Advanced Research on Information Systems Security, an International Journal

GDPR Compliance Made Easier: the BPR4GDPR Project

Georgios Lioudakis [a]*, Eugenia Papagiannakopoulou [a], Maria Koukovini [a], Nikolaos Dellas [b], Kostas Kalaboukas [b], Lorenzo Bracciale [c], Emanuele Raso [c], Giuseppe Bianchi [c], Pierpaolo Loreti [c], Paolo Barracano [d], Spiros Alexakis [e], Renata Medeiros de Carvalho [f], Marwan Hassani [f]

[a]*ICT abovo P.C., Iridanou 20, 11528 Athens, Greece*

[b]*SingularLogic S.A., Achaias 3 & Trizinias, 14564 Kifisia, Greece*

[c]*University of Rome "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy*

[d]*Innovazioni Tecnologiche SRL, Via Arcidiacono Giovanni 43, 70124 Bari, Italy*

[e]*CAS Software AG, CAS Weg 1-5, 76131 Karlsruhe, Germany*

[f]*Eindhoven University of Technology, De Groene Loper 5, Eindhoven, The Netherlands*

[a]*Email: {georgios.lioudakis, mariza.koukovini, eugenia.papagiannakopoulou}@ict-abovo.gr*

[b]*Email: nikolaos.dellas@gmail.com, kkalaboukas@singularlogic.eu*

[c]*Email: {lorenzo.bracciale, emanuele.raso, giuseppe.bianchi, pierpaolo.loreti}@uniroma2.it*

[d]*Email: P.Barracano@intempra.com* [e]*Email: Spiros.Alexakis@cas.de*

[f]*Email: {R.Medeiros.de.Carvalho, M.Hassani}@tue.nl*

Abstract

With the aim to facilitate compliance with the GDPR, particularly for SMEs, this paper summarises the results of the H2020 BPR4GDPR project. With a focus on business processes, the project has proposed a holistic approach able to support compliant processes, while fulfilling requirements covering diverse application domains. The main pillars of the solution are: i) a policy-based access and usage control system, for setting the operational rules; ii) a framework for automatically re-engineering processes, so that they become compliant by design; iii) a run-time environment for the enforcement of privacy constraints and data subjects' rights; iv) a process mining framework, devised for ex post compliance analysis and conformance checking leveraging the process execution traces.

Keywords: Data protection; GDPR compliance; process re-engineering; process mining; PETs; access control.

---

* Corresponding author.

1. Introduction

The digital revolution has resulted in a lag between regulations and the current reality of the social media, Cloud Computing, Internet of Things, Big Data, to mention a few trends that didn't exist merely two decades ago and have resulted in increasingly interconnected systems, amazing processing power (and results thereof), and data proliferation. As dependency to technology increases, so do the information trails left behind following daily activities of people. To this end, the General Data Protection Regulation (GDPR) comprises a milestone step towards filling the "regulatory gap", creating an environment able to cope with the technological and business reality, and provide for the protection of privacy. Apart from the mandate for GDPR compliance—and the non-neglectable financial penalties, compliance is motivated also by the market needs, particularly the growing awareness of people, and their increasing demand that companies protect their information [1].

However, organisations keep declaring difficulties in GDPR provisions' implementation, despite the resources and money spent, whereas particular problems are faced as regards the new requirements GDPR introduces. The challenges, either technical or organisational, include, among others: interpretation of GDPR requirements; operational adaptation towards privacy-aware and compliant business practices; holistic data views and processing actions inventory; enforcement of security means; management of the relations with third parties and the data subjects, and enforcement of rights thereof; last but not least, significant resources are required and, whereas big companies may have human and monetary resources to invest, this does not in general apply for SMEs.

In light of these issues, the H2020 BPR4GDPR project [2], that successfully completed its work in 2021, has aimed at bringing about a new GDPR compliance paradigm, by providing the tools and methodologies for facilitating the implementation of the appropriate technical and organisational measures, particularly by SMEs, to ensure that data collection and processing are performed in accordance with the GDPR. Its technical contribution is highlighted in Section 1.2, after an overview of related work (Section 1.1).

1.1. Related work

Initially studied mainly in the context of computer and network security, privacy engineering has emerged not only as an important research field per se [3], but also as a growing market, fuelled by the compliance needs of organisations, as well as the increasing awareness and demands of users. Beyond cryptography and legacy security technologies, various research areas have spawned, including pseudonymisation [4], anonymisation [5], privacy-aware access control [6], differential privacy [7], privacy assessment [8], location privacy [9], privacy-preserving data analysis [10], and users rights' enforcement [11], among others, whereas in the Business Process Management (BPM) domain, most privacy-related research has focused on the annotation of processes and workflows with authorisation constraints and/or other data protection concerns (e.g., [12][13]). The protection of privacy —and compliance thereof— is also the focus of several European projects, among which BPR4GDPR, together comprising the "GDPR Cluster", and proposing complementary solutions [14].

In the emerging market of privacy solutions, the number of products is also growing; in the 2021 IAPP Tech Vendors Report [15], the number of products exceeds 350, being seven times more than the products discussed in the version of 2017. The types of solutions greatly vary, ranging from tools helping the data subjects to protect their data, to enterprise solutions, typically devised to serve the privacy office needs and the compliance programme management. There are various categorisations of related tools; the IAPP, for instance, identifies eleven types of tools [15].

As described in the next sections, BPR4GDPR not only builds upon the foundational technologies to provide a functional set of solutions, but also contributes to the state-of-the art with several innovative mechanisms. A major innovation is that it considers all prevalent aspects of GDPR compliance in process-aware systems through their suitable distribution in all phases of the process lifecycle. In addition, BPR4GDPR is the first project that leverages process mining with explicit focus on privacy awareness, enabling, on the one hand, the automated identification and documentation of existing organisational procedures and associated vulnerabilities, and, on the other, the assessment of their compliance after-the-fact with respect to modelled behaviour, fostering accountability and traceability. Third, BPR4GDPR goes beyond current approaches in the area of BPM security and privacy that either annotate process models with policies or verify the compliance of the former against the latter by making use of formal methods; instead, BPR4GDPR goes a step further, and provides for the automatic adaptation and transformation of processes in order to comply with privacy policies, both at design time and following execution. Finally, it is noteworthy that BPR4GDPR advances the state-of-the art in Privacy-Enhancing Technologies (PETs), with several important contributions, for instance in collaborative encryption, while providing comprehensive solutions for data management and enforcement of data subjects' rights.

1.2. Contribution and structure of this paper

BPR4GDPR has proposed a holistic compliance framework that covers the whole lifecycle of organisational practices, from process modelling to execution, as well as ex post analysis towards accountability and refinement. Building upon the project vision [16], the main technical contributions, reflected in this paper, are the following:

- A comprehensive policy-based access and usage control framework, conceived on the basis of the GDPR and managing all requirements thereof. It is devised to govern the overall organisational compliance and underlying systems' behaviour, and it relies on a semantic model referred to as the "Compliance Ontology" (Section 3).
- A framework for automatically verifying and re-engineering organisational processes, so that they become *compliant by design*; it is outlined, along with the underlying Compliance Metamodel, in Section 4.
- A set of functional run-time components addressing common needs of stakeholders as regards applied cryptography, access management, and enforcement of data subjects' rights. The easy to deploy, configure and integrate within an organisation's ICT environment "compliance toolkit", is described in Section 5.
- A process mining and discovery framework, devised for ex post compliance analysis and conformance checking leveraging the process execution traces; it is outlined in Section 6.

Prior to delving into each individual BPR4GDPR contribution, some general aspects are provided in Section 2, particularly in terms of the operational phases towards a holistic approach to GDPR compliance, and an overview

of the technical architecture of the project. The paper concludes with some insights concerning the evaluation of the BPR4GDPR solutions and concluding remarks (Section 7).

