are actually implemented into interoperable products. One proposal for sustaining international coherence is the formation of a health data security and privacy "alliance" that would act to consistently implement standards in a shared non-proprietary infrastructure (e.g. the core OpenNCP ecosystem). Common components, interfaces and methodologies would be agreed, and incremental technical and policy developments could take place within implementation projects. The SHiELD consortium's journey shows that this proposition is challenging but given a collaborative ethos it is not impossible. Similar implementation initiatives have already taken place in other health market sectors (e.g. the Continua Health Alliance for personal health devices) resulting in coherent use case management, certification and test, policy alignment, technical working groups and ultimately shared interoperability guidelines (across hundreds of competing companies). In our projects, a wide set of expertise has also been essential to generate and critique cross-community solutions. Figure 4 illustrates how diverse expertise and interests of the current consortium partners have formed around the common OpenNCP infrastructure. This problem is certainly challenging. Of the list of practicable security/privacy standards (Table 2) some of which have reached HL7 International realm-specific Implementation Guide (IG) status for the US realm [38] and practically demonstrated e.g. in the Consent2Share project [16] which is under evaluation in our Connected Health Cities project. Cultural and legal specificities can act as a barrier for direct reuse of standards across national realms and adaptations are necessary to accommodate these and individual's needs.

[4] Despite Brexit, the UK will be implementing essentially all of GDPR into UK national law via the UK Data Protection Bill published on 14 September 2017.

5 Acknowledgement

The authors gratefully acknowledge project funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727301. We also thank Peter Gryffoy for analysis and insights underlying Figure 3 and to Marie Messenger and Michael Walker for proofreading; we also gratefully acknowledge the Connected Health Cities (CHC) programme commissioned by the Northern Health Science Alliance (NHSA) and funded by the UK Department of Health

Abbreviations and Acronyms Used

AI-Artificial Intelligence, **API**-Application Programming Interface, **APPC**-Advanced Patient Privacy Consents, **BPPC**-Basic Patient Privacy Consents, **CEN**-Comité Européen de Normalisation (European Committee for Standardization), **CHC**-Connected Health Cities, **DevOps**-Development Operations, **EFMI**-European Federation for Medical informatics, **GCM**-Generic Component Model, **GDPR**-General Data Protection Regulation, **EHR**-Electronic Health Record, **HCS**-International Healthcare Privacy and Security Classification System, **HIE**-Health Information Exchange, **HL7**-Health Level 7, **FHIR**-Health Level 7 Fast Healthcare Interoperability Resources, **HL7 V2**-Health Level 7 Version 2, **HL7 V3**-Health Level 7 Version 3, **IG**-Information

Governance, **IHE**-Integrating the Healthcare Enterprise, **IMIA**-International Medical Informatics Association, **ISACA**-Information Systems Audit and Control Association, **ISO**-International Standards Organisation, **MHD**-Mobile Health Documents, **NHS**-National Health Service, **NIST**-National Institute of Standards and Technology, **OpenNCP**-Open National Contact Point, **PMAC**-privilege management and access control, **RBAC**-Role-Based Access Control, **R&D**-Research and Development, **SOAP**-Simple Object Access Protocol, **TRE**-Trustworthy Research Environment, **TTP**-Trusted Third Party, **VM**-Virtual Machine, **WS**-Web Services.

References

[1] SHiELD - European Security in Health Data Exchange. http://projectshield.eu/

[2] Connected Health Cities Project. https://www.connectedhealthcities.org

[3] GDPR, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the EU. 2016 L 119, page 1.

[4] Pocs M. Will the European Commission be able to standardize legal technology design without a legal method? Comput Law Secur Rev. 2012; 28: 641-650.

[5] Blobel B. Davis M, Ruotsalainen P. Policy Management Standards Enabling Trustworthy pHealth. Stud Health Technol Inform. 2014; 200: 8-21.

[6] Blobel B, Nordberg R, Davis JM, Pharow P. Modelling privilege management and access control. Int J Med Inform. 2006; 75: 597-623.

[7] International Organization for Standardization. ISO 22600 Health informatics - Privilege management and access control. Geneva: ISO; 2006.

[8] Blobel B. Ontology driven health information systems architectures enable pHealth for empowered patients. Int J Med Inform. 2011; 80: e17-e25.

[9] HL7 International Inc. HL7 Healthcare Privacy and Security Classification System (HCS) - Release 3. Ann Arbor: HL7 International; May 2013.

[10] Yasar H, Kontostathis K. Secure DevOps Process and Implementation. In: Cybersecurity Development (SecDev); 2016 Nov 3-4; Boston, USA: IEEE; 2016. p. 166.

