Accomplishing Transparency within the General Data Protection Regulation

Dayana Spagnuelo[1], Ana Ferreira[2] and Gabriele Lenzini[1]
[1]*SnT - University of Luxembourg, Luxembourg*
[2]*CINTESIS - University of Porto, Portugal*
dayspagnuelo@gmail.com, amlaf@med.up.pt, gabriele.lenzini@uni.lu

Keywords: Transparency, Transparency Enhancing Tools, General Data Protection Regulation, Compliance.

Abstract: Transparency is a user-centric principle proposed to empower users to hold data processors accountable for the usage and the processing of the user's personal data. Accomplishing transparency may come with some resistance because it requires significant architectural changes, but it is mandatory by law under the recently approved General Data Protection Regulation. To help the transition, we systematically review what Transparency Enhancing Technologies can help to accomplish transparency in agreement with technical requirements that we elicited from the Regulation's articles. We discuss our findings in the domain of medical data systems, where accomplishing transparency looks particularly controversial due to sensitivity of the personal medical data.

1 INTRODUCTION

The General Data Protection Regulation (GDPR) is now entirely in force and data processors/controllers need to ensure that processing is *lawful*, *fair* and *transparent*[1]. While the principles of lawfulness and of fairness express legalistic concepts, transparency suggests as a *socio-technical* concept: it should be realised as a technical feature whenever appropriate (Article 29 Working Party, 2018, see paragraphs: 4, 7) but is meant to help data subjects to know how their data is processed and whether this is done lawfully and fairly.

The interest in transparency has grown since the principle appeared in early drafts of the GDPR: it has been discussed as a principle of accountability in the cloud computing domain (Berthold et al., 2013), presented as a privacy goal and precondition for intervenability (Meis and Heisel, 2017), and studied for its meaning in the area of electronic medical systems (Spagnuelo and Lenzini, 2016).

Concomitantly, several Transparency Enhancing Tools (TETs) — system-independent tools intended to help individuals to gain more knowledge about their data— have been proposed; still it is unclear whether they can be adopted to inform users about the lawfulness and fairness of the processing of their data or, from a different perspective, to improve a system's transparency according to the GDPR.

Deciding whether a tool gives a presumption of compliance with a GDPR's principle is an open and hindering problem because the Regulation's provisions are broadly defined and admit several interpretations, but one could attempt the task by leveraging on the existing literature.

We look for correlation between technical requirements to implement transparency (Spagnuelo and Lenzini, 2016) that a few recently proposed TETs realise and the GDPR's articles about transparency. In so doing, we identify the GDPR concepts still in need of more development, and discuss what TETs help, or could help if implemented in a certain way, accomplish transparency. We focus our research on the domain of electronic medical data systems, where highly sensitive personal data are processed; our conclusions though hold in other domains where the personal data are likely to be of a less sensitive nature.

2 RELATED WORKS

A few works attempt to achieve or discuss compliance with the GDPR's principle of transparency by deriving technical requirements from the Regulation's articles. Meis and Heisel (Meis and Heisel, 2017) do so for intervenability i.e., empowering end-users

[1]GDPR, Article 5.1.(a).

to have control over their personal data processing; they extract requirements by reviewing the ISO/IEC 291000 and the GDPR's article about privacy and transparency. No direct correlation between requirements and the Regulation's provision is made.

Bier *et al.* (Bier et al., 2016) derive eight technical requirements for a privacy dashboard they propose directly from a review of the 'right of access' presented by the GDPR, the previous European Data Protection Directive, and the Federal Data Protection Act from Germany. Raschke *et al.* (Raschke et al., 2017), for a similar dashboard, define requirements about four high-level GDPR-related features of the board: the right to access data, obtaining information about involved processors, rectification and erasure of data, and consent review and withdraw.

More in line with our research, Fischer-Hüber *et al.* (Fischer-Hübner et al., 2014) compare and map legal provisions and technical requirements, principles, and designs. The authors review usability principles of Human-Computer Interaction (HCI) in a few selected TETs. They gather requirements from workshops and by reviewing the proposal of the GDPR, the previous European Data Protection Directives, and other documents, e.g., opinions from the Article 29 Data Protection Working Party; they also consider legal provisions for transparency and accountability that have implications about HCI. The requirements are then mapped to three HCI concepts, further discussed in the context of the TETs. The mappings and correlations presented are thoroughly discussed, but the authors do not present a structured procedure that was followed when defining them: it is our interpretation that those correlations were identified manually.

At the time of execution our research we were unaware of the German Standard Data Protection Model (SDM)[2], which classifies GDPR's provision in terms of *data protection goals* (e.g., availability, transparency, intervenability). We will discuss the similarities between our work and the SDM, and clarify where they differ.

3 TRANSPARENCY IN GDPR

Transparency is a property championed by the GDPR as one of the main principles in personal data processing. It qualifies transparency stating that it "requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used"[3]. But how transparency is characterised? Some help to this regard comes from Article 29 of the Data Protection Working Party. It provides interpretative guidelines regarding transparency as enforced by the GDPR (Article 29 Working Party, 2018). But, it gives no characterisation of transparency. For that, one must review the Articles of the Regulation that refer to the principle, even if indirectly. We selected those Articles by following a systematic peer review approach in four rounds: selection; filtering; revision; and validation.

Selection. Two of this paper's authors had, independently, selected a list of the GDPR's Articles that they subjectively judged be talking about transparency. Both authors are expert in transparency and in TETs, so the expectation was that from their combined knowledge it was possible to identify most of the Articles that link to transparency in different technical domains.

Filtering. We combined the two lists resulting from the previous phase and revisited the Articles selected by at least one of the authors. The two authors defended their interpretation of transparency, agreed on a common understanding, and defined Articles that cover that understanding, including also artefacts to implement transparency. The intersection of the two preliminary lists, complemented by a few other Articles, was taken as the output.

