HADA - An Access Controlled Application for Publishing and Discovering Linked Government Data*

Owen Sacco[1], George Thomas[2], and John G. Breslin[1]

[1] Digital Enterprise Research Institute, Galway, Ireland
{owen.sacco}@deri.org, {john.breslin}@nuigalway.ie
[2] U.S. Department of Health and Human Services, Washington D.C. - USA
{george.thomas1}@hhs.gov

Abstract. This paper presents the on-going HADA project, an application for publishing and discovering Linked Data about IT Investments and Assets of the U.S. Government Department of Health and Human Services (HHS). The application extracts, structures and links IT Investment and Asset data residing in various HHS internal repositories which currently are data silos that do not interconnect with each other and require the use of different systems to search and consume this information. Moreover, this project incorporates fine-grained access control for granting or restricting access to specific parts of the data since some data sources are public by nature but other sources are sensitive and require specific authorisation for consuming the data. Hence, in this paper we provide insights to the various challenges and lessons learned whilst developing this project that can be beneficial for publishing and consuming Linked Government Data with access control, especially for publishing sensitive personal citizen information.

Keywords: Semantic Web Applications, Linked Government Data, Access Control, LDA, PPO, PPM, HADA

1 Introduction

The Linked Open Government Data initiative encourages governments to publish their datasets which are publicly and freely available to all citizens. Data.gov[3] and Data.gov.uk[4] are two of many governmental Web portals that publish their datasets using the Linked Open Government Data (LOGD) approach [1], [4], [6], which adhere to the Linked Data best practices [2]. These practices overcome issues of data integration and reusability since they provide guidelines on how to structure data in a standard and open way.

* This work is funded by the Science Foundation Ireland (grant SFI/08/CE/I1380 (Líon 2)) and by an IRCSET scholarship co-funded by Cisco systems.

[3] `http://www.data.gov/`

[4] `http://data.gov.uk/`

This initiative is also proving to be a vital means of communication between governments and citizens as more datasets become publicly available [3]. This makes the government more transparent since citizens can now access the "raw data" and develop their own applications, by also linking to other datasets that enhances the quality of published data.

However, there are datasets that are still not yet published because they contain personal sensitive citizen data that cannot be publicly available. Also there is data that requires a level of authorisation in order to have access to it.

We have overcome these issues by developing an application – HADA: HHS IT Asset Discovery Application – that leverages the advantages of publishing Linked Open Government Data and also adds access control to sensitive data that can be filtered based on authorisation. In this way, with our methods, we encourage governments to not only publish open data but also publish access controlled sensitive data. We apply our methods on the U.S. Government Department of Health and Human Services (HHS)[5] IT Investments and Assets information, which gives us insights and set of best practices to be re-used when publishing and consuming Linked Government Data with access control.

The rest of the paper is organised as follows: section 2 gives an overview of the HHS infrastructure for storing and consuming IT Investments and Assets information. Section 3 outlines the problems with the current infrastructure and discuses the necessary requirements for publishing and discovering access controlled Linked Government Data. Section 4 provides a detailed description of the methods and APIs used to develop the HADA project. Section 5 provide some related work of some access control models. In section 6 we provide our conclusions and insights to this project that might be interesting for governments publishing access controlled Linked Government Data.

2 US Government Department of Health and Human Services (HHS)

The aim of the US Government Department of Health and Human Services (HHS) is to improve the well-being by providing and protecting the health of all American people; by providing all essential human services; and by fostering strong, sustained advances in medicine, public health and social sciences.

The HHS department focuses around three core domains: (1) Health and Human Services; (2) Scientific Research; and (3) Administrative and Management. Furthermore, two cross-functional domains: (1) Shared Services and (2) IT and Security Infrastructure are the foundation domains which are responsible for the effective operation of the core domains. Collectively, these domains support and perform collaborative planning, analysis and decision making for all the department of HHS.

Currently, the data about IT Investments and Assets utilised by all HHS core domains reside in various systems which each have their own repositories or set

[5] `http://www.hhs.gov/`

of documents; namely in the (1) HHS Enterprise Architecture (EA); (2) Capital Planning and Investment Control (CPIC); and (3) Enterprise Performance Life Cycle (EPLC) framework.