2. Overall approach

This section introduces the basic concepts of BPR4GDPR; it begins by identifying fundamental aspects pertaining to specification and execution of organisational processes, reflecting the points of intervention addressed by the project, followed by an overview of the components and interfaces that make up the architecture.

2.1. Operational phases

A fundamental characteristic of the BPR4GDPR approach is that it addresses GDPR compliance in a holistic manner, in the sense that the solutions aim at covering the full process lifecycle, from its initial identification to its enactment and execution, as well as its post-analysis. As illustrated in Figure 1, there are six main stages comprising the BPR4GDPR process lifecycle [16].

Phase 1 concerns the definition of a process model, through either its specification by an administrative user or its discovery based on event logs. **Phase 2** deals with the analysis of a process model in order to identify the risks, flaws and points of non-compliance, on the basis of well-defined policies; it is complemented by **Phase 3**, that provides for the automatic transformation of non-compliant process models, so that they are rendered inherently privacy-aware before being deployed for execution. **Phase 4** fosters effective enactment of GDPR-compliant processes, by entailing a comprehensive set of tools able to support the requirements arising from GDPR, whereas **Phase 5** extends process implementation by ensuring the compliant execution of processes, using a set of run-time tools. **Phase 6** entails process mining for ex post analysis of processes, in order to ensure that policies are enforced, and providing for the improvement of process models over time.

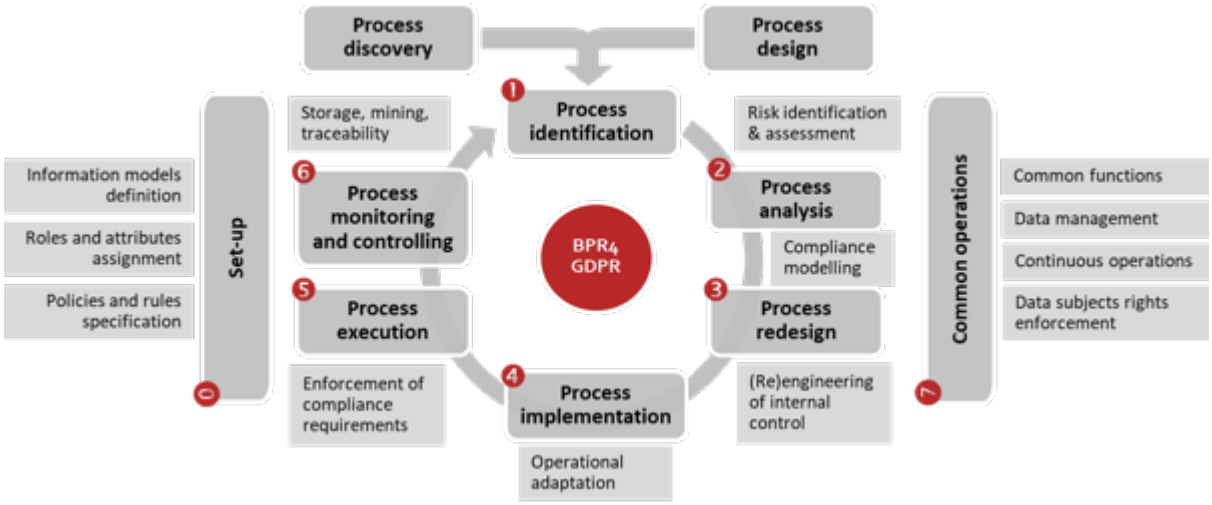

Figure 1: BPR4GDPR operational phases

Furthermore, BPR4GDPR considers two additional phases, vertical to the process lifecycle. **Phase 0** consists in all tasks that concern setting up the base elements of the BPR4GDPR operation, such as the specification of the

information models, the classification of data, systems and other resources, the assignment of roles and attributes to the different entities, the definition of purposes behind data collection and processing, and the specification of policies and underlying rules that should govern the operation of the system components. **Phase 7** refers to operations that are not (necessarily) part of a process lifecycle, but are rather executed asynchronously; they fall in different categories, including, supportive functions (e.g., authorisation mechanisms), enforcement of data subject rights, data management functions, and continuous operations, such as risk estimation, logging, etc.

2.2. Architecture

In order to cover its functional needs towards GDPR compliance and cope with the operational phases described in Section 2.1, BPR4GDPR has adopted the system architecture highlighted in Figure 2. As illustrated, the BPR4GDPR architecture is divided in four "quadrants", reflecting different groups of functionalities. **Governance** provides all functions related to the specification of policies and reasoning thereof, thus representing the Policy Decision Point (PDP) of the system. **Planning** concerns the specification of workflow models and their verification as regards compliance with the GDPR and their subsequent re-engineering, if needed, so that they become compliant *by design*. **Monitoring** deals with process mining and monitoring with the aim to identify discrepancies between compliant and actual behaviour. Finally, **Run-time** provides the means for the run-time system operation, particularly in terms of policy enforcement, data management, privacy-enhancing tools, and interaction with data subjects. The following sections outline the main principles and technical ideas.

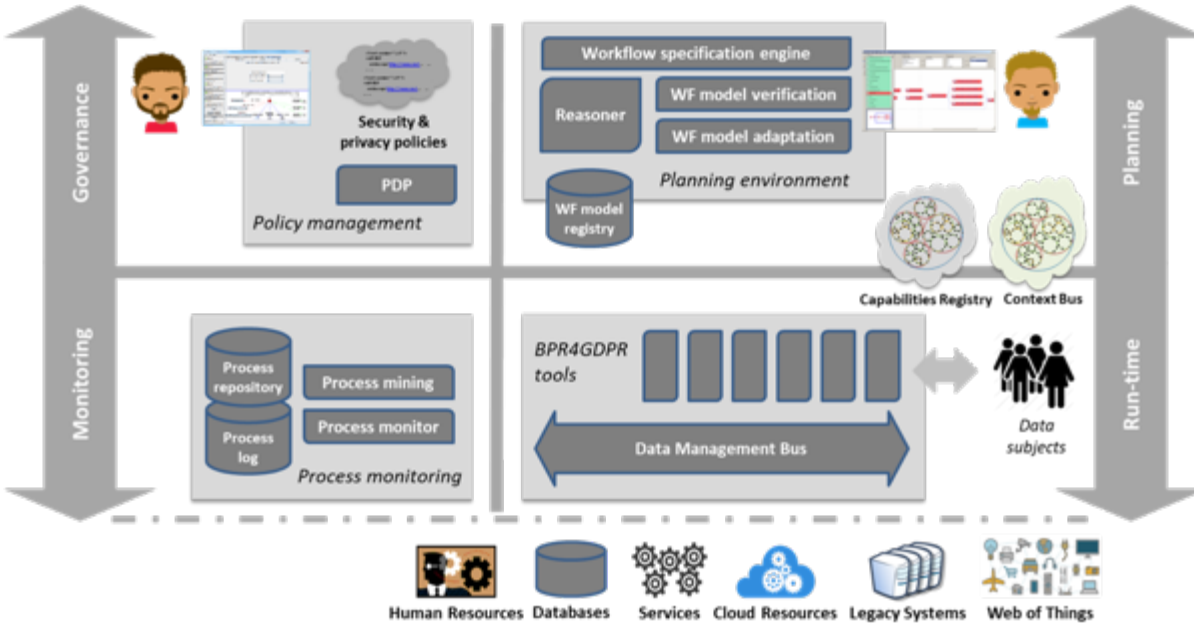

Figure 2: BPR4GDPR architecture

3. Governance framework

Policies are spotlighted at the core of the BPR4GDPR framework, as they comprise the drivers for the compliance-aware process verification and re-engineering, as well as for the run-time operation. Privacy and security policies are incorporated in the processes already during their specification or during the verification and re-engineering

phases. Run-time enforcement is achieved through the Compliance Toolkit, with policies regulating access to and usage of the underlying assets and prescribing the employment of privacy-enhancing mechanisms.