Revision. Two authors, one not involved in the previous rounds of the analysis to reduce selection bias, independently reviewed the guidelines by the Working Party. They selected the Articles that, according to the guidelines and to their judgement were about transparency. The resulting lists —almost identical since the guidelines are less prone to interpretation than the GDPR— were compared and combined.

Validation. We compare the lists obtained in the previous rounds, with the aim to validate our selection using less subjective indications coming from the guidelines. Our final list of selected transparency-related GDPR Articles (paragraphs and sub-paragraphs) comprises 79 items (see Table 3). It only disregards four Articles mentioned in the guidelines (e.g., 12.5, 20, 25.1 and 25.2). In addition, we also compare our list with the presented by the SDM regarding the protection goals of transparency and intervenability (the ones we consider in our work as well). Our selection is more extensive, but disregards five Articles mentioned in the SDM (i.e., 5.1.(d), 5.1.(f), 20, 40 and 42). We consider our list sufficiently relevant.

[2] https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology_V1.0.pdf

[3] *Ibid.* (39)

Table 1: Transparency requirements. IDs refer to the original numbering in (Spagnuelo and Lenzini, 2016), those indexed 1** are ex ante, those 2** are ex post.

Req.	Specification
111.1	The system must provide the user with real time information on physical data storage and data storage location of different types of data.
111.2	The system must inform the user on how data are stored and who has access to them.
111.18	The system must make available a document that describes the ownership of the data.
112.1	The system must provide the user with mechanisms for accessing personal data.
211.3	The system must notify the user in case the policy is overridden (break the glass).
211.4	The system must provide the user with timely notification on security breaches.

4 GDPR AND TECHNICAL REQUIREMENTS

We correlate the selected GDPR's Articles with a list of technical requirements for transparency presented in previous work from the authors (Spagnuelo and Lenzini, 2016). Due to space limitations, in Table 1 we recall just a few requirements to help the reader picture how they look like, but we remand to the original work for full details.

The correlation is done automatically using a simplified parser based on Natural Language Processing (NLP) techniques. We analyse the *text corpora* in order to extract *corpus*-based glossaries. We did not conduct any statistical analysis, nor part-of-speech tagging (techniques applied in more sophisticated NLP algorithms). Instead, we iterated a few times realising small adjustments in our glossaries, reevaluating the results of the parser and, whenever needed, manually adding or removing a correlation.

Our approach is possible because our glossaries are context-based, and focused to the terminology found in the GDPR and in the requirements. There are works that propose interpreting and translating regulations and other legal documents in general (Bartolini et al., 2016; Sathyendra et al., 2017; Nejad et al., 2017). We do not mean to compete with them, but rather state that our parser, in the specific problem herein addressed, has given sufficiently accurate results.

Text Corpora Analysis. The first step was carried out manually. We first analysed the two text *corpora*: the Articles and provisions in the GDPR, and our set of technical requirements. A text *corpus* is described as a "large body of linguistic evidence typically composed of attested language use", but has been used nowadays for defining a variety of text collections (Mitkov, 2005). Our requirements are not a text *corpus* in its typical meaning, as they are based on the literature, and are not composed by standardised terms. Rather, they constitute a text *corpus* in its modern sense: a text collection tailored to a specific domain. The GDPR, on the other hand, represents better a classic text *corpus*, as it is stable and composed by standard legal terminology.

We analysed the text *corpora* and familiarised with the differences between the terminologies, as one comprises technical terms and the other legalistic jargon. The terms found in one *corpus* were interpreted and linked to terms in the other. As a result of this task, we elected potential connections between requirements and GDPR Articles and established a preliminary list of correlations.

Extraction of Corpus-based Glossaries and Parsing. To ensure the consistency of our correlation procedure, we automated the comparisons. Terms found in the GDPR were mapped to their equivalent technical terms, found in the list of requirements. To do that, we revisited our preliminary list of correlations, from where we extracted the key-terms that seem to have triggered each correlation. We identified correlations according to a few textual elements present in the GDPR Articles: the *information* to be provided to the data subject; the *rights* the data subject must have; the *techniques* described in the Article; and few selected *keywords*. We organised each of these in hash tables that represent *corpus*-based glossaries. Due to space limitations, we show in Table 2 only one of our glossaries.

Some key-terms were intentionally marked as *not applicable* (N/A). For instance, the term "transparency" found in Article 5.1(a) is comprehensive and should relate to every single requirement from our list, as it mandates data to be processed transparently. The same applies to the term "shall not apply", which is present in Articles describing an exception to another Article. In other words, it presents the circumstances in which our requirements do not need to be implemented. Hence, correlations found with an Article of this sort are likely to be false-positives. It is important to note that terms marked like this are not the same as terms absent from our glossaries. While the first will force a mismatch between a GDPR Article with that term and any possible requirement in our list, the second will just be disregarded when computing the correlations.

The correlations are computed by an automatic parser. Initially, it parses each GDPR Article to iden-

Table 2: Glossary of Techniques. Information in brackets are contextual and are not part of the term.

GDPR terms	Technical terms
[do not] permit identification	data privacy; to protect [data]; [data] protection; [data is] protected; separation [of data]
appropriate security	to protect
withdraw	revoke
not in a position to identify	N/A
automated decision-making	N/A
obtaining [personal data]	gather; infer; aggregate
copy of personal data	mechanism for accessing [personal data]
automated means	N/A
only personal data which are necessary	data minimisation
record of [processing of data]	accountability; audit
unauthorised	without authorisation
unlawful	vulnerability; breach
accidental loss	data loss; breach
accidental destruction	N/A
accidental damage	N/A
profiling	N/A
data minimisation	N/A
existence of the right	ownership
shall not apply	N/A

tify all the key-terms they contain. Then it searches for requirements which present at least one equivalent term for each key-term found in the Article. Our criteria to match an Article and a requirement is that all key-terms from the first are represented in the second.

