Explaining Disclosure Decisions Over Personal Data

Roghaiyeh (Ramisa) Gachpaz Hamed, Harshvardhan J. Pandit, Declan O'Sullivan, Owen Conlan

ADAPT Centre, Trinity College Dublin, Ireland
{ramisa.hamed,harshvardhan.pandit,declan.osullivan,owen.conlan}@adaptcentre.ie

Abstract. The use of automated decision making systems to disclose personal data provokes privacy concerns as it is difficult for individuals to understand how and why these decisions are made. This research proposes an approach for empowering individuals to understand such automated access to personal data by utilising semantic web technologies to explain complex disclosure decisions in a comprehensible manner. We demonstrate the feasibility of our approach through a prototype that uses text and visual mediums to explain disclosure decisions made in the health domain and its evaluation through a user study.

1 Introduction

While individuals understand the value of their personal data, they are largely concerned with its potential misuse by businesses and governments [4], and bemoan the lack of a trusted entity that affords them control or advice regarding protection of their data [5]. Legislation such as the EU's General Data Protection Regulation (GDPR)[1] aim to preserve privacy by making data processing more transparent and accountable. Recital 71 of the GDPR additionally states the "right to explanation" for automated-decisions with significant effects for individuals. Providing more information and control is beneficial to all stakeholders as individuals who perceive themselves to be in control over the release and access of their private information (even information that allows them to be personally identified) have greater willingness to disclose this information [2].

This research believes that technology can be used for social good by having automated decisions over the access and disclosure of personal data if provided with transparent explanations of the decisions made. To this end, we present a framework for empowering individuals to understand how and why a decision was made regarding access to their personal data. We present the feasibility of our approach through a prototype for a use-case in the health domain. The prototype utilises semantic web technologies to explain complex disclosure decisions of a reasoning process through textual and visual representations.

[1] Regulation (EU) 2016/679. Official Journal of the European Union. L119, 1–88 (2016)

2 Framework & Prototype Implementation

The aim of our framework is to enable individuals to understand automated disclosure decisions over their data by explaining the facts and logic for how the system arrived at a particular decision. The framework is comprised of the following components: (1) **Data Resources:** includes *personal data* and a *knowledge base,* comprised of - domain *ontologies*, *privacy rules* defined by individuals or regulation and *logs* that record decision history. (2) **Actors:** includes *data owners* - individuals to whom the personal data relates, and *data requesters* - entities that request access to *personal data*. (3) **Functional Components:** these include *automatic* and *semi-automatic decision making* units, a *logger* to record all decisions, and a *context-discovery* unit for gathering contextual information about actors for use in the decisions. The *decision maker* unit utilises a semantic *reasoner* over the collected *knowledge-base* and disclosure history to decide the response for access requests. These decisions are intended to be automatic with the *decision explainer* capable of providing an explanation, but can be semi-automatic where the decision maker does not have sufficient information or when the data owner needs to (manually) change a decision - in which case the *confirmer* obtains confirmation to make the disclosure.

We implemented a prototype using semantic web technologies with a focus on explaining the inference of semi-automatic disclosure decisions over personal data. It uses RDF/OWL to define the data graph, with human-readable information using *rdfs:label* for each node. The data disclosure rules are defined using SWRL[2] with a human-readable description. Requests for data access are added as triples using the Apache Jena API[3], and the Pellet reasoner[4] is used to match requests with privacy rules, where matched rules are retrieved using SPARQL[5] and indicate granted access to data.

The explanation of disclosure decisions was provided using human-readable descriptions for entailments in the form of text and visual representation as depicted in Fig. 1. We used OWL Explanation [3] to obtain the entailment as a set of axioms consisting of the minimal subset of the data graph sufficient for the entailment and the corresponding SWRL rules used for making decisions. The axioms were filtered to remove ontological declarations, type information for instances, and domain/range information for properties. The access request was also removed as it is considered the outcome rather than an explanation. The axioms were then reduced by substituting type declarations with the *rdfs:label* of the declared concepts. Finally, text representations were generated by translating each triple into a statement, and visual representation were generated by using GraphViz[6].

[2] I. Horrocks et al., "SWRL : A Semantic Web Rule Language Combining OWL and RuleML," W3C Member submission 21.79 (2004).
[3] http://jena.apache.org/documentation/ontology/
[4] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz,"Pellet: A practical OWL-DL reasoner"Web Semantics: science, services & agents on the World Wide Web 5.2 (2007)
[5] https://www.w3.org/TR/rdf-sparql-query/
[6] https://www.graphviz.org/