3 Current Problems and Requirements

In section 2 we described an overview of the core domains and current infrastructure for storing HHS IT Investments and Assets information. The infrastructure consists of different systems that each contain their own repositories which have their own data models for storing IT Investment and Asset information. Hence, there is no standard data model across all the systems which requires applications to be specifically customised to use a particular repository. The repositories are therefore data silos that are not interoperable with each other and also cannot be used by other Departments. Moreover, external data sources cannot be linked to this data since there is no standard data structure and hence, the systems cannot provide more information other than what is stored in the repositories.

Another problem is that each system has different access control levels since not all the data should be easily accessible due to the sensitive nature of the information. Hence, not all users are granted access to the same data and although they might have access to a particular IT Investment or Asset in one system, it does not necessarily mean that they would have access to the same IT Investment or Asset in another system.

3.1 Requirements

The following outline the requirements for the HADA project that would solve the above problems.

Extracting and Structuring Data Sets. Due to the walled garden surrounding the data silos that make the data sets not interoperable with each other, system specific and cannot be used by external sources; HHS requires means to extract and structure the datasets in RDF so that each information about the same IT Investment and/or asset residing within different data sources could be interoperable and can be accessed from one system rather than from different systems (since RDF provides this advantage). Whilst structuring the data, more metadata can be added to give more meaning to the data which could be easily consumed by machines other than users. All the data should be structured as dereferencable URIs to allow for easy discovery of related resources.

Publishing Data. The advantages of structuring the data in RDF is that the data can be consumed by machines and published in order to leverage the advantages of linking to other Linked Datasets (both internally within the government and from the Web at large) which adds more information about that IT Investment and Asset. Moreover, by publishing the data, other datasets can make use of this information that enhances other government services. However, due to the sensitive nature of this information, care should be taken at how this data is accessed; which makes this project different than other Linked Open Government Data projects.

Browsing Data. An interface is required to provide users with an aggregated view of the information concerning the same IT Investment and/or Asset instead of having to use different systems to acquire all the information of that Investment or Asset. Moreover, the interface should provide users to discover more information about the IT Investment or Asset by navigating through the URIs.

Searching for Information. The application should provide search functionality so that the information is provided on request to the user rather than having to browse or navigate through the data when the user requires a specific IT Investment and/or Asset.

Fine-Grained Access Control. Unlike Linked Open Government Datasets, these datasets contain sensitive information that cannot be publicly available and access should be controlled. Since the datasets have different levels of access, then a fine-grained access control model is required. Moreover, in order to enforce the access control policies; users are required to authenticate themselves in order to identify who they are. Once the user is authenticated, the access control policies are then enforced to filter the data. Filtering data means that when the RDF data is requested, only those triples that a user can access are provided to the user. The following are the access control requirements for HHS:

- P1 Grant/restrict triples stored in a particular *data source* (since some datasets are public by default whereas others require access control);
- P2 Grant/restrict a triple having a particular *subject*;
- P3 Grant/restrict a triple having a particular *predicate*;
- P4 Grant/restrict a triple having a particular *object*; and
- P5 Grant/restrict a triple.

Moreover, since it is hard on the Web to know beforehand who will access the data, it is not feasible to have pre-defined user lists that can access the data; instead the application would restrict users who have the necessary authorisation to access the data. Therefore, an attribute-based access control model is required where users have to satisfy specific attributes to be granted or restricted access to the data.

4 The HADA Project

In this section, we describe all the underlying mechanisms and APIs which we used to develop the HADA project – `http://hprod.dyndns.org/` – HHS IT

Asset Discovery Application; an application to publish and discover access controlled U.S. Department HHS IT Investment and Asset information. The HADA project was developed as a proof of concept to demonstrate that by using a set of APIs, the user can take advantage of Linked Open Government Data sets and also Linked Government Data filtered based on access control policies which safeguard sensitive information from being accessed by non-authorised users. Therefore, the innovative aspect of the HADA project is the combination of Linked Government Data and fine-grained access control policies that demonstrate how sensitive information can still be published without being "openly" accessible. Thus, this will encourage governments to not only publish non-sensitive information, but also to publish sensitive information to which access is controlled based on privacy policies.

4.1 Architecture

Figure 1 depicts a high level overview of the HADA project. The lower part of the architecture consists of the data layer; the HHS repositories where the system-specific HHS IT Investment and Asset data is stored. These are pre-extracted and structured in RDF using HADA vocabularies in the content extraction layer. The data is structured as N-Quads[6] since the source (where the data was originally stored) is stored as the *context*. The quads are stored in an OpenRDF Sesame RDF repository.