The correlation computation is realised in steps: we run the same parsing algorithm for each glossary, and later we merge the results of each comparison in one final list. By doing so, we maintained the correlating criterion decoupled, and simplified the process of reevaluation of the equivalent terms. This has also helped in balancing the asymmetry between GDPR Articles and technical requirements, as the Articles are generally more verbose and encompass too many key-terms. Separating the terms into four glossaries ensured our criterion is not too restrictive, and that Articles can be correlated by one or several categories of textual elements.

Final Adjustments. After computing the correlations, we reviewed the resulting list and compared with our preliminary one. Each correlation was analysed, but we focused on the discrepancies. For those, we semantically analysed the article and requirement marked as correlating to understand the context in which the key-terms appeared, and whether they had similar meaning. We realised this procedure in a peer review manner. The final correlations (see Table 3) were then adjusted accordingly. We highlight here a few of the manually adjusted correlations.

According to our initial list, requirement 111.2 on information about how data are stored and who has access to them, should correlate with Article 15.1.(c), about the data subject's rights of obtaining from the controller the recipients of personal data. The requirement and the Article have a clear correlation. However, it was being disregarded by our parser as the Article contains the key-term "third countries" which does not appear in the requirement. As this key-term is responsible for several other well-fitted correlations, we opted for adjusting this exception manually.

We also adjusted Articles and requirements that were marked as correlating by key-terms but had different meanings. Such as requirement 111.3, mandating the system informs users on the purchase of services, and Article 19, which requires the controller to communicate the erasure or rectification of personal data to its recipients.

5 TRANSPARENCY ENHANCING TOOLS

We conducted a literature review looking for scientific works or projects indexed by the keywords "transparency enhancing tools". We restricted our search to works published since 2014, the year the GDPR started to be strongly supported by the European Parliament[4]. By adding this time restriction, we consider only tools potentially designed in line with the GDPR principles. We then broadened our study with works cited in our initial pool. A few works surveying TETs helped us defining a list of tools for our study (Ferreira and Lenzini, 2015; OPC, 2017; Bier et al., 2016; Zimmermann, 2015; Siljee, 2015).

To select the relevant tools we classified them according to the categories proposed in (Zimmermann, 2015). This categorisation, TETCat, takes into account whether the information provided by these tools is trusted and can be verified (*assurance level*), the *application time* of the tool (*ex ante*, *ex post* or *real time*) and *interactivity level* the tools offer. As a result of this exercise, we found 27 tools that are potentially linked to the transparency principle. We present them classified by their TETs category. We only describe

[4]http://europa.eu/rapid/press-release_MEMO-14-186_de.htm

Table 3: Final list of correlated GDPR Articles and technical requirements. 72% of the requirements are correlated (26 out of 36).

GDPR	Requirements	GDPR	Requirements	GDPR	Requirements
5.1.(a)		14.1.(c)	111.19	18	
5.2	111.16, 111.20, 221.1, 221.2, 221.3, 221.4, 221.5, 221.7, 221.8	14.1.(d)	221.6	19	111.2, 111.4
6.1.(a)	221.7	14.1.(e)	111.2, 111.3, 111.4	21.1	
7.1		14.1.(f)	111.4, 11.11, 221.3	21.2	
7.2		14.2.(a)		21.3	
7.3	221.7	14.2.(b)	111.3, 111.4, 111.14	21.4	111.18
9.2.(a)		14.2.(c)	111.18	21.5	
11.2		14.2.(d)	111.18	22.1	
12.1		14.2.(e)		22.2.(c)	
12.3		14.2.(f)	221.6	25.3	
12.4		14.2.(g)		26.1	111.14
12.7		14.3.(a)	211.5	26.2	111.14
13.1.(a)	111.1	14.3.(b)		26.3	111.14
13.1.(b)	111.15	14.3.(c)		30.1	221.5, 222.1, 232.1
13.1.(c)	111.19	14.4		30.2	221.5, 222.1, 232.1
13.1.(d)	111.3, 111.4, 111.14	15.1.(a)	111.19	30.3	
13.1.(e)	111.2, 111.3, 111.4	15.1.(b)	221.6	30.4	
13.1.(f)	111.4, 111.11, 221.3	15.1.(c)	111.2, 111.4	32.3	
13.2.(a)		15.1.(d)		33.1	111.7, 211.1, 211.4, 221.8
13.2.(b)	111.18	15.1.(e)	111.18	33.2	111.7, 211.1, 211.4, 221.8
13.2.(c)	111.18	15.1.(f)		33.3	111.7, 111.15, 211.1, 211.4, 221.8
13.2.(d)		15.1.(g)	221.6	33.4	211.4
13.2.(e)		15.1.(h)		33.5	111.7, 211.1, 211.4, 221.8
13.2.(f)		15.2	111.4, 111.11, 221.3	34.1	111.7, 211.1, 211.4, 221.8
13.3		15.3	112.1	34.2	
14.1.(a)	111.1	16			
14.1.(b)	111.15	17	221.7		

the characteristics needed for the understanding of this work. The full categorisation is made available in (Spagnuelo et al., 2018).

Assertion Tools. Tools are classified as the assertion type whenever the correctness and completeness of the information they provide cannot be verified. They can only provide users with information on the controller's alleged processing practices. The TETCat does not further distinguish between assertion tools as their trustworthiness remains unaffected even under different manifestation of other parameters. As a consequence, this category covers tools with diverse goals.

Examples of assertion tools are third-party tracking blockers. These tools are commonly implemented as web-browser plug-ins. They help the users becoming aware of trackers gathering information about them while browsing the web, for the purpose of, e.g., advertising or analytic. These tools also allow the users to interact and block such trackers. Mozilla Lightbeam[5] (ML), Disconnect me[6] (DM), and Privacy Badger[7] (PB) are examples of tracking blocker tools.

Those tools offer very similar features to the users, with the exception of Privacy Badger (PB). When this tool detects a new third-party script is tracking the user in three different websites, it automatically blocks them, without the need for the users to configure their preferences.