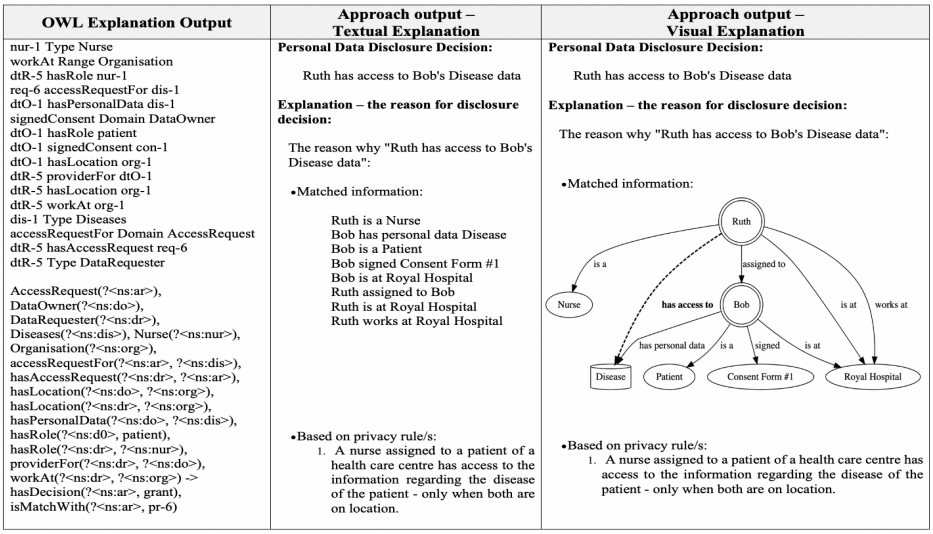

Fig. 1. Overview of textual and visual representation for explanation

3 User-Study

The user-study consisted of participants being shown two scenarios with disclosure decisions and their explanation for issues in the health-domain [6]. Each user was shown a disclosure decision and its explanation with textual representation for one scenario and a visual representation for the other which allowed us to compare the understandability of textual and visual mediums for explanations. The users first had to provide explicit informed consent for the study, after which a questionnaire solicited users' perceptions about disclosure decisions and access to data. The users were then shown two scenarios with their disclosure decision explained using a random permutation of textual and visual mediums. Each scenario was followed by a SUS [1] questionnaire assessing usability, three comprehension questions and plus one attention question, followed by an ASQ[7] questionnaire assessing satisfiability.

Three comprehension questions enquired user's understanding of explanation for disclosure decision, where two were multiple choice (MCQ) and one was multiple choice with multiple answers (MA). Scoring was from 0 to +3 where higher indicated better comprehension, and was based on awarding +1 for correct choice in MCQ and +1/n for MA where n was total number of options.

We used Prolific[8] to recruit 21 participants which consisted of paying £2.50 for 25mins required to complete the tasks. One participant was rejected for failing to answer the attention question, and results were analysed for remaining 20 participants, as shown in Table 1. The results for SUS indicate good usability being above 71.4 [1], and those for ASQ (lower is better)[7] similarly indicate acceptable satisfaction regarding

[7]After-Scenario Questionnaire - Lewis, J. R. (1991): Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ

[8] https://prolific.ac/

provided explanations in both mediums. The scores for comprehension similarly represent sufficient understanding of the explanations, though values for the comparatively simpler Scenario 1 reflect better understanding as compared with the more complex Scenario 2. The lack of sufficient difference between scores for text and visual mediums indicates similar comprehension and understanding for both scenarios, and show that the values are not conclusive to decide their comparative effectiveness.

Table. 1. Overview of textual and visual representation for explanation

	S1 Text		**S1 Visual**		**S2 Text**		**S2 Visual**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SUS	77.25	23.44	72.75	15.61	71.5	10.55	74.25	17.25
Comprehension	2.43	0.49	2.23	0.78	1.93	0.8	1.98	0.98
ASQ	2.53	1.19	2.17	0.97	2.67	1.2	2.2	0.61

4 Conclusion and Future Work

This paper addressed privacy concerns regarding automated disclosure decisions over personal data using a framework that provides explanations to empower users to understand the reason of access to their personal data. The paper showed feasibility of the framework through a prototype implemented using semantic web technologies to explain disclosure decisions using textual and visual mediums. A user-study evaluation of showed acceptable comprehension of explanations based on the prototype.

For future work, we plan to undertake further user-studies involving greater number of participants and complex scenarios in order to compare the effectiveness of text and visual mediums on the ability of participants to comprehend disclosure decisions. We also plan to incorporate graph-summarisation techniques and queries to simplify complex scenarios and improve explanations of decisions.

Acknowledgement. This research is supported by the ADAPT Centre for Digital Content Technology, is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

1. Bangor, A., Staff, T., Kortum, P., Miller, J., & Staff, T. Determining What Individual SUS Scores Mean : Adding an Adjective Rating Scale, 4(3), 114–123. (2009).
2. Brandimarte, L., Acquisti, A., & Loewenstein, G. . Misplaced Confidences: Privacy and the Control Paradox. Social Psychological and Personality Science, 4(3), (2013)
3. Horridge, M., Parsia, B., &x Sattler, U.. The OWL Explanation Workbench : A toolkit for working with justifications for entailments in OWL ontologies, 1, 1–5. (2009)
4. Morey, T., Forbath, T., & Schoop, A. Customer Data : Designing for Transparency and Trust. (2016).
5. Web Foundation. PERSONAL DATA: An overview of low and middle-income countries. (2017)
6. Yüksel, B., Küpçü, A., & Özkasap, Ö. (2017). Research issues for privacy and security of electronic health services. Future Generation Computer Systems, 68, 1–13. (2017)