Starting from the thorough analysis of the regulatory provisions, particularly the GDPR, BPR4GDPR has provided a comprehensive framework for the specification of sophisticated security and privacy policies, able to capture the complex concepts stemming from the legislation, as well as the stakeholders' needs. The Compliance Ontology (Figure 3) [17] provides a high-level codification of GDPR into concepts that need to be considered by the policy framework, in the context of both run-time enforcement and process re-engineering. The Compliance Ontology includes, for instance, the types under which personal data fall, roles of the entities collecting and processing personal data and purposes thereof, operations and services performed over personal data, attributes of all entities, among others. It also considers the interrelations among identified concepts and provides for their thorough semantic structuring, by specifying hierarchies reflecting partial relations, such as generalisation/particularisation.

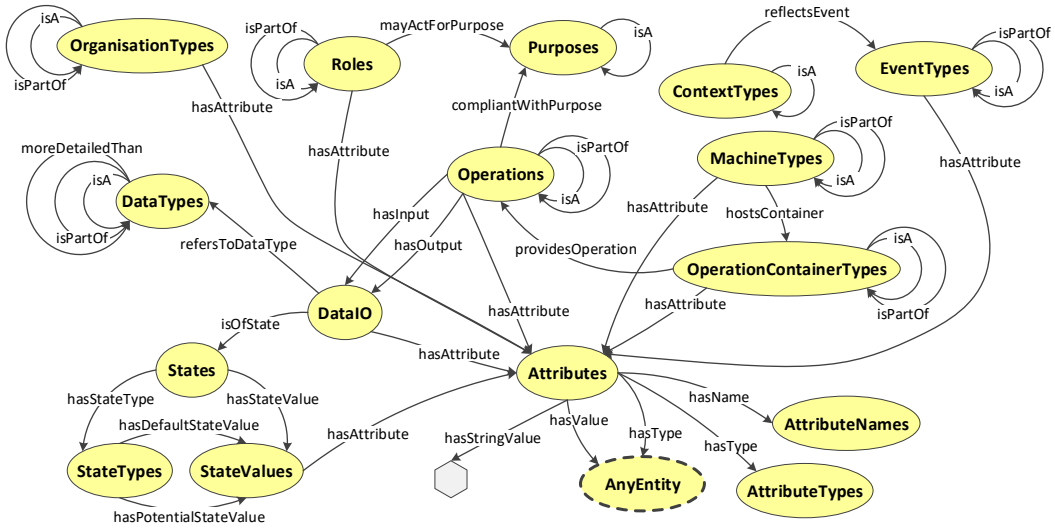

Figure 3: Compliance Ontology

The Compliance Ontology is extended with policies, formalised as sophisticated and fine-grained access and usage control rules [18]. The central concept here is the *action*, reflecting some *operation* performed by an *actor* over a *resource* in an *organisation*. More specifically, an action *act* is a tuple $\langle a, op, res, org \rangle$, such that *a* is an actor; *op* is an operation; *res* is a resource; and *org* is the organisation within which an action takes place. An action can be either *atomic* or *composite*, depending on whether the associated operation can be decomposed to more elementary operations or not. The elements of an action can be specified as *enhanced entities* that include, apart from the entity's semantic type, expressions over its attributes and/or sub-concepts, thus refining the concept definition, towards specifying attribute-based constraints and access and usage control rules. Upon the concept of actions, rules are specified as *permissions*, *prohibitions,* and *obligations* over actions:

$$\left.\begin{matrix} Permission \\ Prohibition \\ Obligation \end{matrix}\right\} (pu, act, preAct, cont, postAct),$$

where *act* is the action that the rule applies to, *pu* is the purpose for which act is permitted/prohibited/obliged to be executed, *cont* is a structure of contextual parameters, *preAct* is a structure of actions that should have preceded,

and *postAct* refers to the action(s) that must be executed following the rule enforcement.

For the realisation of the policies, BPR4GDPR has opted for an ontological approach, starting from the academic work described in [19]. The resulting Policy Model Ontology (PMO) (Figure 4) leverages the Compliance Ontology for semantically representing the domain entities, upon which it builds an expressive rule-based policy framework. Its expressiveness allows the specification of complex interrelations and dependencies between loosely-related actions and the enforcement of Separation and Binding of Duty (SoD/BoD) constraints, while hierarchies provide for comprehensive and simpler specification and formalisation of fine-grained security and data protection requirements, at any abstraction level. The policies allow for attribute-based data collection and handling, as well as for managing all associated constraints, including retention periods and the application of protection measures.

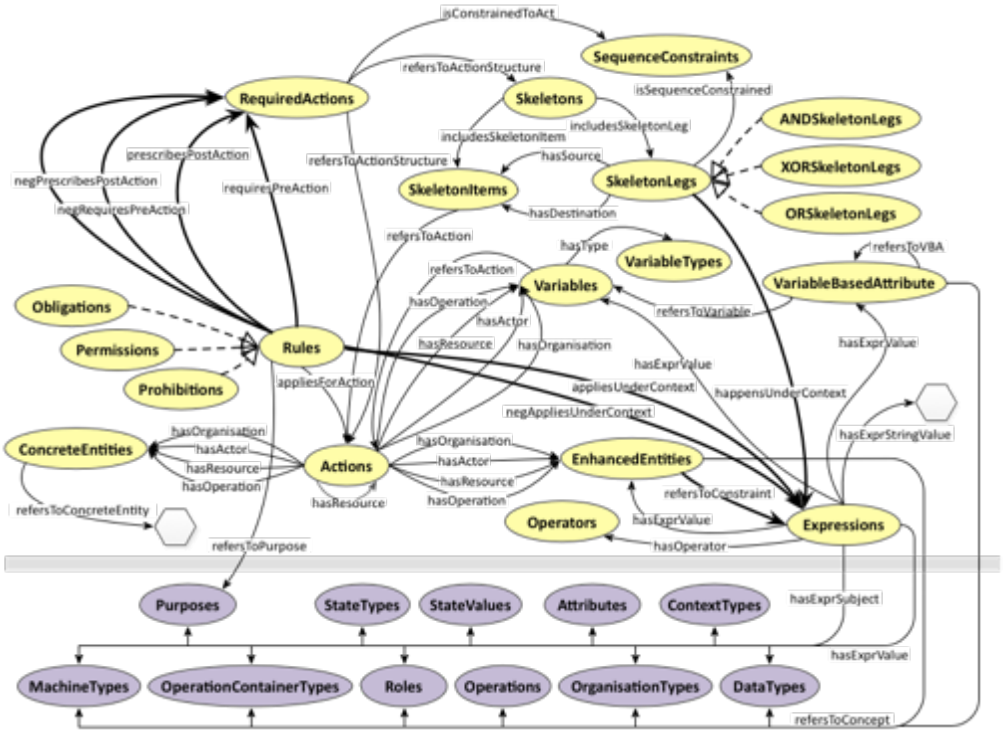

Figure 4: Policy Model Ontology (PMO)

BPR4GDPR has provided all necessary software for serving the dual role of the policy framework (cf. [18]): on the one hand, for being a functional PDP at runtime, supporting the XACML standard [20] for handling requests; on the other hand, for driving the process re-engineering aspect of the project, as described in the next Section.

4. Process planning and re-engineering

Fostering GDPR compliance at the level of process models, BPR4GDPR has focused on two aspects, reflecting important needs towards compliance: i) on how to incorporate GDPR constraints in the process models; ii) on how to verify and eventually re-engineer process models, in order to make them compliant with GDPR provisions.