Tools that educate users on matters related to privacy protection are also considered assertion tools. One example of such a tool is the Privacy Risk Analysis (PRA) (De and Le Métayer, 2018), which allows users to express their preferences, and visualise the impact on privacy risks through a user-friendly interface. Another example is Me and My Shadow[8] (MMS) which informs users about what happens to their data, how traceable they are on the internet and gives tips about existing privacy tools and digital shadow.

Privacy Score[9] (PS) is also classified as an assertion tool. It tests and ranks websites according to their

[5]https://www.mozilla.org/lightbeam

[6]https://disconnect.me/

[7]https://www.eff.org/privacybadger

[8]https://myshadow.org/

[9]https://privacyscore.org/

security features (tracking, encryption of traffic and messages, protection against attacks). For each feature tested it also presents brief explanations which serve as material to educate users on privacy protection subjects.

Finally, Access My Info[10] (AMI) also falls under the assertion category. AMI is a web application that helps users to create legal requests for copies of their data. This tool differs from the previous ones as it does not *per se* provide information nor educates the users on the service's practices. Yet, we consider it under the same category as it cannot ensure the service providers will properly process the requests. The tool guides users in requesting data from dating, fitness, and telecommunications services, according to the Canadian privacy legislation (PIPEDA). Despite being a web application, it is implemented as a script running on the users' browser, and it collects no information unless explicitly authorised by them.

Awareness Tools. This is the first type of tools providing information verifiable for completeness and correctness. Different terminology is suggested for tools which provide technical means to verify its information (i.e., *Trusted*), and tools which information can be verified manually by a user or an auditor (i.e., *Semi Trusted*). However, for the TETCat, they do not distinguish between the two assurance levels. Similarly, we refrain from evaluating this aspect of the tools. We only distinguish between *Not trusted* and *(Semi) Trusted*, being the last given to tools that provide somewhat trustworthy information.

Awareness tools provide *Ex ante* transparency, and interactivity level of *Read only*. Tools in this category help the user becoming aware of the privacy policy of the service provider but do not provide the users with controls over the processing of data. Examples of such tools are machine readable or interpreted policy languages and certification seals and marks.

Platform for Privacy Preferences Project[11] (P3P) is an example of machine-readable language tool. It proposes a language for describing a website's privacy policy in a standard format which can be retrieved and interpreted automatically by web-browsers. It enables the users to be informed of the website intentions towards the use and collection of their data in a consistent way, without requiring them to read the entire privacy policy of each website they visit. Even though works on P3P are currently suspended, we include it in our study as it has strong support from the academic community.

[10]https://openeffect.ca/access-my-info/

[11]https://www.w3.org/P3P/

On a different approach, the Usable Privacy Project[12] (Sathyendra et al., 2017) proposes a tool that automatically annotates privacy policies. The tool eases the reading of policies by interpreting it and highlighting parts of the text according to a fine-grained annotation scheme.

Other examples of awareness tools are the certification seals. European Privacy Seal (EuroPriSe), for example, provides a privacy compliance certification of IT products and services with European data protection regulations (EuroPriSe, 2017). Another example is the TrustArc (TArc), which provides a trust mark on privacy practices and data governance. It follows certification standards based upon recognised laws and regulatory standards, such as OECD Privacy Guidelines, and GDPR (TrustArc, 2018).

Declaration Tools. These tools are very similar to awareness tools, but they offer some level of interactivity. In our pool only one tool falls under this category: PrimeLife Policy Language (PPL) (Fischer-Hübner and Martucci, 2014). With this tool, users can interact and negotiate policies.

PPL is comparable to the awareness tool P3P as it proposes a machine-readable language for privacy policies. However, PPL further supports the description of privacy preferences from the users. So the service provider's declared intentions can be matched and checked for compliance with the user's preferences. PPL expresses policy in terms of authorisations, e.g., for what purposes the service provider will use the data, and obligations it is willing to fulfil for collected data items (e.g., to delete the data after a certain period, or to log all accesses to the data).

Audit Tools. Audit TETs present users with *Ex post* or *Real time* transparency. Tools in this category include those that allow for access and verifiability of data, but do not provide means for the users to interact and intervene with the data processing (i.e., *Read only* tools).

Data Track[13] (DT) (Fischer-Hübner et al., 2016) is a user side ex post transparency tool that displays what personal data the service provider has stored, which was received from the user explicitly, implicitly, or derived. The tool is a proof-of-concept that parses location history from Google take-out. Personal Data Table (PDT) (Siljee, 2015) is a similar tool, however in an earlier stage of maturity. PDT is a transparency design pattern. It describes a standardised table containing information on personal data

[12]https://explore.usableprivacy.org/

[13]https://github.com/pylls/datatrack

handled by the service provider, such as, the reasons for collection, and who has access to it.

Digi.me[14] is an application for retrieving a copy of personal data from several different services (e.g., social media, finance, and health). The application does not store any personal data, it copies the data into the storage of the user's preference. By using this tool the user can visualise, search, and choose to share these data with other apps. Even though Digi.me allows the users to share, and consequently to control the collection and usage of personal data, it is not considered interactive. That is because it does not provide means for the users to control processing from the source where the data is retrieved.

Finally, the Blue Button[15] is an initiative to standardise the right to access personal medical data in the USA. Blue Button-enabled portals display a logo, which symbolises that users are allowed to visualise and download their data.

On the verifiability side, there are discussions regarding transparency, but tools of this type are still in the idealisation phase. Privacy Evidence (PEv) (Sackmann et al., 2006), for example, proposes the generation of pieces of evidence based on structured policies (P3P and NAPS), secure logs (with hash chain scheme to guarantee confidentiality and integrity, for instance), and logs view that allow to scan through the logs and match with the policy. Transparent Accountable Data Mining (TAMI) (Weitzner et al., 2008) similarly proposes *a posteriori* privacy compliance checks on data mining. It is intended to check for data usage that is logically allowed to happen, but that legally should not be used in support to a given conclusion (inference). The proposed architecture is composed of a policy-aware logs module, a policy language framework, and a policy reasoning tool. Both tools are thoroughly discussed, but we found no implementation of them.