The upper part of the architecture consists of the HADA interface, whereby the user interacts with the HHS IT Investment and Asset information. When the user requests particular information, the query is sent to the Privacy Preference Manager API which is responsible for enforcing the privacy policies. The Privacy Preference Manager API checks the privacy preferences (see section 4.6) and retrieves only those triples which the user is allowed to access. The filtered triples are sent back to the interface; which are then displayed.

4.2 Data Extraction and Mapping

In the preliminary stages of the project, we first extracted, mapped and linked all the IT Investment and Asset data into RDF. Therefore, we created a "snapshot" of the datasets, converted them to RDF and stored the data in the OpenSesame RDF store. The data are mapped to properties based on the column names, as suggested by [3] and [13]. This is also because within the HHS there are documents and standards on how the data models should be structured. In this way, the extraction and mapping to concepts is done automatically without having to do intensive manual extraction and mapping.

In our prototype, we use `http://hprod.dyndns.org` as our *base_uri*. For every entity in the data, we create dereferencable URIs in the following manner: `{base_uri}/{hada}/{entity_name}/{entity_id}`; where *entity_name* is for instance *ITInvestment* and *entity_id* is a unique identifier example *"123"*. When

[6] N-Quads – `http://sw.deri.org/2008/07/n-quads/`

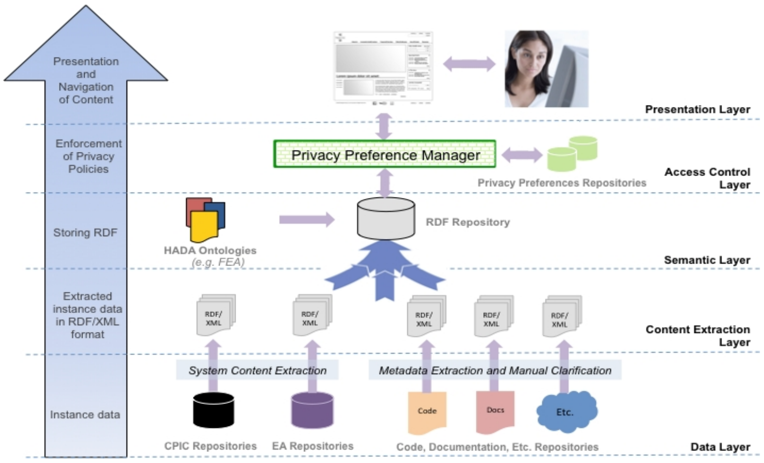

Presentation and Navigation of Content
Enforcement of Privacy Policies
Storing RDF
Extracted instance data in RDF/XML format
Instance data
Presentation Layer
Privacy Preference Manager
Privacy Preferences Repositories
Access Control Layer
HADA Ontologies (e.g. FEA)
RDF Repository
Semantic Layer
RDF/ XML
RDF/ XML
RDF/ XML
RDF/ XML
RDF/ XML
Content Extraction Layer
System Content Extraction
Metadata Extraction and Manual Clarification
Code
Docs
Etc.
CPIC Repositories
EA Repositories
Code, Documentation, Etc. Repositories
Data Layer

Fig. 1. HADA - High Level Architecture

the user dereferences the URI in the browser, will return the data for that particular *entity_id*.

Since the project is still in a prototype stage, the "snapshot" RDF data is not updated but once the prototype project is complete and the project goes live, then another conversion of the data will take place to have an updated version of the data and it will be used as the primary data source.

4.3 Publishing Data

The Linked Data API[7] is used to publish RDF data. It is a RESTful API over RDF graphs that acts as a proxy over SPARQL endpoints. The Linked Data API provides the following features:

- Generates documents for publishing Linked Data;
- Provides data querying and data abstraction without the user having to enter SPARQL queries; and
- Provides different data formats such as RDF/XML, turtle and json.

For this project, the Puelia API[8], a PHP implementation of the Linked Data API, is used since the project was developed using PHP. This API is mainly used to handle the incoming user requests by reading configuration files which in turn converts these requests into SPARQL queries which retrieve the RDF data from SPARQL endpoints (defined in the configuration file). Hence, when a user clicks on a dereferencable URI, the request is handled by this API and it

[7] Linked Data API (LDA) – `http://code.google.com/p/linked-data-api/`

[8] Puelia API – `http://code.google.com/p/puelia-php/`

returns a generated document containing the information of that particular IT Investment or Asset.