As regards the first aspect, BPR4GDPR is grounded upon a Compliance Metamodel (Figure 5), that leverages the

ontological implementation of processes through workflows, proposed in [21]. The metamodel presents several innovative features, including: i) it enables the comprehensive specification of workflow elements, providing extensive coverage of core workflow perspectives [22]; ii) it introduces the novel concept of *assets*, as a means for representing the entities being subject to the execution of tasks; iii) it allows modelling of both control and data flows; iv) its expressiveness supports the expression of complex and varying security and privacy constraints.

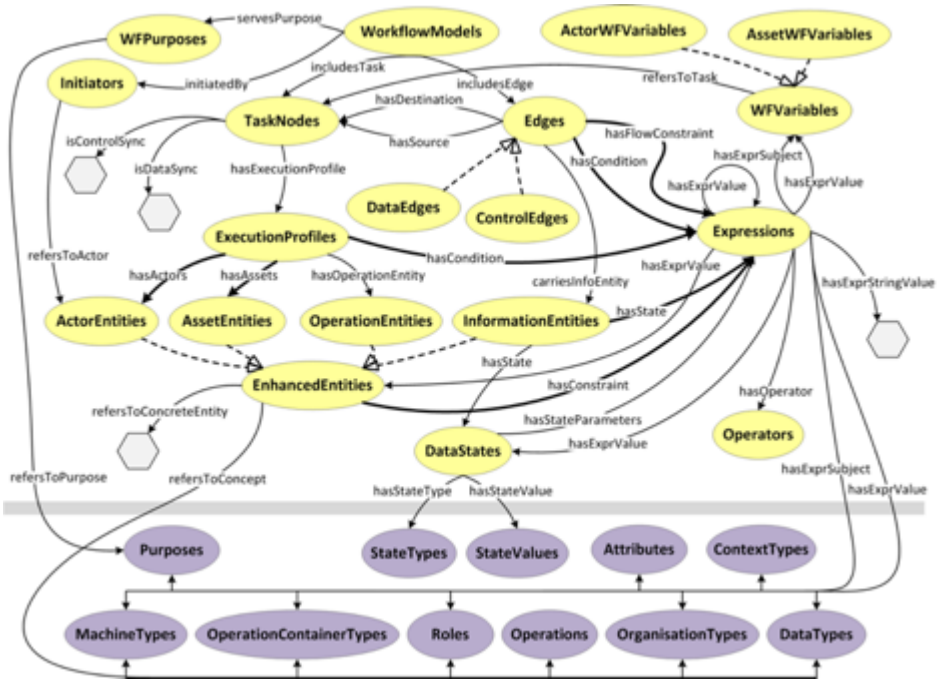

Figure 5: Compliance Metamodel

The most fundamental artefacts of a process model are *tasks* and *flows*. The former represents actions to be executed within the workflow, each describing the *operation* performed by an *actor* on an *asset*. Flows, or *edges*, express dependencies between tasks and are of two types: *control* and *data*. A model is complemented by the operational *purposes* it is meant to serve, and the potential *initiators*, denoting entities authorised to initiate the process.

A core feature of the Compliance Metamodel is the introduction of the concept of *enhanced entities*, devised for the comprehensive representation of actors, assets, operations, and information passed among tasks. Leveraging enhanced entities, each relevant concept is represented by its semantic type along with constraints that may refer to its attributes, contextual conditions, or relations among concepts. This provides for specifying SoD/BoD constraints inside the process model, thus satisfying an important authorisation requirement for workflows [23].

At the task level, conditionality is achieved through the innovative mechanism of *execution profiles*, that allows defining alternative ways to execute a task. This actually concerns two aspects: differentiated execution based on some conditions, and capturing the dependencies between the task's actors, assets and operation constraints, that is, precisely defining their valid combinations. On the other hand, at the level of flows, conditionality is realised by means of constraints reflecting flow properties and conditions that refer to environmental attributes.

On the basis of the Compliance Metamodel, BPR4GDPR has put in place a framework for the verification and compliance-aware re-engineering of process models. To this end, a functional model checker has been developed, in order to evaluate, and appropriately transform models to compliant ones, taking into consideration a variety of aspects and criteria, including: i) the validity of underlying purpose(s); ii) the authorisation of stated initiator(s); iii) the validity of each task's specification (e.g., in terms of authorised combination of actor, operation and asset); iv) the validity of each workflow edge representing a flow (e.g., in terms of permitted data exchange); v) the presence —positive or negative— of tasks in the process, considering also their absolute or relative position (e.g., whether consent has been provided); vi) whether the input and output of a task are valid against the underlying provisions, for instance whether the data handed to a task are proportional and in valid state (e.g., anonymised). For the verification and re-engineering to take place, the initial process model is analysed at various levels of granularity (instance subgraph, workflow case, bilateral task association, individual tasks, and edges, etc.), and eventually recomposed in order for the compliant model to be generated. The necessary knowledge upon which reasoning takes place is stemming from the Compliance Ontology and the PMO (cf. Section 3), whereas the PDP provides the model checker with the knowledge extraction and reasoning services.

Whereas the algorithmic aspects of verification and re-engineering are described in detail in [24], Figure 6 illustrates an example of process re-engineering, originating from the BPR4GDPR trials, and corresponding to the process of a medical prescription creation and dispensation. Indicative changes in the process model to highlight include: i) a `ProvideConsent` task is added prior to display the list of a patient's prescriptions to a doctor, to ensure that the patient (data subject) has been asked to provide consent thereof; ii) three instances of the `LogOperation` task are inserted for accountability purposes; iii) an `AnonymiseData` task prevents the disclosure to the pharmacist of certain data types contained in the prescription, as irrelevant to its dispensation.

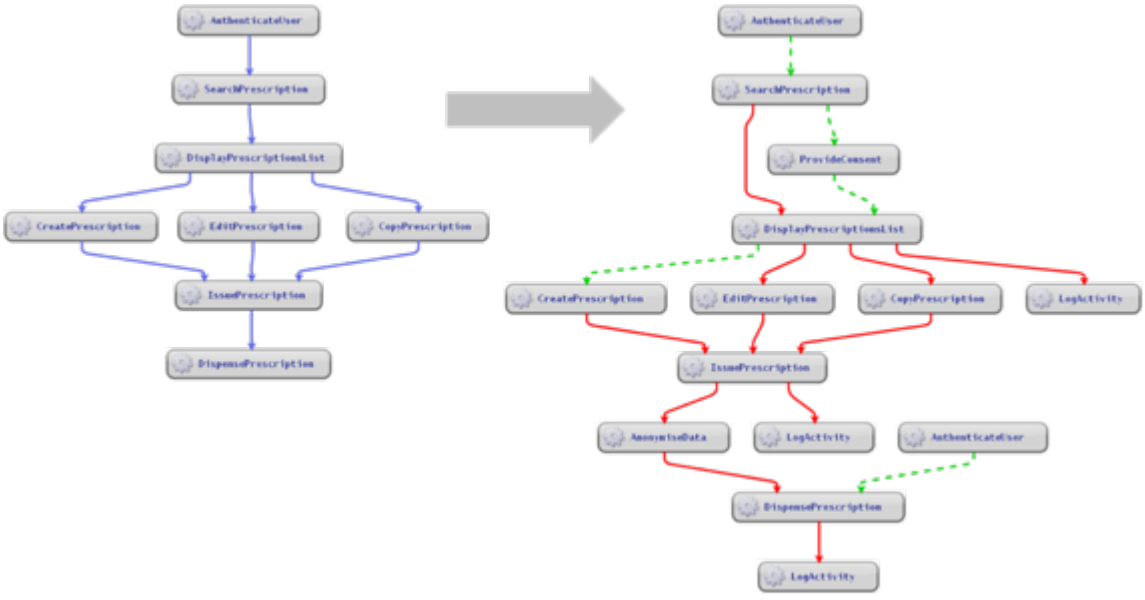

Figure 6: Process re-engineering

5. Runtime "compliance toolkit"

In order to facilitate the deployment of appropriate technical measures, as required by the GDPR, BPR4GDPR

has provided a set of functional components addressing common needs of stakeholders. The "toolkit" consists of modular functions that are easy to deploy, configure and integrate within an organisation's ICT environment, by leveraging appropriate virtualisation and integration means [25]. The provided tools are clustered in three groups, notably data management, privacy-enhancing technologies, and user-centred tools, presented in the following.