Finally, Private Verification of Access (PVA) (Idalino et al., 2017) also proposes a scheme for *a posteriori* access control compliance checks, but that operates under a data minimisation principle. The scheme suggests the use of a third-party tool which can operate on encrypted data access logs to check for matches (or mismatches) with a given access policy. The tool allows for a private independent audit of a system.

Intervention Tools. Tools in this category allow users to verify properties about the processing of their data. They differ from audit tools as they also provide means for them to interact and control the terms of data collection and usage.

Privacy Through Transparency (PTT) (Seneviratne and Kagal, 2014), for example, proposes the use of a Provenance Tracker Network (PTN) which stores the logs for any transaction realised in a personal data flagged as *sensitive*, and allows for *a posteriori* audits. The logs are distributed in a network of trusted peers, preventing a single point of failure. Every sensitive data has a usage restriction associated with it, and every use of data needs to be justified by a *usage intention*. This tool allows data owners to analyse logs and search for mismatches between the usage restrictions and intentions. They are allowed to request for explanations in case mismatches are found. The model assumes a non-prohibitive access control mechanism and supports Break-the-Glass (BTG) policies.

Privacy eSuite[16] (PeS) is a web-service consent engine that centralises consent and access rules with support to purpose of use. This tool also supports the integration with other services, such as the myConsentMinder, a web application that allows patients to manage their privacy preferences, and the Universal Audit Repository, that logs all accesses and attempts, notifies when a BTG happens, and simplifies audits through searches and report capabilities.

Remediation Tools. These are the most comprehensive tools according to the TETCat. They comprise functionality to exercise control over data collection and usage, and also to modify and delete personal data stored by a data controller. Tools in this category are usually found in the format of privacy dashboards or data vault/marketplace applications.

PrivacyInsight (PI) (Bier et al., 2016) and GDPR Privacy Dashboard[17] (GPD) (Raschke et al., 2017) are both examples of privacy dashboards within this category of TETs. PrivacyInsight is a dashboard whose main feature is to visualise (as a provenance graph) the flow of personal data into, through and out of an organisation. PrivacyInsight also provides full access to all personal data and allows users to exercise their rights over that data (e.g., erasure, and rectification). The GDPR Privacy Dashboard is intended to help users visualising, and requesting rectification or erasure of the data stored by a service provider. Both tools are designed to be easily adopted by any organisation.

Google Dashboard[18] (GD), and Microsoft Dash-

[14] https://digi.me/

[15] https://www.healthit.gov/topic/health-it-initiatives/blue-button

[16] http://hipaat.com/privacy-esuite/

[17] http://philip-raschke.github.io/GDPR-privacy-dashboard

[18] https://myaccount.google.com/dashboard

board[19] (MD) are also examples of such tools, however they serve only their own organisation. Both tools allow the users to manage privacy settings, and to see, download, and manage personal data stored in their account.

Finally, the openPDS (oPDS) (de Montjoye et al., 2014), and Meeco[20] (Mee) are examples of data vault/marketplace applications. The openPDS tool is a meta-data storage. Combined with SafeAnswers it allows for the privacy of the users by computing answers on the client side, and only sending third-party applications anonymous results. The results can also be aggregated with the ones from other users. Meeco, on the other hand, is a personal data marketplace which allows users to add, organise, edit, and progressively share their information. The Meeco client stores the terms the user agreed to, and records events of interaction with personal data in an event chain.

6 DISCUSSION

To have a general picture of transparency's development, we compared the selected TETs with our requirements. Doing so enabled us to understand the extent in which TETs can realise transparency in medical systems, and also to have, by transitivity, a list of TETs which can help to achieve compliance with the GDPR's provisions.

The TETs categorisation facilitated the comparison between the tools and the list of requirements. Mainly, we pre-selected the tools and requirements by their application time (i.e., *ex ante*, *ex post*/*real time*) and matched them manually according to the other categories, and descriptions whenever needed. The result of this effort is shown in Table 4. Exceptions were made only in two specific cases: regarding requirement 112.1, and when comparing certification seals tools.

The first is concerning requirement 112.1 on the provision of mechanisms for accessing personal data. In the context of medical systems, this requirement is considered *ex ante* as the data about the patients are typically generated by other users in the system, rather than being provided by the patients themselves. As a consequence, allowing these patients to access their data can be interpreted as a mandatory pre-condition for them to anticipate what will happen to their data. However, in the context of TETs, tools which allow for the access of personal data are considered *ex post*. In this specific case, we understand there is a close correlation between the requirement 112.1 and those tools, even if their application times do not match.

The second exception is regarding certification seals. We consider them *ex ante*, but admit a correlation between them and *ex post* requirements. Certification seals are tools which can serve as convincing evidence that a system complies with a given criterion. If the criteria regards the processing of data, these seals can help users anticipating what will happen to their data, and whether it will be processed in a secure manner. However, from the perspective of the system when evaluated for the certification, the processing of data is already happening. For this reason, we accept the correlation between ex ante certification tools and a few relevant ex post requirements.

We determine which tools can help to achieve compliance with the GDPR's provisions by transitivity: for each tool matched to a given requirement, we set that this tool is also closely linked to the GDPR Articles that given requirement matches (see Table 3). This exercise highlighted, for example, the transparency aspects which are not yet covered by TETs. Due to space constrains, in Table 4 we summarised the results of this comparison. We make available a full report where we expand the GDPR Articles relevant to this work (Spagnuelo et al., 2018). In what follows, we comment on our findings concerning the technical and legal aspects of transparency.

