Ontology-based Negotiation and Enforcement of Privacy Constraints in Collaborative Knowledge Discovery

Lauri Tuovinen[1,2] [0000−0002−7916−0255] and
Alan F. Smeaton[1] [0000−0003−1028−8389]

[1] Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland
lauri.tuovinen@insight-centre.org, alan.smeaton@dcu.ie
[2] Biomimetics and Intelligent Systems Group, University of Oulu, Oulu, Finland

Abstract. Many people could benefit from collecting and analyzing their own personal digital data, but most do not possess the necessary expertise to do so. Remote collaboration with knowledge discovery experts who do possess this expertise is a possible solution to this conundrum but raises a number of issues of its own, one of which is preserving the data owner's privacy. It is up to the data owner to decide how much data to share with a data analyst, but withholding too much will make the analyst unable to help the data owner effectively, so it is necessary to find a trade-off between these two conflicting interests. We propose a solution whereby the data requirements imposed by analysis tasks and the access restrictions imposed by privacy constraints are encoded formally using an ontology, enabling automatic detection of conflicts. Once a conflict has been identified, the data owner and the data analyst can negotiate a resolution, possibly by transforming the data using a method that makes it no longer sensitive from the data owner's perspective while sufficiently preserving its utility from the data analyst's perspective. Using such an ontology, data owners and data analysts tap into a knowledge base of privacy-preserving data transformations, each with known effects on the utility of the transformed data for analysis. This makes it easier to find an acceptable trade-off between privacy and utility in future collaborations.

Keywords: Knowledge discovery · Data mining · Personal data · Privacy · Collaborative systems · Ontologies

1 Introduction

Digital data is nowadays often referred to by expressions such as "the new oil", emphasizing the potential business value of data. Social media platforms, for instance, commonly depend on collecting and analyzing their users' personal data for revenue. Analysis results are packaged into business-to-business products and services such as targeted advertising.

Something that the data-as-resource rhetoric tends to overlook, however, is that analyzing personal data could also be of considerable value to individuals themselves. This value could be for health and wellness reasons such as monitoring dietary intake or activity levels as part of some weight loss or exercise programme, or it could be for sports training with a focus on performance improvement or perhaps for rehabilitation after an accident or an illness. The problem is that extracting valuable knowledge hidden in data requires special skills that not many people have, nor can they all be expected to acquire them.

A possible solution to the challenge of people exploring their own personal data which we explore here is collaborative knowledge discovery, where the owner of the data works together with an expert who is willing to help the owner analyze the data. If the data owner and the data analyst are friends, the analyst might do this as a favor; if not, the analyst might do it for a fee, or for the opportunity to gain access to some real-world research data. In any case, some of the data owner's personal data will need to be shared with the data analyst, which means that some measures will need to be taken to preserve the data owner's privacy.

Depending on the data owner's privacy preferences, some of the data that the owner would rather not disclose to the data analyst may be data that would in fact be useful in achieving the owner's knowledge discovery objectives. This can lead to a conflict of interest where, for some data items, it can be argued both that they should be shared with the analyst, to enable effective collaboration, and that they should not be shared, to avoid disclosure of sensitive information to an untrusted person. A resolution to the conflict may be found through negotiation, but first, the owner and the analyst need to be aware that there is a conflict. Furthermore, after a suitable degree of privacy protection has been agreed upon, it must be enforced throughout the collaborative knowledge discovery process.

To enable the detection of conflicts and the enforcement of data access restrictions, a mechanism is needed that allows participants in the collaboration to encode their points of view in a formal, machine-readable representation. For this purpose, we propose an approach based on an ontology that data analysts and data owners can use to represent actionable knowledge about data analysis tasks, privacy constraints, and the datasets that these operate on. The analysis tasks are described in terms of which data items they require and the privacy constraints in terms of which data items are to be protected, enabling a reasoner to detect inconsistencies and notify the participants that they need to begin negotiations to resolve the conflict.

Furthermore, an ontology also enables representation of knowledge about the methods used to implement different types of analysis tasks and privacy constraints, including how applying a given privacy protection method (e.g. presenting a sensitive variable in an aggregated form) affects the utility of a given data analysis method. If there is sufficient information available on the characteristics of different analysis methods, alternative methods can be suggested based on the ontology to minimize the reduction of utility in conflict-of-interest situations. In the best case, a mutually acceptable resolution to the conflict can be generated algorithmically without any additional input from the data owner or the data analyst.