However, we have modified the sequence of which the Linked Data API handles the request in order to add fine grained access control. When a request is made, the Linked Data API handles the request by converting it into the SPARQL query, but instead of executing the query on the SPARQL endpoint, the query is sent to the Privacy Preference Manager (PPM) so that the privacy preferences are checked. Once the PPM enforces the privacy preferences and retrieves the filtered RDF data, the filtered RDF data is sent back to the Linked Data API which in turn generates the information document based on this filtered result set.

4.4 Browsing and Discovering Information

The Linked Data API, generates information documents that facilitates the development of User Interfaces for presenting Linked Data, in this case Linked Government Data. Figure 2 shows the HADA Welcome page that displays the various categories of IT Investments and Assets. Whenever a user clicks on a URI of an entity, the Linked Data API, after the Privacy Preference Manager enforces the privacy preferences, generates an information document similar to figure 3 which contains all the information retrieved from the RDF store. Each property and object contained in that page are dereferenceable URIs so whenever the user clicks on those URIs, more information about that URI can be discovered. Moreover, the Linked Data API provides various formats for representing the Linked Government Data which can be retrieved from links on the top right hand side of the generated information document. Whenever the user requests a different format, the Linked Data API first sends the query to the Privacy Preference Manager to enforce the privacy preferences before it provides back the formatted RDF data.

4.5 Searching Data

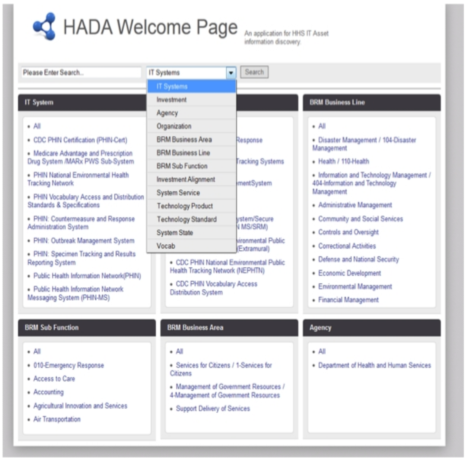
The HADA application uses the SIREn API[9] [8] for indexing and searching RDF data. This API is a Lucene plugin that efficiently indexes and queries RDF, as well as any textual document with an arbitrary amount of metadata fields. After the SIREn API searches and retrieves back the result set, the result set is sent to the Linked Data API which generates the information document.

The HADA application provides various types of search, including: (1) matching all of the words; (2) matching the exact search phrase; and (3) matching any of these words.

4.6 Fine-Grained Access Control for HHS Linked Government Data

In this section we present the Privacy Preference Ontology (PPO) and the Privacy Preference Manager (PPM) which provide the creation and enforcement of

[9] SIREn – `http://siren.sindice.com/`



IT System

- All
- CDC PHIN Certification (PHIN-Cert)
- Medicare Advantage and Prescription Drug System /MARx PWS Sub-System
- PHIN National Environmental Health Tracking Network
- PHIN Vocabulary Access and Distribution Standards & Specifications
- PHIN: Countermeasure and Response Administration System
- PHIN: Outbreak Management System
- PHIN: Specimen Tracking and Results Reporting System
- Public Health Information Network(PHIN)
- Public Health Information Network Messaging System (PHIN-MS)

Response

Tracking Systems

mentSystem

ystem/Secure
N MS/SRM)

ironmental Public
(Extramural)

- CDC PHIN National Environmental Public Health Tracking Network (NEPHTN)
- CDC PHIN Vocabulary Access Distribution System

BRM Business Line

- All
- Disaster Management / 104-Disaster Management
- Health / 110-Health
- Information and Technology Management / 404-Information and Technology Management
- Administrative Management
- Community and Social Services
- Controls and Oversight
- Correctional Activities
- Defense and National Security
- Economic Development
- Environmental Management
- Financial Management

BRM Sub Function

- All
- 010-Emergency Response
- Access to Care
- Accounting
- Agricultural Innovation and Services
- Air Transportation