Technical Aspects. Three requirements regarding terms and conditions seem not to be addressed by any TET: 111.1 on information regarding the physical location where data is stored; 111.4 on the existence of third-party services and sub-providers; and 111.14 on clarifications of responsibility in case of the existence of third-party services. We believe this information could be provided together with the terms and conditions of service. Even though the tool provided by the Usable Privacy Project (UP) aims at facilitating the reading of information provided in the terms and conditions, we did not identify tags for the requirements above. For this reason, we do not consider these requirements as addressed.

There are other relevant developments on the reading of terms and conditions, and policies, such as the CLAUDETTE project[21], which uses artificial intelligence to automatically evaluate clauses of policies for their clarity and completeness in the light of the GDPR provisions. Another relevant functionality in this regard is the Lost in Small Print[22] from Me and

[19] https://account.microsoft.com/account/privacy

[20] https://www.meeco.me/

[21] https://claudette.eui.eu/

[22] https://myshadow.org/lost-in-small-print

Table 4: Transparency Enhancing Tools (TETs), the technical requirements and GDPR Articles they help realising (* added manually). Articles not addressed by TETs: 11, 12, 16, 18, 22, 25, 26, 32, 34.

TET	Requirements	GDPR Articles
Mozilla Lightbeam	211.5, 221.6	14, 15
P3P	111.2, 111.3, 111.16, 111.18, 111.19	5, 13, 14, 15, 19, 21
PrimeLife Policy Language	111.2, 111.3, 111.16, 111.18, 111.19	5, 13, 14, 15, 19, 21
Data Track	112.1, 221.5, 221.6, 221.7	5, 6, 7, 14, 15, 17, 30
Privacy Insight	112.1, 221.4, 221.5, 221.6, 221.7	5, 6, 7, 14, 15, 17, 30
Privacy Risk Analysis	111.9, 111.13	
GDPR Privacy Dashboard	112.1, 211.5, 221.4, 221.6, 221.7	5, 6, 7, 14, 15, 17
Personal Data Table	112.1, 211.2, 211.3, 211.5, 221.4, 221.6, 221.7	5, 6, 7, 14, 15, 17
Disconnect me	211.5, 221.6	14, 15
Me and My Shadow	111.8, 111.13, 111.16, 111.19	5, 13, 14, 15
EuroPriSe	111.16, 221.1, 221.3, 221.4	5, 13, 14, 15
Privacy Score	111.6, 111.12, 111.13	
Google Dashboard	112.1, 211.5, 221.6, 221.7	5, 6, 7, 14, 15, 17
Privacy Evidence	221.1, 221.4, 221.5, 222.1, 232.1	5, 30
TAMI Project	211.2, 211.3, 211.5, 221.1, 221.4, 222.1, 232.1	5, 14, 30
Privacy Through Transparency	211.2, 211.3, 221.1, 221.4, 221.5, 222.1, 232.1	5, 30
Private Verification of Access	211.2, 211.3, 221.1, 221.4, 222.1, 232.1	5, 30
Privacy Badger	211.5, 221.6	14, 15
Access My Info	112.1, 221.6	14, 15
TrustArc	111.16, 221.1, 221.3, 221.4	5, 13, 14, 15
openPDS	211.5, 221.6, 221.7	5, 6, 7, 14, 15, 17
Digi.me	221.6, 221.7	5, 6, 7, 14, 15, 17
Microsoft Dashboard	112.1, 211.5, 221.6, 221.7	5, 6, 7, 14, 15, 17
Privacy eSuite	221.1, 221.5, 221.7, 222.1, 232.1	5, 6, 7, 9*, 17, 30
Meeco	221.6, 221.7	5, 6, 7, 14, 15, 17
Blue Button	112.1, 221.6	14, 15
Usable Privacy	111.5, 111.10, 111.11, 111.15, 111.17, 111.19	13, 14, 15, 33

My Shadow (MMS), which reveals and highlights relevant information in the policy of a few popular services. We decided not to include those tools in our study as the first only evaluates the quality of a policy, without aiding data subjects understanding its contents, and the second for only providing a few examples of policies. Nevertheless, it is possible to see the matter is already subject of attention. We expect to see a different scenario concerning tools for terms and conditions in the future.

Another set of requirements which seem to have gained less attention is regarding security breaches and attacks. They constitute the majority of requirements not addressed by any TET: 111.7, 211.1, 211.4, 221.2, and 221.8. As security breaches are unforeseen events, it does not come as a surprise that there are no tools for aiding the understanding of issues related to them. Nonetheless, it is important to notice that the GDPR reserves two Articles to provisions on personal data breaches (Art. 33 and 34), one of which is dedicated to describing how to communicate such matters to the affected data subjects. Being the health-care industry among the ones with most reported breaches, and being medical data in the top three most compromised variety of data (for more details, see results of the data breach investigation (Verizon, 2018)), we consider this to be an area in need of further development.

Legal Aspects. Only a few Articles from the GDPR do not seem to be covered by any of our selected transparency tools. We consider an Article as not covered when none of its paragraphs or sub-paragraphs is correlated to at least one TET. Examples of this are the Articles related to certification; Article 25 regards data protection by design and by default, Article 32 has provisions on security of processing, but both mention that compliance with such Articles may be demonstrated through the use of approved certification mechanisms referred to in Article 42.

Despite having included two certification seals in our study (i.e., EuroPriSe, and TrustArc), we cannot confirm they are approved certification mechanisms. According to EuroPriSe, their criteria catalogue has not been approved pursuant to Article 42(5) GDPR, and they have not been accredited as a certification

body pursuant to Article 43 GDPR yet[23]. Regarding TrustArc, we did not find enough information about this matter.

A few transparency quality and empowerment related Articles are also not addressed by our selected tools. Article 12, for example, qualifies the communications with the data subject and states that it should be concise, easily accessible, using clear and plain language, and by electronic means whenever appropriate. In our understanding, this Article does not correlate to any specific tool because it is transverse to all of them. This Article has provisions regarding the quality of communications; hence, all tools communicating information to data subjects should be affected by it. There are works in the literature discussing metrics for transparency which, in line with this reasoning, consider the information provided to final users "being concise", or "being easily accessible" as indicators that transparency is properly implemented (Spagnuelo et al., 2017).