Our proposed ontology is an early work-in-progress being developed as part of a project to create a new collaborative knowledge discovery software platform with special focus on personal analytics. The main benefits of the ontology-based approach, once fully implemented and validated, are the following:

- It enables the accumulation of a knowledge base where various methods used to transform datasets in collaborative knowledge discovery are also characterized in terms of their impacts on privacy.
- It allows specification of constraints concerning individual data items, specific categories or subsets of data, or entire datasets; furthermore, different sets of constraints can be applied depending on the circumstances of the collaboration and the identity of the data analyst.
- It allows data owners to control not just the original datasets, but also all derivatives generated over the course of the collaboration, enabling more comprehensive awareness of privacy implications and therefore better-informed decisions regarding data disclosure.
- It can be generalized to cover scenarios with varying numbers of data owners and data analysts (although for simplicity, the paper mainly considers a scenario where there is just one data owner and one data analyst).

The remainder of the paper is organized as follows: Section 2 presents background information and discusses related work on collaborative knowledge discovery, lifelogging (the practice of collecting, archiving and analyzing data about one's daily life) and ontologies. Section 3 gives an overview of the requirements and structure of the proposed ontology. Section 4 demonstrates how the ontology would be applied in practice to enable the negotiation, representation and enforcement of privacy constraints in a collaborative knowledge discovery effort. Section 5 discusses the validity of our approach, open issues and future work. Section 6 concludes the paper.

2 Background

For the purposes of this paper, collaborative knowledge discovery can be defined as the process of two or more individuals, who may be geographically distributed and previously unknown to one another, forming a team and working together to extract useful knowledge from datasets of personal data owned by some of the participants. The participants may work synchronously or asynchronously, and will use an online collaboration platform that enables them to share their data and execute analytical algorithms on shared data. This is a relatively new concept, and while there are systems such as LabBook [12] that support this type of knowledge discovery, applying them in the context of personal analytics is largely unexplored territory. As a result, handling privacy issues in collaborative knowledge discovery is not very well understood.

What makes collaborative knowledge discovery unique in this respect is the nature of the relationship between data owners and data analysts. This relationship is defined by the following properties:

- The data owners and data analysts are engaged in an *ongoing, direct collaboration*, working synchronously or asynchronously and probably not colocated, and the objectives and constraints of the collaboration are dynamically determined through negotiation.
- The knowledge discovery results are to be used *primarily for the benefit of the data owners*, with the participation of the data analysts motivated by some incentive other than the generated knowledge itself.
- There is at least some degree of *separation of roles* between data owners and data analysts, that is, the data to be analyzed and the expertise required to analyze it are not held by the same individuals.

The existence of a direct relationship between data owners and analysts distinguishes collaborative knowledge discovery from simple data sharing (e.g. publication of personal data as open data), where privacy issues are generally handled by purging the dataset to be shared of any information that could be used to associate some of the data with a specific identified individual. Once this operation has been carried out, there is no more need for the data owners to be involved in controlling access to the data. However, the data transformation methods developed for this purpose constitute a major part of the toolkit for privacy protection in collaborative knowledge discovery. A recent survey of such methods can be found in [16].

Another situation that is similar but not fully analogous to collaborative knowledge discovery scenario is when an individual shares some personal data with a service provider in exchange for access to the service. If the service allows its users to specify their personal privacy preferences, this can be viewed as a kind of negotiation between the data owner (service user) and the data analyst (service provider), but it is primarily the data analyst, not the data owner, who benefits from the knowledge extracted from the data.

Finally, there is a form of distributed knowledge discovery often referred to as collaborative, where the local datasets of multiple data owners are combined into a single database and served to the participating data owners via a protocol that enables them to mine it for global knowledge while preventing them from directly accessing the original datasets [24]. Here it is the data owners who are the direct beneficiaries of their own respective knowledge discovery activities, but the roles of data owner and data analyst are not separated, and collaboration among the data owners is limited to a special form of data sharing.