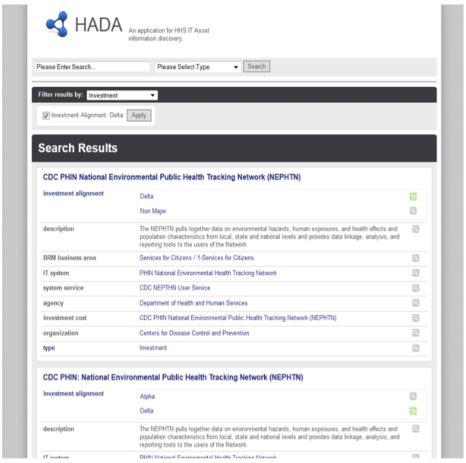
BRM Business Area

- All
- Services for Citizens / 1-Services for Citizens
- Management of Government Resources / 4-Management of Government Resources
- Support Delivery of Services

Agency

- All
- Department of Health and Human Services

Fig. 2. HADA - Welcome Page



☑ Investment Alignment: Delta

Search Results

CDC PHIN National Environmental Public Health Tracking Network (NEPHTN)

investment alignment	Delta Non Major
description	The NEPHTN pulls together data on environmental hazards, human exposures, and health effects and population characteristics from local, state and national levels and provides data linkage, analysis, and reporting tools to the users of the Network.
BRM business area	Services for Citizens / 1-Services for Citizens
IT system	PHIN National Environmental Health Tracking Network
system service	CDC NEPTHN User Service
agency	Department of Health and Human Services
investment cost	CDC PHIN National Environmental Public Health Tracking Network (NEPHTN)
organization	Centers for Disease Control and Prevention
type	Investment

CDC PHIN: National Environmental Public Health Tracking Network (NEPHTN)

investment alignment	Alpha Delta
description	The NEPHTN pulls together data on environmental hazards, human exposures, and health effects and population characteristics from local, state and national levels and provides data linkage, analysis, and reporting tools to the users of the Network.
IT system	PHIN National Environmental Health Tracking Network

Fig. 3. HADA - Browsing a Resource

fine-grained access control policies for the HHS IT Investment and Asset Linked Government Data stored in multiple sources as specified in the requirements in section 3.1.

The Privacy Preference Ontology (PPO). The Privacy Preference Ontology (PPO) [11] - `http://vocab.deri.ie/ppo#` - is a light-weight Attribute-based Access Control (ABAC) vocabulary that allows users to describe fine-grained privacy preferences for restricting or granting access to non-domain specific Linked Data elements. Among other use-cases, PPO can be used to restrict part of Linked Government Data records only to users that have specific attributes. It provides a machine-readable way to define settings such as "Provide a particular IT System Investment record only to IT System personnel" or "Grant write access to a particular IT Asset only to IT Asset personnel".

As PPO deals with RDF(S)/OWL data, a privacy preference, defines: (1) the resource, statement, named graph, dataset or context it must grant or restrict access to; (2) the conditions refining what to grant or restrict (for example defining which resource as subject or object to grant or restrict); (3) the access control type; and (4) a SPARQL query, (`AccessSpace`) *i.e.* a graph pattern representing what must be satisfied by the user requesting information. The access control type is defined by using the extended Web Access Control (WAC)[10] vocabulary which defines the `Create`, `Read` and `Write` (which also includes `Update`, `Delete` and `Append`) access control privileges.

As can be noted, the PPO satisfies all the HADA access control requirements as specified in section 3.1.

[10] WAC — `http://www.w3.org/ns/auth/acl`

Privacy Preference Manager (PPM). The Privacy Preference Manager (PPM)[11] [10], is a privacy preference manager for the Web of Data. Our aim is to illustrate how PPO can be applied to create privacy preferences for the HHS IT Investment and Asset data; and also how Linked Government Data, in particular the HHS IT Investment and Asset data can be filtered based on those preferences. The PPM is an API that can be accessed by the HADA application but also contains a Web interface that users can interact with it independently from the HADA application.

The PPM allows users (in particular administrators) to manage the privacy preferences and also grants or denies access to Linked Government Data when requested by users. Using the PPM, users can (1) authenticate to the HADA PPM instance and create privacy preferences for IT Investment and Asset data, only if they are administrators (as defined in a configuration settings file for the PPM); and (2) authenticate to the HADA PPM instance and access the filtered IT Investment and Asset data based on the privacy preferences.

The architecture of the PPM consists of: (1) WebID Authenticator: handles user sign-on using the FOAF+SSL protocol [12]; (2) Structured Data Retriever and Parser: retrieves and parses RDF data from SPARQL Endpoints or from URIs or data passed directly to the manager from other applications such as HADA; (3) Privacy Preferences Creator: for defining privacy preferences described using PPO; (4) Privacy Preferences Enforcer: queries the RDF data store to retrieve and enforce privacy preferences; (5) User Interface: provides users the environment whereby they can create privacy preferences and to view filtered IT Investment and Asset data; and (6) RDF Data store: an ARC2[12] RDF data store to store the privacy preferences[13].