Article 12 also has provisions regarding the data subject's rights, as do Articles 16, 17, 18, 19, 21, 22, and 26. Articles 17, 19 and 21 do relate to some tools as transparency and empowerment are closely linked. But generally speaking, empowerment related Articles are either partially or not addressed at all by TETs. There are relevant developments in this topic though (Meis and Heisel, 2017). The authors discuss the privacy goal of empowerment (or intervenability) and its relationship to transparency. For instance, Article 12 relates to their requirement T4 and T5, and Article 17 relates to requirement I10. The analysis of these requirements and their relationship with TETs falls out of this work's scope.

It is important to notice that a few Articles which appear not to be covered by any TET, are disregarded from this analysis because they do not match their key-terms with any of our requirements (i.e., Article 9, and 11). We investigate Article 9 manually. This Article has provisions on data subject's consent for data processing of special categories of personal data, including data concerning health. Privacy eSuite tool (PeS) is a web-service consent engine specifically tailored to collect and centralise consent for the processing of health data. This tool is connected with Article 9, and in the interest of completeness, we manually added this correlation to Table 4. However, PeS is a proprietary tool designed in line with the Canadian regulations. We found no means to determine to which extent this tool can help achieving the GDPR provisions.

Being consent one of the basis for lawful processing of personal data described in the GDPR, the number of tools addressing this subject seems suspiciously low. This fact does not imply that medical systems and other services are currently operating illegally. We are aware that collecting consent for processing data is a practice. However, we are interested in tools designed that facilitate the task of collecting consent and help users to be truly informed and aware of the consequences of the consent they are giving.

We investigated this more closely and searched for tools aiming at informed consent. Among our findings there are mostly tools and frameworks aiding the collection of informed consent for digital advertising[24]. We also found mentions to the EnCoRe (Ensuring Consent and Revocation) project, which presents insights on the role of informed consent in online interactions (Whitley and Kanellopoulou, 2010). The project appears to have finalised, and we found no tool proposed to address the collection of informed consent.

One could claim that informed consent can be collected when the user agrees with the terms and conditions, or privacy policies, for which there are tools proposed (e.g., P3P, PPL, and UP). While that may be one possible solution, special attention is required that the request for consent is distinguishable from other matters (as per GDPR Article 7). It is also important to note that consent to the processing of personal data shall be freely given, specific, informed, and unambiguous[25]. In such case, implicitly collecting consent is arguably against the provisions in the GDPR, a viewpoint also defended in (Whitley and Kanellopoulou, 2010). In that work, the authors discuss the extent to which terms and policies are even read and understood. In this sense, consent is unlikely to be truly informed and freely given.

7 CONCLUDING REMARKS

We systematically reviewed the literature of Transparency Enhancing Tools (TETs) in search of what exists to implement the principle of transparency as described by the GDPR. Our selection was guided by a subset of existing technical requirements for medical systems that we have found be semantically correlated with the Articles of the GDPR that define, directly or indirectly, transparency. Out of the 21 GDPR Articles we study here, 12 seem to be, at least partially, addressed by our selected TETs.

The SDM presents a list of classified GDPR Articles, which could replace our selection described in

[23] See https://www.european-privacy-seal.eu/EPS-en/Criteria.

[24] See Conversant, IAB Europe, and ShareThis.

[25] GDPR Article 4 (11).

section 3. Our selection does not contradict the list presented by the SDM, it is simply more detailed. The majority of Articles listed by the SDM are also considered in our selection. With the exception of Articles 5.1.(d), 5.1.(f), and 20 — regarding accuracy of data, security of personal data, and portability of data— which contain provisions on the quality of the data provided by transparency, and should be verified for compliance in every tool. Article 40, referring to the design of codes of conduct for controllers and processors, and could hardly be accomplished through the use of TETs. And Article 42, on certification, which are considered in section 5.

The selection mediated by technical requirements may look like a limitation of our approach, but by doing so we managed to have insights on issues that have less attention in works tailored to discuss compliance with the provisions of the GDPR. For instance, we noticed that matters related to Break-the-Glass (BTG) are not correlated with any provisions from the GDPR. This may be a topic of specific interest in the medical systems domain. Examples of requirements on the subject are 211.2, on informing users when authorities access their data, and 211.3, on informing in case policy is overridden. Although we did not find a clear correlation between these requirements and GDPR Articles, we also have no reasons to believe BTG is out of the scope when discussing data protection principles. A clear indication of that is the number of TETs addressing the subject (e.g., PDT, TAMI, PTT and PVA): although in their early stage of development, which suggests the subject is new, but of interest for the TETs community. Works defending the exercise of access control in one single point ignore the genuine possibility that data is available or can be inferred from somewhere else. Adopting BTG is a suitable alternative that will help to emphasise the importance of individual accountability towards the usage of data (Weitzner et al., 2008).

Similarly, we also found TETs which only correlate to our requirements (e.g., PRA, and PS). We believe they may serve as an inspiration to fill the gap we identified regarding tools for consent and security breaches. Privacy Score (PS), which tests web pages for known security vulnerabilities and provide a short explanation of them, could serve as inspiration for the development of tools explaining security breaches to a broad public. While Privacy Risk Analysis (PRA), which explains the risks and consequences of having a piece of personal data disclosed, could be adapted to help users in giving *de facto* informed consent.

The SDM also comments on *technical measures* that help to guarantee transparency, such as, documentation of procedures, logging of access and modifications. These measures relate to our requirements, but are more high-level. We believe our requirements could be classified according to them, allowing us to select TETs that can accomplish transparency as described by the SDM. We leave this task to future works.