5.1. Data management

One of the very first user requirements that emerged during the project has been that of unified data management across multiple and heterogeneous data sources, in order to: i) have an effective Policy Enforcement Point (PEP) regarding access to and usage of data; ii) facilitate the implementation of data subjects' rights; and iii) implement fundamental for GDPR compliance data management functions, related to, e.g., retention periods and storage.

For data management, BPR4GDPR has developed a generic policy-driven semantic-oriented middleware for handling data access and usage requests, employing the necessary mechanisms for controlling data collection, pre- and post-processing, storage and dissemination in a fine-grained way. It is flexible to support a range of concerns and environments, through the transformation of the Platform Independent Model (PIM), as defined by means of the Compliance Ontology, to Platform-Specific Models (PSM). The core functionality is provided by the Data Management Bus (DMB), that can be extended by plug-in tools devised for specific tasks, such as retention periods' management. The main functional features and APIs of the DMB are the following:

- **Policy Enforcement Point (PEP)**: The DMB provides the mechanism for enforcing access and usage policies. It interacts with the PDP following the XACML protocol [20], whereas, based on the PDP response, some tools may be invoked (through the appropriate invoker) for the enforcement of the policy decision.
- **Policy Information Point (PIP)**: The DMB gathers information from external sources (e.g., databases) about attributes of the entities participating in the access requests, along with their values, by utilising appropriate data connectors; this kind of information is necessary for the PDP in order to come up with the access decisions.
- **Messaging**: The DMB is responsible for real-time data delivery between BPR4GDPR components and tools, as well as external entities (e.g., databases). Further, based on the PDP instructions, it is responsible for invoking the appropriate tools, e.g., for data processing, leveraging information stored into the Capabilities Registry.
- **Data connectors**: In order to support the variety of possible data sources, the DMB provides for connecting to data sources, and for effective transformation between data models. To this end, Apache NiFi [26] has been adopted as the proposed solution, although other implementations are not excluded.
- **Capabilities registry**: The DMB is used for storing information needed for the dynamic invocation of tools and data connectors. For each tool, it stores semantic information about the provided operations and parameters thereof (such as inputs, outputs, service parameters, etc.), as well as the endpoint connection details.
- **Data Subjects API**: This provides support for the enforcement of data subject rights (e.g., right to access).
- **Querying API**: It supports querying of data that belong to multiple data sources in a SQL-like format. All queries are based on the BPR4GDPR Compliance Ontology and should provide the underlying purpose(s).
- **Tools API**: It provides a common way to instruct tools to perform some operation (e.g., to anonymise data).
- **PIP API**: Through this interface the PDP can request additional attribute values; the requests are forwarded to external data sources (e.g., databases, LDAP, environment resolvers) by utilising the core DMB functionality.

5.2. Privacy enhancing technology for GDPR

The GDPR explicitly mention the use of encryption for improving security (Article 32) and mitigate risks (Recital 83). However, the mere data encryption is only a part of the story. How to distribute cryptographic keys, how to update/revoke keys and, in general, how to properly handle cryptographic keys in domain-specific use cases can make the difference from usable encryption and the inappropriate (or useless) use of cryptography. The BPR4GDPR approach was devoted to focus on such problems, specifically on exploring innovative solutions of cryptographic key management and on the use of Privacy Enhancing Technologies (PETs) to create privacy-preserving systems. BPR4GDPR developed four tools using PETs to propose four different solutions:

CoProtect. Albeit cloud is driving the enterprise workloads, according to a Microsoft survey, over 90% of the public and business leaders recognise data as the most critical company asset and are worried about their security, availability, and privacy in the cloud [27]. CoProtect comprises an encryption tool based on the collaboration between companies and cloud providers for cryptographic key management: the encryption key is split into fragments held by each of them. On the one hand, this gives the companies the control of their data, and on the other hand it offers disaster recovery and protection against accidental key loss or theft by any of the actors. It allows companies to be the sole responsible for their data disclosure and foster the construction of data access and modification logging service (required, in some cases, by GDPR), other than implementing their own access control policies independently. More information about the tool is available here [28].

AbeBox. A straightforward solution to decouple access control by the data service provisioning is to simply encrypt data before being uploaded to the cloud, and then decode the data just after the download. However, if this can be satisfactory from a security standpoint, it can be devastating from an operational point of view since dynamic management of the cryptographic keys inside an organization can be far from being trivial. BPR4GDPR proposed AbeBox [29], a solution which uses Attribute-Based Encryption (ABE), in order to provide a management model for secure service provisioning. The tool can run on top of existing filesharing system.

CYRVM. BPR4GDPR applied PETs also in the domain of risk assessment, mentioned in both Articles 25 and 35 of the GDPR. The tool, called CYRUM (Cyber Risk and Vulnerability Assessment) [30], implements a privacy-preserving collaborative recommending system for Cyber Risk Assessment, improving the quality and the accuracy of the assessment using collaboration among companies which can be potential competitors. It uses off-the-shelf arithmetic, performs similarly to a non-protected system, and employs trivial-to-explain cryptographic techniques which even a layman person may understand.

DiffPriv tool. Anonymisation aims at eliminating personal data so that data subjects can no longer be identified. Thus, anonymised data do not fall within the GDPR (Recital 26) as data are no longer considered “personal data”. However, performing proper anonymisation is a really complex process. Removing users’ identifiers (e.g., the name) can be completely ineffective, since users can be possibly recognised by the combination of some other parameters (e.g., the address), or from complex behaviour attributes called quasi-identifiers [31]. Recently, a new mechanism to anonymise data, the “Differential Privacy”, has become the dominant way to perform data anonymisation, because it provides a quantitative measure of privacy, has a rigorous mathematical foundation [7],

and the strong benefit of regulating privacy budget numerically by just one parameter (epsilon). For these reasons, very recently, this approach has been followed also by big over-the-top providers, such as Apple and Google. The BPR4GDPR DiffPriv tool is based on this technique and its goal is to simplify access to Differential Privacy by providing an easy-to-use and easy-to-deploy system to anonymously release data.

5.3. User-centred tools

The GDPR includes a wide range of existing and new rights for the data subjects, such as information and notification, provision and withdraw of consent, the rights of access, to erasure ("right to be forgotten"), to restrict processing, to data portability, to object, to rectification, the right not to be profiled. In order to help organisations to cope with these requirements, BPR4GDPR approach has been the development of the user-centred tools, that implement the corresponding services, and reduce the organisations' effort to implement these common functionalities. The user-centred tools provide a toolkit and software architecture to enable organisations to easily integrate them, so they are reusable in multiple application fields across the boundaries of organisations.

The toolset allows an organisation to inform a data subject about the personal data processing policies, as they are updated, and to acquire the various types of consent, putting the data subject in the position to easily manage consent. The data subject is therefore able to exercise all rights in relation to the GDPR, thanks to a practical interface, the allows to know what data are in the systems of the organisation, who has had or not access to the data, to request correction or cancellation, etc.