As a specific example of a domain where collaborative knowledge discovery is potentially beneficial, but only if the privacy concerns associated with it can be addressed adequately, lifelogging is considered in this paper. As discussed in [9], lifelogging is something of an umbrella term, but the concept is characterized as persistent archiving of multi-modally captured data about a single individual person and an aspiration to represent a comprehensive history of that person's, the lifelogger's, experiences and actions, with a variety of potential applications such as activity monitoring or memory assistance. Lifelogging in this sense is actually a specialist interest and is also sometimes referred to as the quantified self movement. The popularity of personal sports/activity and sleep trackers shows that many people are interested in self-monitoring as a means of enhancing one's own well-being, provided that the process of collecting and manipulating their own data is made sufficiently convenient for them. Collaborative knowledge discovery, supported by a novice-friendly collaboration platform, could provide the solution in situations where a lifelogger wants to explore their own data and go beyond the straightforward dashboards typically generated by commercial off-the-shelf products but does not have the necessary skills to develop and/or apply the required data mining algorithms without expert assistance. An analysis of privacy in the context of lifelogging can be found in [6] and [11].

Using ontologies for access control and privacy preservation is a topic that has been studied extensively, but the domain of knowledge discovery is seldom considered in such studies. Conversely, while a considerable number of knowledge discovery ontologies have been proposed, these do not address the privacy issues involved in knowledge discovery on personal data. Most importantly, at the intersection of privacy, knowledge discovery and collaborative systems there is a noteworthy gap in the state of the art where these are brought together and unified. However, there are some research results in this area that, despite not addressing the specific problem discussed in this paper, are pertinent enough to be mentioned here. The Privacy Preference Ontology (PPO) [20], for instance, is a general-purpose ontology for expressing and enforcing privacy preferences for personal data on the web, and in [4] an ontology is proposed for the enforcement of privacy legislation. The privacy ontology of [1] addresses collaborative environments specifically, albeit from an organizational perspective, not an individual one.

One of the goals of the ontology-based approach taken in this paper is to enable access control where, instead of simply allowing or denying access, the requested data may be automatically transformed to make it less sensitive before it is returned to the requester. Similar ideas have been discussed in [7, 10], and also related to this topic is work in [2], which discusses automatic generation of generalization hierarchies that can be used to anonymize a dataset through abstraction. The utility of ontologies in anonymizing non-numerical data is demonstrated in e.g. [15, 19]. Another goal is that data owners should be able to control access to derivative datasets as well as the original ones, which is one of the features of the PriArmor framework [8]. Also related is the concept of knowledge level privacy presented in [21].

One particularly relevant feature of the PriArmor framework is that it enables data owners to specify their own personal privacy policies. Empowering data owners in such a way is crucial to the success of collaborative knowledge discovery; besides the technical capability to create policies, this requires that the process of creating privacy policies that accurately represent their privacy preferences be accessible enough to enable data owners to carry it out themselves regardless of their computing background or lack thereof. Other works related to this goal include the privacy negotiation service of [14], the risk-detecting Privacy Oracle of [3] and the Semantic Weighted Context Tagging Engine (SWCTE) of [18], which aims to automatically identify sensitive data based on semantic context. Also interesting are studies on empowering patients to control access to their own medical records in the healthcare domain (e.g. [13, 22]), since such records are widely considered to contain sensitive information but the individuals concerned cannot be expected to be versed in the technical aspects of privacy and access control.

3 Overview of the ontology

The principal requirements of the ontology being developed here are enumerated below:

1. The ontology should enable data analysts to specify which data items they require access to, and data owners to specify which data items they wish to remain private.
2. The ontology should enable detection of conflicts where a data item designated as required by a data analyst is designated as private by the data owner.
3. The ontology should enable the representation of information about how different privacy protection measures affect the utility of data and how the performance of different data analysis methods is impacted as a result.
4. The ontology should enable representation of knowledge about which privacy protection measures and data analysis methods can be considered as interchangeable, allowing recommendation of alternatives in case of a conflict.
5. The ontology should enable descriptions of data analysis tasks in terms of which derivative datasets they produce, allowing data owners to control their privacy throughout the collaborative knowledge discovery process.