[11] Fonseca M, Karkaletsis K, Cruz IA, Berler A, Oliveira IC. OpenNCP: a novel framework to foster cross-border e-Health services. Stud Health Technol Inform. 2015; 210: 617-621.

[12] Moharra M, Almazán C, Decool M, Nilsson AL, Allegretti N, Seven M. Implementation of a cross-border health service: physician and pharmacists opinions from the epSOS project. Fam Pract. 2015; 32: 564-567.

[13] Integrating the Healthcare Enterprise (IHE). Basic Patient Privacy Consents (BPPC). https://wiki.ihe.net/index.php/Basic_Patient_Privacy_Consents

[14] HL7 International Inc. HL7 FHIR Consent Directive - current version of consent resource. http://build.fhir.org/consent.html

[15] Consent2Share open source software supporting an online consent process. https://bhits.github.io/consent2share/

[16] Blobel B. Standardization for Mastering Healthcare Transformation - Challenges and Solutions. Eur J Biomed Inform. 2017; 13: 09-15.

[17] Article 29 Working Party on Data Protection (WP29). Working Document setting up a table with the elements and principles to be found in Binding Corporate Rules (updated) (WP 256). Brussels 2017.

[18] Article 29 Working Party on Data Protection (WP29). Working Document setting up a table with the elements and principles to be found in Processor Binding Corporate Rules (updated) (WP 257). Brussels 2017.

[19] miConsent: Patient-controlled information sharing (legacy specification project combining HL7 and IHE standards). http://www.hl7.org.uk/marketing/archive/2012/120518_miconsent.asp

[20] Integrating the Healthcare Enterprise. IHE Advanced Patient Privacy Consents (APPC). https://wiki.ihe.net/index.php/Advanced_Patient_Privacy_Consents

[21] OASIS. A Brief Introduction to XACML.https://www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.html

[22] Delaney BC, Curcin V, Andreasson A, Arvanitis T, Bastiaens H, Corrigan D, et al. Translational Medicine and Patient Safety in Europe: TRANSFoRm - Architecture for the Learning Health System in Europe. Biomed Res Int. 2015; 2015: 961526.

[23] Kuchinke W, Ohmann C, Verheij R, van Veen E, Arvanitis T, Taweel A, et al. A standardised graphic method for describing data privacy frameworks in primary care research using a flexible zone model. Int J Med Inform. 2014; 83: 941-957.

[24] Lea N, Nicholls J, Dobbs C, Sethi N, Cunningham J, Ainsworth J, et al. Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research. JMIR Med Inform. 2016; 4: e22.

[25] Sintero Server - Simplifying interoperability for distributed collaborative health care. http://docplayer.

net/6258059-Sintero-server-simplifying-interoperability-fordistributed-collaborative-health-care.html

[26] Conley E. Realising the Interoperable ‚Web-of-Care'. Eur J ePractice. 2013; 19: 48-64.

[27] ISACA -Information Systems Audit and Control Association. http://www.isaca.org/

[28] Article 29 Working Party on Data Protection (WP29). Working Document 01/2012 on epSOS (WP 189). Brussels 2012.

[29] Article 29 Working Party on Data Protection (WP29). Working Document on the processing of personal data relating to health in electronic health records (EHR) (WP 131). Brussels 2007.

[30] Charter of Fundamental Rights of The European Union (2000/C 364/01) http://www.europarl.europa.eu/charter/pdf/text_en.pdf

[31] CEN/TC 251 Health informatics. https://standards.cen.eu/dyn/www/f?p=204:7:0::::FSP_G_D:6232&cs=18CA078392807EDD402B798AAEF1644E1

[32] CEN-CENELEC/JTC 13 Cybersecurity and data protection. https://standards.cen.eu/dyn/www/f?p=204:7:0::::FSP_G_307986&cs=1E7D8757573B5975ED287A29293A34D6B

[33] Blobel B, Lopez DM, Gonzalez C. Patient privacy and security concerns on big data for personalized medicine. Health Technol. 2016; 6: 75-81.

[34] Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017; 69: S36-S40.

[35] The 100,000 Genomes Project. https://www.genomicsengland.co.uk/the-100000-genomes-project/

[36] National Institute of Standards and Technology. NIST FIPS 188 - Standard Security Label for Information Transfer. Gaithersburg, Maryland, USA; 1994.

[37] HL7 International Inc. HL7 Version 3 Standard: Privacy, Access and Security Services; Security Labelling Service, Release 1. 2014

[38] HL7 International Inc. HL7 Implementation Guide: Data Segmentation for Privacy (DS4P), Release 1. Ann Arbor: HL7 International. 2014.