The user-centred toolset facilitates management of GDPR processes between the organisation and its stakeholders. It enables the organisation to administer the acquisition of consent, its possible subsequent changes, together with the requests of the data subjects in relation to their rights. The toolset therefore allows the designated manager of the organisation, to administer all data subject requests in a practical and effective way, saving both time and efforts needed for these processes. Further, when full automation is not possible, the toolset allows management of a data subject request in a "manual" way, by sending the request to the designated data source administrator, who will interact with the specific BPR4GDPR interface, managing the execution of the request.

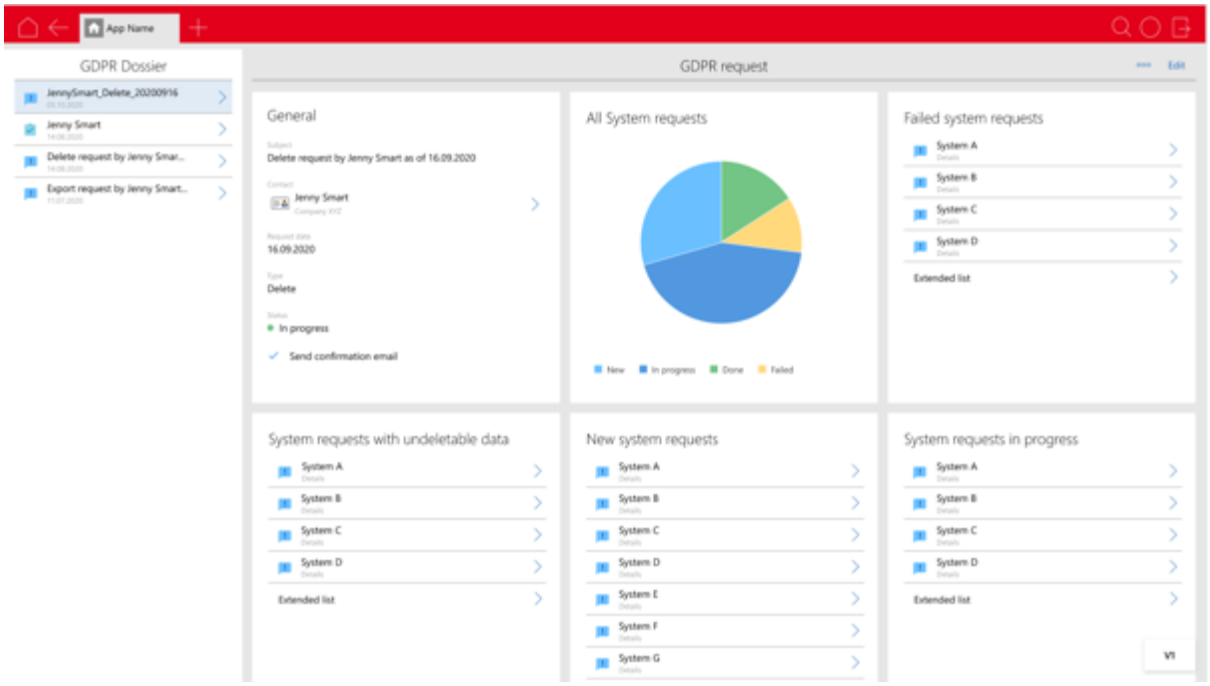

Figure 7: User-centred tools – data subject view

So, the toolset consists of three main Interfaces which allow all the involved parties to manage their tasks and options: the user (data subject) area, the administration area (for the DPO or main privacy manager), and the linked sources manager (IT managers or responsible for management of the third-party software). The toolset has been implemented in two versions, respectively for two of the project pilots' cases. Figure 7 illustrates the data subject view of the CAS UI [25], presenting a GDPR request from Jenny Smart. The view is separated in several slots, displaying all GDPR-related requests from Jenny (GDPR Dossier), general information about some Jenny's request (such as its status), a diagram with the status of all requests, and specific areas concerning new requests, requests in progress, failed requests, and requests with undeletable data.

6. Ex-post compliance assessment

BPR4GDPR has also focused on the ex-post compliance assessment because many information systems were already able to store event data before the GDPR took place. For that, data generated by such systems is confronted with models that represent the regulations in place to determine to what extent the latter were/are already followed, and to support decisions on what is missing and/or should be changed so that such regulations are now respected. In this context, BPR4GDPR advanced several existing Process Mining techniques to be applied over event logs.

In the Process Mining context, a trace represents all the events related to the same context (or case) ordered by its occurrence in time. Furthermore, compliance checking techniques of Process Mining precisely locate and localise the non-compliance on the level of individual traces with respect to a business rule, some behavioural constraint(s), or a regulation. What makes compliance checking more challenging is the fact that non-compliance can be in the form of excessive behaviour as well as missing behaviour, i.e., (i) what should have not happened but may have happened, and (ii) what should have happened but may have not happened. In addition to the existence or non-existence of some particular behaviour known as control-flow aspect of process behaviour, the

temporal aspect of the behaviour, i.e., when ideally or strictly something should have happened, is also important.

Many GDPR provisions like the Data Subject Rights can be translated to a control-flow behavioural requirement on the incumbent process. For instance, if a customer (data subject) makes a request to the data controller for deletion of personal data (under Article 17- right to be forgotten), the personal data shall be deleted and the controller shall provide information on action taken on the request to the data subject without undue delay and in any case within one month of receipt of the request (as per Article 12 - Transparent information, communication and modalities for the exercise of the rights of the data subject). Not adhering to this behaviour is considered as non-compliance, unless allowed by the exceptions mentioned in the Article 17. In business process context, the activity related to receipt of a customer data deletion request shall eventually be followed by an activity related to a notification to the customer regarding the fulfilment of his request, to be referred to as pattern in the rest of the section. Some other activities may possibly be executed in between these two pattern activities.

Compliance checking requires: (i) an event log, and (ii) a formal representation, which for control-flow compliance checking, is a special kind of process model termed as *rule model* (cf. Figure 8) with embedded required pattern. The activities other than the pattern activities are masked with Ω (label Omega) in the rule model, as they do not carry much information for the compliance.

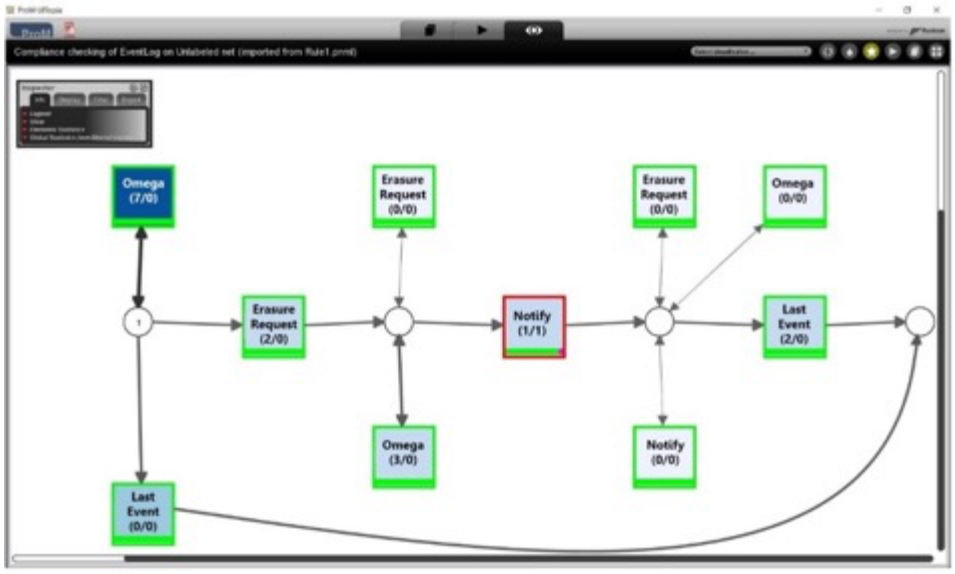

Figure 8: Compliance checking

The semantics of the rule model and trace replay are important to understand. In the background rule model in the Figure 8, place "1" (the circle with label 1) is the starting point of the rule model and contains a token so that the transitions Ω, Erasure Request, and Last Event are enabled. Any number of Ω activities can happen in the trace which are replayed on the Omega transitions in the rule model. If no pattern related activity is observed in the trace, then the case finishes with execution of the Last Event transition. If a pattern related activity is observed then the trace must fulfil the pattern requirement, i.e., there should be an activity related to erasure request by the data subject. With or without intermediate activities, i.e., Ω activities, an activity related to conveying erasure information to the data subject must happen, finally completing the case with Last Event transition. Further details

about this work together with a demo can be found in [32] and [33].