The ontology designed to satisfy these requirements is shown as a graph in Figure 1 and described in detail in the following paragraphs. In the description, the names of ontology classes are written in **boldface** when first mentioned and with Initial Capitals throughout.

The core classes of the ontology are **Collaboration**, **Data Owner**, **Data Analyst**, **Dataset**, **Analysis Task** and **Privacy Constraint**. A Collaboration has some Data Owners and Data Analysts as participants, and involves some Analysis Tasks, Datasets and Privacy Constraints. Each Dataset is controlled by a Data Owner and consists of some **Data Items**, which in turn may belong to any number of **Data Categories**. Datasets and Data Categories share a common superclass, **Data Grouping**; a Data Grouping may have other groupings as subgroupings. For Datasets, these are subsets, which can be used to represent, for example, individual variables within the Dataset, whereas for Data Categories, they are used to express that one category, e.g. “medical data”, is

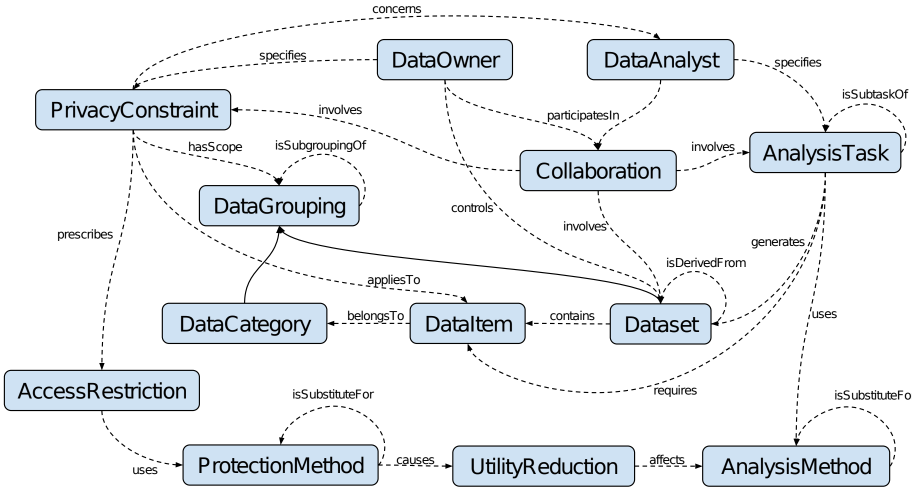

Fig. 1. An ontology of data analysis tasks and privacy constraints in collaborative knowledge discovery on personal data. Subclass-superclass relationships are represented by arrows with a solid line and object properties (relationships between individuals) by arrows with a dashed line pointing from the domain to the range.

a subcategory of another category, e.g. "sensitive data". Membership in a given grouping implies membership in all groupings that have it as a subgrouping.

A Data Analyst participating in a Collaboration may specify Analysis Tasks, each of which requires access to some Data Items and uses some **Analysis Methods**. An Analysis Task may be composed of multiple subtasks. Similarly, a Data Owner participating in a Collaboration may specify Privacy Constraints, each of which concerns some Data Analysts, applies to some Data Items and prescribes the application of an **Access Restriction**. Instead of individual Data Items, a Privacy Constraint may refer to a Dataset or a Data Category (via the hasScope property), in which case the Data Items concerned can be inferred via property chains defined for this purpose. Therefore for any given Data Item, it is possible to infer whether it is the subject of both an Analysis Task and a Privacy Constraint and whether the constraint refers to the same Data Analyst who specified the task. Requirements 1 and 2 are thus satisfied.

An Access Restriction uses a **Protection Method** to implement the restriction. This may be, for example, a specific way of aggregating Data Items to generate a more abstract (and therefore less sensitive) representation of the Dataset containing them. Denying access entirely can be viewed as a special case of Protection Method. The application of a Protection Method causes a **Utility Reduction**, which may be different for each individual kind of Analysis Method, enabling the characterization of Protection Methods and Analysis Methods in terms of how compatible they are with one another. Furthermore, each Protection Method and Analysis Method can be designated as being a substitute for other Protection Methods and Analysis Methods, respectively. This enables the enumeration of all valid combinations of Protection Method and Analysis Method and the selection of the combination that results in the least Utility Reduction. Requirements 3 and 4 are thus satisfied.