Creating Privacy Preferences for HADA. The system's interface allows users to create privacy preferences. Users first authenticate using the WebID protocol. Once the PPM identifies that the user is an administrator, it provides the user with two text boxes; one to enter the SPARQL query and another to enter the location of the SPARQL endpoint. With this feature, the administrator can create privacy preferences on data residing in SPARQL endpoints. Once the PPM retrieves the result set from the SPARQL endpoint, the interface displays the data and the user can select/enter: (1) on which data source or context the privacy preferences are going to apply to; (2) to which named graph, resource or statement the privacy preference applies to; (3) the conditions for the privacy preference (such as defining which property the privacy preference applies to); (4) a SPARQL query that users when requesting the data must satisfy; and (5) the access control privilege. Once the choices are validated, the corresponding PPO preferences are created and stored in the system.

[11] Screencast online – `http://vmuss13.deri.ie/ppmv2/screencast/screencast.html`

[12] ARC2 — `http://arc.semsol.org`

[13] Although ARC2 was used for the implementation of PPM, any RDF store can be used.

Figure 4 shows an example of a privacy preference created for HADA, which reads: apply the privacy preference when the resource `http://hprod.dyndns.org/hada/Investment/90000001` is requested; grant the `Read` access control privilege to users who are interested in `Assets`.

```
PREFIX ppo: <http://vocab.deri.ie/ppo#> .
PREFIX hada: <http://hprod.dyndns.org/> .

hada:pp1 a ppo:PrivacyPreference;
 ppo:appliesToResource
  <http://hprod.dyndns.org/hada/Investment/90000001>;
 ppo:hasAccess acl:Read;

 ppo:hasAccessSpace
  [ ppo:hasAccessQuery
    "ASK {?x foaf:topic_interest
        <http://hprod.dyndns.org/hada/vocab/Asset>}"].
```

Fig. 4. Example of a privacy preference defined using PPO

Requesting and Enforcing Privacy Preferences in HADA The sequence in which privacy preferences are requested and enforced consists of: (1) a requester authenticates to HADA using the WebID protocol[14]; (2) the requester requests a particular IT Investment or Asset in HADA; (3) the Linked Data API (integrated in HADA) converts the request into a SPARQL query which is passed to the privacy preference manager; (4) the privacy preference manager queries the privacy preferences to identify which preference applies to the request; (5) the access space preferences are matched according to the requester's profile to test what the requester can access; (6) the requested HHS IT Investment and Asset is retrieved from the SPARQL endpoint based on what can be accessed; and (7) the requester is provided with the data s/he can access.

5 Related Work

The eXtensible Access Control Markup Language [7] is an XML based language for expressing a large variety of access control policies. Although the XACML is widely used, it does not provide the necessary elements to define fine-grained access control statements for structured data. It also does not contain enough semantics to describe what the actual access restriction is about and also does not semantically define which attributes a requester must satisfy.

[14] For demonstration purposes, the WebID authentication is disabled but we provide 3 users "already authenticated" by means of a drop down box

The Web Access Control (WAC) vocabulary[15] describes access control privileges for RDF data. This vocabulary defines the `Read` and `Write` access control privileges (for reading or updating data). This vocabulary is designed to specify access control to the full RDF document rather than specifying access control properties to specific data contained within the RDF document.

The Platform for Privacy Preferences (P3P)[16] specifies a protocol that enables Web sites to share their privacy policies with Web users. This platform does not ensure that Web sites act according to their publicised policies. Moreover, since this platform aims to enable Web sites to define their privacy policies, it does not solve our aim of enabling users to define their own privacy preferences.

The authors in [5] propose a privacy preference formal model consisting of relationships between objects and subjects. The proposed formal model relies on specifying precisely who can access the resource and therefore, our approach provides a more flexible solution which requires the user to specify attributes which the requester must satisfy.

The authors [9] propose a similar access control vocabulary and manager that uses SPARQL queries to test requesters whether they satisfy specific attributes. However, their model applies only to named graphs, unlike our model which we apply to statements and resources; hence providing finer-grained access control. Moreover, this model does not provide properties to specify which specific dataset to apply the rules; does not provide nested logical operators and does not provide the negation operator.