The GDPR along with necessitating specific response behaviour (pattern) in specific situations also impose bounds on the response time of this response behaviour. For temporal checking, the compliance checking process is divided into two phases. First, the control-flow compliance is checked as explained, followed by the second phase where the temporal constraints are checked. As with control-flow checking, the temporal checking constraints are also embedded as a pattern in a rule-model. The rule-model for temporal checking also contains Ω transitions and the transitions referring to the activities relevant to the control-flow pattern and *guards* on (some of) these transitions. The guards on the transitions provide the mechanism to constrain temporal aspects. The visualization in Figure 9 shows a combination of both the control-flow and the temporal check. We use different colours to show the different types of non-compliance. The trace on top has a temporal violation depicted in white – both the request for deletion and the action to inform the customer were done but not in the specified time –, while the second trace has a control-flow violation depicted in purple – the request for deletion took place but no action to inform the customer was registered in the data.

In addition to the example described here, other GDPR provisions have their own rule model and could serve to identify non-conformance. The diagnosed non-compliant behaviour may be analysed by the domain experts for devising remedial actions like changing or repairing the model, training the personnel, updating the policies etc.

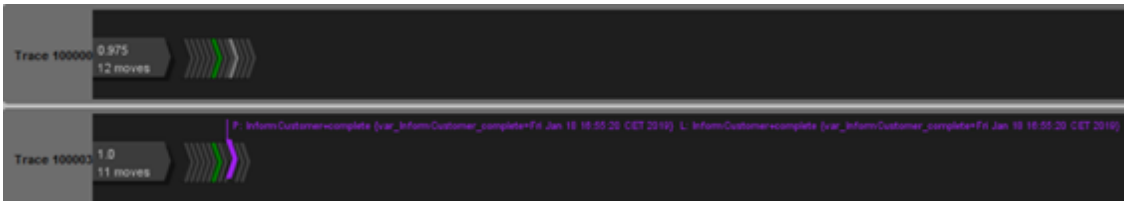

Figure 9: Combined result of control-flow and temporal compliance checking

Besides the presented contribution on compliance checking, other efforts have taken place in the BPR4GDPR context. To be able to detect more complex non-compliance, [34] and [35] propose a technique to represent control-flow, data, and privacy aspects all together as reference model. Such technique, implemented as a plug-in to the presented tool [36], is able to find non-compliance in one of these perspectives or resulting from a combination of them. Furthermore, to enable the compliance to data minimisation, we suggested several data forgetting policies [37] that guarantee a successful ending of the process. In [38], we further reduced the forgetting effect by imputing missing information in orphan events.

7. Assessment and conclusions

The assessment strategy of BPR4GDPR was an essential instrument to evaluate and measure quality and impact of the developed tools, to support a mutual understanding of end users' requirements and to facilitate continuous improvement. As BPR4GDPR is targeting different industries and customer types, we have based the evaluation on the operation improvement in the frame of three distinct, but replicable and typical trial scenarios, each spanning various use cases. The first trial case dealt with a governmental organisation in the healthcare and social security domains, collecting, storing, and processing very sensitive data leveraging own infrastructure. The second

trial concerned the needs of multiple stakeholders participating in a network of car dealerships, with complex, inter-company service provision chains; this implies also significant complexity as regards data protection and enforcement of data subjects' rights, considering the distribution of data and processing operations. Finally, the third trial case dealt with very small organisations typically covering their needs through cloud-provided systems and services. For the latter two cases, the provision of "Compliance-as-a-Service" has been an important requirement and challenge.

For each trial we have identified sets of operative business processes (use cases) affected by the BPR4GDPR regulation, broke them down to test cases and defined a set of meaningful key indicators (KPIs), both qualitative and quantitative, while some are exploitation-oriented, addressing commercialisation aspects. Each KPI has been mapped to project requirements, which reflect the maximum "wishlist" of potential BPR4GDPR users, regardless of their cost. Therefore, at the end of the evaluation, we were in the position to analyse how many of these requirements were addressed and how effective they have been tackled.

An extensive regulatory analysis, result by result, showed that BPR4GDPR delivered a tool portfolio supporting compliance with the crucial obligations set forth by the GDPR. Data subjects' rights, security, privacy by design and by default, accountability obligations can be improved by the use of BPR4GDPR tools. Even if not all tools directly process personal data, the data protection by design and by default approach of BPR4GDPR was evident.

Nevertheless, the tools provided by BPR4GDPR do not exclude the contribution of legal experts in evaluating the legal consequences of specific events, for instance the claim of data subjects, the proceeding before the Supervisory Authority, as well as the review of the data processing agreement to be concluded with a data processor. In such cases, BPR4GDPR is not sufficient per se, but may act as a helpful tool for the legal expert to collect the necessary information and easily reconstruct how the data flow is managed. BPR4GDPR has not be developed to replace the human intervention at all, but to serve data controllers and processors in organizing the processing of personal data in a consistent and compliant way, considering the main regulatory requirements and obligations to which they are subject.

During the trials' execution, we have collected interesting lessons learnt that might be of interest for any project implementing privacy with tool support. We summarise the most important lessons learnt below:

- **User experience**: the participants have highlighted the difficulties that a user may experience while interacting with the BPR4GDPR solutions. We have drawn useful conclusions, such as the need to have templates for easier modelling.
- **Applicability in GDPR compliance**: the difficulty to collect all information needed for modelling an organisation and its processes, as well as to define appropriate rules, proved to be a main blocking factor.
- **Multi-system environment**: the respondents of all trials highlighted the fact that multiple systems are composing their IT landscape and the need to take this into account to find the data of a particular data subject.
- **GDPR compliance**: the pilot partners expressed their insecurity, of not being fully compliant to the GDPR. Pilot partners have appreciated the support of the tools when assessing whether their company is really in compliance with the new regulation.

- **More automation**: for future developments the users wish more automation and less human action when performing GDPR related processes.

Concluding, what the trials have shown is that SMEs can have significant gains as regards GDPR compliance with the adoption of the tools provided by the BPR4GDPR. These tools provide solutions that concern a broad spectrum of needs and the corresponding operational stages of SMEs' practices and data processing activities: a) starting from a phase of planning, policies are defined, risks are assessed and process models are re-engineered to become compliant; b) continuing to run-time, tools provide for data management and protection, as well as for the enforcement of data subjects' rights; c) the operational loop closes with an ex post "reality check", where the evaluation and assessment of actual execution traces provides for verification of compliant business behaviour and the appropriate indications for improvement.

Acknowledgements

This research is being supported by the European Commission, in the frame of the H2020 BPR4GDPR project (Grant No. 787149). The authors would like to express their gratitude to the BPR4GDPR Consortium.

References

[1] European Commission, DG for Communication, "e-privacy", *Flash Eurobarometer 443*, December 2016.