Requirement 5 is satisfied by associating each Analysis Task with the Datasets it generates. The generated Datasets are inferred to be derivatives of the input Datasets via a property chain that connects them to the Datasets containing the Data Items used in the Analysis Task. Dataset derivation is represented by a transitive property, so that if Dataset B is a derivative of Dataset A and Dataset C is a derivative of Dataset B, then it can be inferred that Dataset C is also a derivative of Dataset A. The control of a Data Owner over a Dataset is extended via another property chain to cover the derivatives of that Dataset generated during the Collaboration, so the Data Owner can specify Privacy Constraints concerning the derivatives as well as the original Dataset.

The ontology was modeled in the OWL 2 Web Ontology Language using the Protégé ontology development tool [17] and the FaCT++ reasoner plugin [23]. To improve the readability of Figure 1, some entities are omitted from the graph. These include the classes **Collaborator** (superclass of Data Owner and Data Analyst), **Requirement** (superclass of Analysis Task and Privacy Constraint) and **Transformation Method** (superclass of Analysis Method and Protection Method); the object properties isSubsetOf and isSubcategoryOf, which are subproperties of isSubgroupingOf; and a large number of inverse properties, which have been defined for most of the properties shown. Additionally, there are two object properties, isRequestedBy and isDeniedTo, with domain Data Item and range Data Analyst. The former means that the Data Item in question is required by an Analysis Task specified by the Data Analyst, the latter that the Data Item is subject to a Privacy Constraint concerning the Data Analyst. The existence of these relationships is inferred via property chains, enabling identification of Data Items that are subject to conflicting interests.

4 Applying the ontology

When using the ontology to support a collaborative knowledge discovery effort, the objectives of the collaboration are encoded by creating individuals of the Analysis Task class. To represent what data is to be used in the task, the Analysis Task individual is linked to individuals of the Data Item class via the requires property, and to represent how the data is to be processed, it is linked to individuals of the Analysis Method class via the uses property. The outputs of the task are specified by linking the Analysis Task individual with individuals of the Dataset class via the generates property, causing relationships to be established from the outputs to the inputs via the isDerivedFrom property.

Similarly, constraints are encoded by creating individuals of the Privacy Constraint class. To represent the data and collaborators covered by the constraint, the Privacy Constraint individual is linked to individuals of the Data Item and Data Analyst classes via the appliesTo and concerns properties, respectively. The effect of the constraint is specified by linking the Analysis Task individual to an Access Restriction individual via the prescribes property, which in turn is linked to a Protection Method individual via the uses property. The Protection Method individual represents the transformation to be applied to the data before it can be released for analysis.

To illustrate how an ontology-based approach would work in practice, let us consider a scenario where a lifelogger would like to develop an application that maintains a log of his physical exercise activities without requiring him to enter them manually. To achieve this, the lifelogger needs a computational model capable of recognizing such activities directly from his lifelogging data, specifically from wearable sensors which record his activity levels and some physiology characteristics. To obtain such a model, the lifelogger collaborates with a researcher who is studying activity recognition and classification and is willing to help in exchange for the ability to test and refine her models on the lifelogger's real-world data.

Among the data the lifelogger has collected are the outputs of some wearable activity monitors, providing readings such as heart rate and acceleration, and GPS coordinates. The researcher would like to use all of this data to train the model, believing that the GPS data would be useful in calculating distance travelled and in improving classification accuracy by using the lifelogger's nearness to known sports facilities and known running/cycling routes as one of the input variables. However, the lifelogger would prefer to keep the GPS data private and creates a constraint to that effect. As a consequence of this, the ontology-based collaboration platform detects a conflict and triggers a negotiation between the lifelogger and the researcher.

To resolve the conflict, the collaborators agree that instead of the researcher being permitted to access the raw GPS coordinates, the data will be sanitized using the method described in [5] where semantic locations such as *home* are substituted for exact locations. This reduces the utility of the data with respect to the calculation of distance travelled, but retains sufficient utility for the purpose of classification of activities. Having determined this to be an acceptable trade-off between privacy and utility, the lifelogger and the researcher proceed with the collaboration.