This research is about the tools a system can leverage to accomplish transparency as a technical principle: it is not about ensuring legal compliance with the GDPR. However, we identified future developments which we believe will contribute to better coverage of transparency. We present in this work several tools tailored to one single use case (a few even designed for one specific organisation). Although they are comprehensive in addressing transparency, other systems could not immediately apply them (e.g., Google and Microsoft Dashboards). Those tools should serve as a role model for a possible generic Transparency Enhancing Tool. Other tools are already prepared for general use but are designed with a focus on other regulations in mind. One example of such a tool is the Privacy eSuite (PeS), which is tailored to Canadian regulations. Similarly, Usable Privacy (UP) intends to highlight the most relevant parts of a privacy policy for the American public. Adapting those tools to the GDPR's provisions seems to be an interesting future development for state of the art in transparency.

ACKNOWLEDGEMENTS

Spagnuelo and Lenzini's research is supported by the Luxembourg National Research Fund (FNR), AFR project 7842804 TYPAMED and CORE project 11333956 DAPRECO, respectively.

REFERENCES

Article 29 Working Party (2018). Guidelines on transparency under regulation 2016/679. http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227. Accessed in Aug 2018.

Bartolini, C., Giurgiu, A., Lenzini, G., and Robaldo, L. (2016). A framework to reason about the legal compliance of security standards. In *Proc.of the 10th Int. Workshop on Juris-informatics*.

Berthold, S., Fischer-Hübner, S., Martucci, L., and Pulls, T. (2013). Crime and punishment in the cloud: Accountability, transparency, and privacy. In *Int. Workshop on Trustworthiness, Accountability and Forensics in the Cloud*.

Bier, C., Kühne, K., and Beyerer, J. (2016). PrivacyInsight: the next generation privacy dashboard. In *Annual Privacy Forum*, pages 135–152. Springer.

De, S. J. and Le Métayer, D. (2018). Privacy risk analysis to enable informed privacy settings. In *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE.

de Montjoye, Y.-A., Shmueli, E., Wang, S. S., and Pentland, A. S. (2014). OpenPDS: Protecting the privacy of metadata through safeanswers. *PloS one*, 9(7).

EuroPriSe (2017). Europrise certification criteria (v201701). https://www.european-privacy-seal.eu/EPS-en/Criteria. Accessed in Oct 2018.

Ferreira, A. and Lenzini, G. (2015). Can transparency enhancing tools support patient's accessing electronic health records? In *New Contributions in Inf. Systems and Technologies*, pages 1121–1132. Springer.

Fischer-Hübner, S., Angulo, J., Karegar, F., and Pulls, T. (2016). Transparency, privacy and trust–technology for tracking and controlling my data disclosures: Does this work? In *IFIP Int. Conf. on Trust Management*, pages 3–14. Springer.

Fischer-Hübner, S., Angulo, J., and Pulls, T. (2014). How can Cloud Users be Supported in Deciding on, Tracking and Controlling How their Data are Used? In *Privacy and Identity Management for Emerging Services and Technologies*, volume 421, pages 77–92. Springer.

Fischer-Hübner, S. and Martucci, L. A. (2014). Privacy in social collective intelligence systems. In *Social collective intelligence*, pages 105–124. Springer.

Idalino, T. B., Spagnuelo, D., and Martina, J. E. (2017). Private verification of access on medical data: An initial study. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, pages 86–103. Springer.

Meis, R. and Heisel, M. (2017). Computer-aided identification and validation of intervenability requirements. *Information*, 8(1):30.

Mitkov, R. (2005). *The Oxford handbook of computational linguistics*. Oxford University Press.

Nejad, N. M., Scerri, S., and Auer, S. (2017). Semantic similarity based clustering of license excerpts for improved end-user interpretation. In *Proc. of the 13th Int. Conf. on Semantic Systems*, pages 144–151. ACM.

OPC (2017). Privacy Enhancing Technologies - A Review of Tools and Techniques. https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/. Accessed in Aug 2018.

Raschke, P., Küpper, A., Drozd, O., and Kirrane, S. (2017). Designing a GDPR-Compliant and Usable Privacy Dashboard. In *IFIP Int. Summer School on Privacy and Identity Management*, pages 221–236. Springer.

Sackmann, S., Strüker, J., and Accorsi, R. (2006). Personalization in privacy-aware highly dynamic systems. *Comm. of the ACM*, 49(9):32–38.

Sathyendra, K. M., Wilson, S., Schaub, F., Zimmeck, S., and Sadeh, N. (2017). Identifying the provision of choices in privacy policy text. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2774–2779.

Seneviratne, O. and Kagal, L. (2014). Enabling privacy through transparency. In *Privacy, Security and Trust, 12th Annual Int. Conf. on*, pages 121–128. IEEE.

Siljee, J. (2015). Privacy transparency patterns. In *Proc. of the 20th European Conf. on Pattern Languages of Programs*, page 52. ACM.

Spagnuelo, D., Bartolini, C., and Lenzini, G. (2017). Modelling metrics for transparency in medical systems. In *Int. Conf. on Trust and Privacy in Digital Business*, pages 81–95. Springer.

Spagnuelo, D., Ferreira, A., and Lenzini, G. (2018). Accomplishing transparency within the general data protection regulation (auxiliary material). http://hdl.handle.net/10993/37692.

Spagnuelo, D. and Lenzini, G. (2016). Transparent medical data systems. *Journal of Medical Systems*, 41(1):8.

TrustArc (2018). Enterprise privacy & data governance practices certification assessment criteria. https://www.trustarc.com/products/enterprise-privacy-certification/. Accessed in Oct 2018.

Verizon (2018). 2018 data breach investigations report. https://www.verizonenterprise.com/verizon-insights-lab/dbir/. Accessed in Oct 2018.

Weitzner, D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., and Sussman, G. J. (2008). Information accountability. *Comm. of the ACM*, 51(6):82–87.

Whitley, E. A. and Kanellopoulou, N. (2010). Privacy and informed consent in online interactions: Evidence from expert focus groups. In *ICIS*, page 126.

Zimmermann, C. (2015). A categorization of Transparency-Enhancing Technologies. *arXiv preprint arXiv:1507.04914*.