[2] "H2020 BPR4GDPR", https://www.bpr4gdpr.eu/ [October 26, 2021]

[3] S. Spiekermann, L. Faith Cranor, "Engineering Privacy", *IEEE Transactions on Software Engineering*, Vol. 35, No. 1, pp. 67-82, January 2009.

[4] G. Kermezis, K. Limniotis, N. Kolokotronis, "User-Generated Pseudonyms Through Merkle Trees", in *Proceedings of the 9th Annual Privacy Forum (APF 2021)*, Oslo, Norway, June 17–18, 2021.

[5] J. Salas, V. Torra, "A General Algorithm for *k*-anonymity on Dynamic Databases", in *Proceedings of the 13th Data Privacy Management Workshop (DPM 2018)*, Barcelona, Spain, September 6–7, 2018.

[6] E. I. Papagiannakopoulou, et al, "Privacy-Aware Access Control", in *Encyclopedia of Information Science and Technology*, Mehdi Khosrow-Pour Ed., IGI Global, 2015, pp. 4403-4411.

[7] C. Dwork, "Differential Privacy: A Survey of Results", in *Proceedings of the 5th International Conference on Theory and Applications of Models of Computation (TAMC 2008)*, Xian, China, April 25-29, 2008.

[8] S.J. De, D. Le Métayer, "PRIAM: A Privacy Risk Analysis Methodology", in *Proceedings of the 11th Data Privacy Management Workshop (DPM 2016)*, Heraklion, Greece, September 26–27, 2016.

[9] H. Jiang, et al, "Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey", *ACM Computing Surveys*, Vol. 54, No. 1, January 2022.

[10] M. Joye, F. Petitcolas, "PINFER: Privacy-Preserving Inference - Logistic Regression, Support Vector Machines, and More, over Encrypted Data", in *Proceedings of the 14th Data Privacy Management Workshop (DPM 2019)*, Luxembourg, September 26-27, 2019.

[11] E. Politou, E. Alepis, C. Patsakis, "Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions", *Journal of Cybersecurity*, Vol. 4, No. 1, 2018.

[12] C. Wolter, A. Schaad, C. Meinel, "Task-based entailment constraints for basic workflow patterns", in *Proceedings of the 13th ACM Symposium on Access Control Models and Technologies (SACMAT '08)*, Estes Park, USA, June 11 - 13, 2008.

[13] M. Windrich, A. Speck, N. Gruschka, "Representing Data Protection Aspects in Process Models by Coloring", in *Proceedings of the 9th Annual Privacy Forum (APF 2021)*, Oslo, Norway, June 17–18, 2021.

[14] R.M. de Carvalho, et al, "Protecting Citizens' Personal Data and Privacy: Joint Effort from GDPR EU Cluster Research Projects", *SN Computer Science*, Vol. 1, No. 4., July 2020.

[15] International Association of Privacy Professionals (IAPP), "2021 Privacy Tech Vendor Report", https://iapp.org/media/pdf/resource_center/2021TechVendorReport.pdf, [October 26, 2021]

[16] G. Lioudakis, et al, "Facilitating GDPR Compliance: The H2020 BPR4GDPR Approach", in *Proceedings of the Workshop on Trust and Privacy Aspects of Smart Information Environments (TPSIE)*, September 18–20, 2019, Trondheim, Norway, IFIP AICT Vol 573, Springer Nature.

[17] G. Lioudakis, D. Cascone (eds.), "Compliance Ontology", *BPR4GDPR Deliverable D3.1*, February 2019.

[18] E. Papagiannakopoulou (ed.), "Final specification and prototyping of the policy framework", *BPR4GDPR Deliverable D3.3*, August 2020.

[19] E. I. Papagiannakopoulou, "Semantic Access Control Model for Distributed Environments", Ph.D. thesis, National Technical University of Athens, School of Electrical and Computer Engineering, Athens, 2014.

[20] Organization for the Advancement of Structured Information Standards (OASIS), "eXtensible Access Control Markup Language (XACML) Version 3.0.", OASIS Standard, January 2013.

[21] M. N. Koukovini, "Inherent privacy awareness in service-oriented architectures", Ph.D. thesis, National Technical University of Athens, School of Electrical and Computer Engineering, Athens, 2014.

[22] S. Jablonski, C. Bussler, *Workflow Management: Modeling, Concepts, Architecture and Implementation*, London: International Thomson Computer Press, 1996.

[23] R. A. Botha and J. H. P. Eloff, "Separation of duties for access control enforcement in workflow environments", *IBM Systems Journal*, Vol. 40, No. 3, pp. 666-682, 2001.

[24] R.M. de Carvalho, G. Lioudakis (eds.), "Final specification and prototyping of the process re-engineering framework", *BPR4GDPR Deliverable D4.2*, November 2020.

[25] L. Bracciale (ed.), "Final specification and prototyping of the compliance toolkit", *BPR4GDPR Deliverable D5.2*, November 2020.

[26] "Apache NiFi", https://nifi.apache.org/ [October 26, 2021]

[27] A. Singh and C. Kakali, "Cloud security issues and challenges: A survey", *Journal of Network and Computer Applications*, Vol. 79, pp.88-115, 2017.

[28] L. Bracciale, P. Loreti, E. Raso, M. Naldi, G. Bianchi, "CoProtect: Collaborative Management of Cryptographic Keys for Data Security in Cloud Systems", in *Proceedings of the 6th International Conference on Information Systems Security and Privacy (ICISSP 2020)*, Valletta, Malta, February 25-27, 2020.

[29] E. Raso, L. Bracciale, P. Loreti, G. Bianchi, "ABEBox: A data driven access control for securing public cloud storage with efficient key revocation", in *Proceedings of the 16th International Conference on Availability, Reliability and Security (ARES 2021)*, Vienna, Austria, August 17-20, 2021.

[30] P. Russo, L. Bracciale, G. Bianchi, "Dare-to-Share: Collaborative privacy-preserving recommendations with (almost) no crypto", *Security and Privacy*, Vol. 4, No. 3, May-June 2021.

[31] A. Narayanan and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets", in *Proceedings of the 2008 IEEE Symposium on Security and Privacy (SP 2008)*, Oakland, California, USA, May 18-21, 2008.

[32] R. Zaman, A. Cuzzocrea, M. Hassani: "An Innovative Online Process Mining Framework for Supporting Incremental GDPR Compliance of Business Processes." *IEEE BigData 2019: 2982-2991*

[33] R. Zaman, M. Hassani, "On Enabling GDPR Compliance in Business Processes Through Data-Driven Solutions", *SN Computer Science*, Vol. 1, No. 4., July 2020.

[34] A.S. Mozafari Mehr, "Compliance to data protection and purpose control using process mining technique" in *Proceedings of the 17th International Conference on Business Process Management (BPM 2019)*, September 1-6, 2019, Vienna, Austria.

[35] A.S. Mozafari Mehr, R.M. de Carvalho, B. van Dongen, "Detecting Privacy, Data and Control-Flow Deviations in Business Processes", in *CAiSE Forum 2021*, Melbourne, Australia, June 28 – July 2, 2021.

[36] A.S. Mozafari Mehr, R.M. de Carvalho, B. van Dongen, "MLA: A Tool for Multi-Perspective Conformance Checking of Business Processes". 3rd International Conference on Process Mining – Demo (to appear).

[37] R. Zaman, M. Hassani, B.F. van Dongen: "Data Minimisation as Privacy and Trust Instrument" in Business Processes. Business Process Management Workshops 2020: 17-29

[38] R. Zaman, M. Hassani and B.F. Van Dongen: "Prefix Imputation of Orphan Events in Event Stream Processing". *In Frontiers in Big Data Vol. 4, 2021.*