To model this scenario in Protégé, the ontology was populated with dummy individuals of each of the classes shown in Figure 1. The reasoner was then activated to confirm that the following inferences are made:

- All Data Items are inferred to be members of the correct Data Groupings based on asserted subgrouping relationships.
- All Data Items asserted as required by an Analysis Task specified by the Data Analyst (researcher) are inferred to be requested by the Data Analyst.
- All Data Items asserted as belonging to a Data Category representing GPS coordinate data are inferred to be subject to a Privacy Constraint that has as its scope a supercategory of that category (representing location data in general).

– All Data Items inferred as being subject to the aforementioned Privacy Constraint are also inferred to be denied to the Data Analyst (because the constraint refers to the analyst via the concerns property).
– The Dataset asserted as generated by the Analysis Task is inferred to be derived from each of the Datasets containing some of the Data Items required by the task, and the Data Owner (lifelogger), having been asserted to control the original Datasets, is inferred to also control the generated Dataset.

It is worth noting here that not all of the information required to resolve the conflict is necessarily already available in the knowledge base. However, once the collaborators have agreed on the proper course of action, they can record the new information on which the agreement is based as new expert knowledge for the knowledge base, where it will remain available for use in future collaborations. If the knowledge base is shared with other users of the collaboration platform, this enables an implicit form of collaboration where each user can benefit from all of the process knowledge accumulated over all of the collaborations carried out using the platform. Also, as more and more knowledge is accumulated in the knowledge base, the encoded expertise makes data owners increasingly independent of data analysts, allowing them to use the knowledge base to carry out tasks for which they previously would have required a human expert collaborator.

5 Discussion

The ontology as presented in this paper is an early-stage model that has not yet been implemented in full, it is a work-in-progress. As such, however, it is a plausible solution to a problem that, as shown in Section 2, is not adequately addressed by existing work. The analysis of the ontology structure in Section 3 shows that the ontology can support the kind of reasoning required by the problem, and the worked scenario presented in Section 4, while somewhat simplistic, demonstrates the potential of the ontology-based approach in detecting and resolving privacy-related conflicts in real-world knowledge discovery collaborations.

In its present state, the ontology is mainly useful for detecting conflicts between the data requirements of analysis tasks and the access restrictions imposed by privacy constraints where the constraints are of the straightforward denial-of-access type. Technically, the ontology can be used to represent any type of constraint, but it provides limited means to express the semantics of different types. Most of the actual reasoning concerning the compatibility of a given constraint with a given analysis task is therefore left up to human experts, with the knowledge base possibly able to provide implementation options but unable to differentiate between them in terms of which of them would be preferable in the current situation.

The short-term future work to be done on the ontology thus largely involves creating new classes and properties to enable more detailed representation of the concepts that connect the privacy and analysis sides of the ontology at implementation level: Access Restriction, Protection Method, Utility Reduction and

Analysis Method. These are crucial enablers for more advanced functionality where the collaboration platform is able to not just point out a conflict of interest, but to analyze the nature of the conflict and discover possible ways to resolve it, possibly using its own knowledge base of how past conflicts were resolved. It would also be important for data owners to have the means to express their privacy preferences semantically rather than in terms of specific privacy-preserving transformations. Much of this can probably be accomplished by re-using existing privacy and knowledge discovery ontologies, but the concept of Utility Reduction, in particular, can still be expected to require a substantial research effort.

Apart from addressing these open issues, the most immediate item of future work is to validate the ontology in an environment where it is not an isolated entity but part of a system capable of supporting real-world collaborative knowledge discovery efforts. This will involve building a prototype collaborative knowledge discovery platform enabling multiple users to set up a collaboration and specify analysis tasks and privacy constraints, with the ontology providing the underlying formal representation of these objects and the capacity to identify data access conflicts through algorithmic reasoning. To evaluate the prototype platform, the simplistic application scenario discussed in this paper will be replaced with more realistic lifelogging case study scenarios.

Our long-term plan for the ontology is not just to solve the aforementioned problems, but to expand the ontology and link it with other relevant ontologies such that eventually the entire process of collaborative knowledge discovery, from finding of collaborators to deployment of results, is covered. Similarly, we are developing not just a software tool for negotiating privacy constraints, but a suite of tools capable of supporting all collaborative knowledge discovery tasks. Some of these tools may be ones created by other authors, provided that they adequately address the special requirements of collaborative knowledge discovery on personal data: protection of privacy and empowerment of data owners regardless of their technical skills.