6 Discussion and Conclusion

In this paper we present HADA, an access controlled application for publishing and discovering U.S. Department of HHS IT Investments and Assets information. With HADA, we provided insights how Linked Government Data can be extracted, structured and published in standard formats such as RDF that enables linking between multiple data sources that provide enriched information to government stakeholders. We also demonstrated how we leverage the integration of the Linked Data API + the Privacy Preference Ontology + the Privacy Preference Manager to provide a platform for publishing access controlled Linked Government Data.

From this experience, we have learnt that HHS are emphasising a need for ‘securing data, not just the device’. Also, from this project, we examine that as other access control models approaches are generally user and request path oriented, the PPO/PPM approach gives the ability to be driven by data models. There are other well known approaches that also begin with RDF/OWL and leverage XACML, as described in section 5, but SPARQL ASK seems a more ‘native’ and flexible approach for the HHS.

The advantages for HHS to use Semantic Web technologies are the possible WebID augmentation of X.509 certificates in existing federal government Identity

[15] WAC — `http://www.w3.org/ns/auth/acl`

[16] P3P — `http://www.w3.org/TR/P3P/`

Management (IdM) initiatives and deployments [17] and the incredible flexibility of the role and attribute specification in a service requestors graph at their URI coupled with the policy declared as an instance of the PPO against any generic or domain specific data model. However, the disadvantages for HHS to use Semantic Web technologies are that WebID is very new and no one knows about it, and similarly, Linked Data is still new to the federal government at large, so there's very little awareness of it, much less understanding and preference.

In conclusion, HHS wished not only to correlate disparate stores of information across it's own myriad of agencies that each have a piece of the decentralised view of the puzzle that connects all the health domain entities but also to enable a way for the network effect to be applied to HHS data from other related open government data domains, such as the environment, food, etc. The impact could be transformative in realising the goals of shifting the cost of care from volume to value.

References

1. T. Berners-Lee. Putting government data online.
2. C. Bizer, T. Heath, T. Berners-Lee. Linked Data - The Story So Far. In *International Journal on Semantic Web and Information Systems*, 2009.
3. L. Ding, T. Lebo, J. S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. G. Zheng, Z. Shangguan, J. Flores, D. L. McGuinness, and J. Hendler. Twc logd: A portal for linked open government data ecosystems. In *Journal of Web Semantics*, 2011.
4. H. Alani, D. Dupplaw, J. Sheridan, K. O'Hara, J. Darlington, N. Shadbolt, C. Tullo. Unlocking the potential of public sector information with semantic web technology. In *ISWC/ASWC*, 2007.
5. P. Kärger and W. Siberski. Guarding a Walled Garden Semantic Privacy Preferences for the Social Web. *The Semantic Web: Research and Applications*, 2010.
6. L. Ding, D. Difranzo, A. Graves, J. Michaelis, X. Li, D.L. McGuiness, J. Hendler. Twc data-gov corpius: Incrementally generating linked government data from data.gov, in. In *WWW'10 (developer track)*, 2010.
7. Oasis. eXtensible Access Control Markup Language (XACML) Version 3.0. 2009.
8. R. Delbru, S. Campinas, G. Tummarello. Searching web data: an entity retrieval and high-performance indexing model. In *Journal of Web Semantics*, 2011.
9. S. Villata, N. Delaforge, F. Gandon, A. Gyrard. Social Semantic Web Access Control. In *Procs of the 4th International Workshop Social Data on the Web, SDoW2011*, 2011.
10. O. Sacco and A. Passant. A Privacy Preference Manager for the Social Semantic Web. In *Proceedings of the 2nd Workshop on Semantic Personalized Information Management: Retrieval and Recommendation, SPIM2011*, 2011.
11. O. Sacco and A. Passant. A Privacy Preference Ontology (PPO) for Linked Data. In *Proceedings of the Linked Data on the Web Workshop, LDOW2011*, 2011.
12. H. Story, B. Harbulot, I. Jacobi, and M. Jones. FOAF + SSL : RESTful Authentication for the Social Web. *Semantic Web Conference*, 2009.
13. T.Lebo, G.T. Williams. Converting governmental datasets into linked data,. In *6th International Conference on Semantic Systems, I-SEMANTICS'10*, 2010.

[17] See `http://www.gsa.gov/portal/content/105233`