Concerning empowerment, one aspect to consider is providing novice-friendly user interfaces that data owners can use to specify their expectations in terms of the concepts of the underlying ontology. Besides user interface concepts and technologies that are currently mainstream, more experimental technologies such as natural language interfaces may prove useful here. However, it is equally important to ensure that the data owners are fully *aware* of the privacy risks they need to protect their data against. Here, the proposed ontology enables an interesting possibility: because the ontology keeps track of derivative datasets, data owners can be presented with a preview of all the information that will be disclosed to the data analyst over the course of the collaboration. In the case of derivatives, the preview obviously cannot display individual data items, but at the very least the data owner would be able to see what kind of information is included in the derivatives and raise questions about its sensitivity. This would make the knowledge discovery process less of a black box to the data owner, allowing the data analyst and the data owner to collaborate on more equal terms.

6 Conclusion

Collaborative knowledge discovery is a relatively new concept where geographically distributed data owners and data analysts use an online platform to collaboratively extract useful knowledge from the data owners' datasets. One of the potential benefits of collaborative knowledge discovery is that it enables non-expert individuals to utilize their personal data by collaborating with experts, but this may lead to conflicts of interest where the data owner's privacy preferences clash with the data requirements of the analysis tasks to be carried out. To support the successful resolution of such conflicts, we proposed in this paper an ontology-based solution where analysis tasks and privacy constraints are encoded in a formal representation and a reasoner is used to identify conflicts between them.

Instead of simply denying access, a privacy constraint may specify that a sensitive subset of the data be transformed using a method that removes the sensitive aspect through, for example, abstraction or perturbation of the data. Such transformations generally have a negative effect on the utility of the data, but the outcome may be an acceptable trade-off between the data owner's privacy preferences and knowledge discovery goals. The ontology enables the collaborators to seek such trade-offs by providing concepts that can be used to represent the effects of different kinds of transformations on the utility of different kinds of data analysis methods. However, this part of the ontology in particular is a work-in-progress and requires further elaboration before it can be applied in practice.

References

1. Allison, D.S., Kamoun, A., Capretz, M.A.M., Tazi, S., Drira, K., Elyamany, H.F.: An ontology driven privacy framework for collaborative working environments. International Journal of Autonomous and Adaptive Communications Systems **9**(3–4), 243–268 (2016). https://doi.org/10.1504/IJAACS.2016.079624
2. Ayala-Rivera, V., Murphy, L., Thorpe, C.: Automatic construction of generalization hierarchies for publishing anonymized data. In: Knowledge Science, Engineering and Management, pp. 262–274. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47650-6_21
3. Barhamgi, M., Yang, M., Yu, C.M., Yu, Y., Bandara, A.K., Benslimane, D., Nuseibeh, B.: POSTER: Enabling end-users to protect their privacy. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. pp. 905–907 (2017). https://doi.org/10.1145/3052973.3055154
4. Bekara, K., Laurent, M., Nguyen, T.H.: Technical enforcement of European privacy legislation: An access control approach. In: 2012 5th International Conference on New Technologies, Mobility and Security (NTMS). pp. 1–7 (2012). https://doi.org/10.1109/NTMS.2012.6208724
5. Dang-Nguyen, D.T., Zhou, L., Gupta, R., Riegler, M., Gurrin, C.: Building a disclosed lifelog dataset: Challenges, principles and processes. In: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing. pp. 22:1–22:6 (2017). https://doi.org/10.1145/3095713.3095736

6. Ferdous, M.S., Chowdhury, S., Jose, J.M.: Analysing privacy in visual lifelogging. Pervasive and Mobile Computing **40**, 430–449 (2017). https://doi.org/10.1016/j.pmcj.2017.03.003
7. Fornara, N., Marfia, F.: Modeling and enforcing access control obligations for SPARQL-DL queries. In: Proceedings of the 12th International Conference on Semantic Systems. pp. 145–152 (2016). https://doi.org/10.1145/2993318.2993337
8. Ghorbel, A., Ghorbel, M., Jmaiel, M.: PRIARMOR: An IaaS solution for low-level privacy enforcement in the cloud. In: 2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). pp. 119–124 (2017). https://doi.org/10.1109/WETICE.2017.64
9. Gurrin, C., Smeaton, A.F., Doherty, A.R.: LifeLogging: Personal big data. Foundations and Trends in Information Retrieval **8**(1), 1–125 (2014). https://doi.org/10.1561/1500000033
10. Hartmann, S., Ma, H., Vechsamutvaree, P.: Providing ontology-based privacy-aware data access through web services. In: Advances in Conceptual Modeling, pp. 74–85. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25747-1_8
11. Jacquemard, T., Novitzky, P., O'Brolcháin, F., Smeaton, A.F., Gordijn, B.: Challenges and opportunities of lifelog technologies: A literature review and critical analysis. Science and Engineering Ethics **20**(2), 379–409 (2014). https://doi.org/doi.org/10.1007/s11948-013-9456-1
12. Kandogan, E., Roth, M., Schwarz, P., Hui, J., Terrizzano, I., Christodoulakis, C., Miller, R.J.: LabBook: Metadata-driven social collaborative data analysis. In: 2015 IEEE International Conference on Big Data (Big Data). pp. 431–440 (2015). https://doi.org/10.1109/BigData.2015.7363784
13. Khan, A., McKillop, I.: Privacy-centric access control for distributed heterogeneous medical information systems. In: 2013 IEEE International Conference on Healthcare Informatics. pp. 297–306 (2013). https://doi.org/10.1109/ICHI.2013.42
14. Kwon, O., Lee, Y., Sarangib, D.: A Galois lattice approach to a context-aware privacy negotiation service. Expert Systems with Applications **38**(10), 12619–12629 (2011). https://doi.org/10.1016/j.eswa.2011.04.050
15. Martínez, S., Valls, A., Sánchez, D.: Semantic anonymisation of categorical datasets. In: Advanced Research in Data Privacy, pp. 111–128. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-09885-2_7
16. Mendes, R., Vilela, J.P.: Privacy-preserving data mining: Methods, metrics, and applications. IEEE Access **5**, 10562–10582 (2017). https://doi.org/10.1109/ACCESS.2017.2706947
17. Musen, M.A.: The Protégé project: A look back and a look forward. AI Matters **1**(4), 4–12 (2015). https://doi.org/10.1145/2757001.2757003
18. Nethravathi, N.P., Rao, P.G., Desai, V.J., Shenoy, P.D., Venugopal, K.R., Indiramma, M.: SWCTE: Semantic Weighted Context Tagging Engine for privacy preserving data mining. In: 2016 International Conference on Data Science and Engineering (ICDSE). pp. 1–5 (2016). https://doi.org/10.1109/ICDSE.2016.7823968
19. Rodriguez-Garcia, M., Batet, M., Snchez, D.: A semantic framework for noise addition with nominal data. Knowledge-Based Systems **122**, 103–118 (2017). https://doi.org/10.1016/j.knosys.2017.01.032
20. Sacco, O., Breslin, J.G.: PPO & PPM 2.0: Extending the Privacy Preference Framework to provide finer-grained access control for the Web of Data. In: Proceedings of the 8th International Conference on Semantic Systems. pp. 80–87 (2012). https://doi.org/10.1145/2362499.2362511

21. Saripalle, R.K., De la Rosa Algarin, A., Ziminski, T.B.: Towards knowledge level privacy and security using RDF/RDFS and RBAC. In: Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015). pp. 264–267 (2015). https://doi.org/10.1109/ICOSC.2015.7050817
22. Sicuranza, M., Ciampi, M.: A semantic access control for easy management of the privacy for EHR systems. In: 2014 Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing. pp. 400–405 (2014). https://doi.org/10.1109/3PGCIC.2014.84
23. Tsarkov, D., Horrocks, I.: FaCT++ description logic reasoner: System description. In: Automated Reasoning. Third International Joint Conference, IJCAR 2006. pp. 292–297 (2006). https://doi.org/10.1007/11814771_26
24. Zhan, J.: Privacy-preserving collaborative data mining. IEEE Computational Intelligence Magazine **3**(2), 31–41 (2008). https://doi.org/10.1109/MCI.2008.919071