# Addressing exploitability of Smart City data

Enrico Daga, Mathieu d'Aquin, Alessandro Adamou, Enrico Motta
The Open University,
Milton Keynes, United Kingdom
Email: {enrico.daga, mathieu.daquin, alessandro.adamou, enrico.motta}@open.ac.uk

***Abstract*—Central to a number of emerging Smart Cities are online platforms for data sharing and reuse: Data Hubs and Data Catalogues. These systems support the use of data by developers through enabling data discoverability and access. As such, the effectiveness of a Data Catalogue can be seen as the way in which it supports 'data exploitability': the ability to assess whether the provided data is appropriate to the given task. Beyond technical compatibility, this also regards validating the policies attached to data. Here, we present a methodology to enable Smart City Data Hubs to better address exploitability by considering the way policies propagate across the data flows applied in the system.**

## I. Introduction

Smart Cities can be seen as composite systems in which information rapidly flows through multiple devices and data silos [7, 10]. The amount of data to be managed is rapidly increasing, together with the scenarios, use cases and applications that rely on such shared data. Our work is placed within the context of the MK Data Hub [4], the data sharing platform of the MK:Smart project[1], which explores the use of data analytics to support Smarter Cities, taking the city of Milton Keynes in UK as a testbed. The main purpose of the MK Data Hub is to support applications that combine different city data in innovative scenarios. Since the data as a result of the MK Data Hub APIs might be combined from diverse datasets, different parts of the data might have different exploitability conditions or requirements, propagated from the licences and policies associated with the original datasets. Data consumers might need to check which original sources of the data need to be acknowledged because of an attribution requirement, and even whether the form of exposure or re-distribution they employ is allowed according to the policies attached to each individual piece of data they might obtain from the Data Hub. The issue of exploitability is therefore one that directly relates to providing the right level of information regarding the rights and policies that apply to the data being delivered by data hubs. However, while technologies that can help the representation and reasoning on policies do exist, they are not part of current data cataloguing approaches. The core contribution of this paper is a methodology through which the administrators of a Smart City Data Hub can support exploitability with state-of-the-art technical solutions.

In the next Section we introduce the problem, using the MK Data Hub as a case study. Section III surveys the state of the art, showing that, while technologies that can help the representation and reasoning on policies exist, they are not part of current data cataloguing approaches. Section IV presents the methodology, which we evaluate in the MK Data Hub case study in Section V, where we show how such methodology can be implemented in the design of a Data Catalogue through existing technical and non-technical resources.

[1] see http://mksmart.org

## II. The MK Data Hub

A Smart City **data hub** is an infrastructure that manages a wide range of data sources and methods of delivering them, with the aim of providing users with services that rely upon data taken from the sources it manages. The MK:Smart project aims to provide citizens and companies with access to a wide range of data sources about the city of Milton Keynes (MK). These data sources include sensor data, public data extracted from the Web as well as data provided by public institutions and other organisations, such as the Milton Keynes Council. These data sources, however, come with a set of policies regulating their usage. For example, the "Bletchley and Fenny Stratford" ward is a British electoral division that corresponds to an area in the South of the city. Located within this ward are a number of sensor devices that push data of varied nature to the Data Hub, including Air quality and Soil moisture (see an example in Figure 1a). The National Museum of Computing is located in Bletchley Park, and it is often a topic of interest in social platforms like Flickr. The Milton Keynes Council provides the MK Data Hub with statistics about population growth, crime, marital status, religion and employment, among others. All these data sources are catalogued, consumed and stored as *datasets* by the Data Hub in order to provide the end-user with services that intensively rely upon these data. One of these services is the *Entity-Centric API* (ECAPI) of the MK Data Hub. The ECAPI offers an entity-based access point to the information offered by the Data Hub, aggregating data from multiple data sources around 'real world entities' such as geographical regions, buildings, bus stops etc [1]. The aforementioned ward (see Figure 1b for some example data) and museum in Milton Keynes are examples of named entities the ECAPI may be queried for. More in general, any arbitrary geographical area within a fixed radius of given geospatials coordinates (e.g. `51.998,-0.7436` in decimal degrees) could be an entity for an application to try to get information about (see Figure 1c for example data). The ECAPI will return a collection of items that are relevant for that location, selected from the appropriate datasets. However, the parts of the returned data have been collected (and processed) from

978-1-5090-1846-8/16/$31.00 ©2016 IEEE


(a) Water Level Gauge - Bletchley - E21505. (b) Example of a British ward. (c) Example of a geographical point.

Fig. 1: MK:Smart Data Hub

sources that have different usage policies. This makes the exploitation of the data problematic.

Using the MK Data Hub as an example of multi-source Smart City data infrastructure, we consider data **exploitability** to specify the compatibility of the policies attached to the delivered data – obligations, permissions and prohibitions – with the requirements of the user's task. The problem we aim to address is: *How to support the end-user in assessing exploitability for data offered by a Smart City Data Hub?*

## III. State of the Art

We record very limited support for exploitability assessment in existing data cataloguing approaches. Systems like *CKAN*, one of the best-known data cataloguing platforms, and *Dataverse*[2], support the attachment of a license to datasets. *Socrata*[3] supports the specification of roles and permissions for the management workflow, while data terms of use are exclusively in human-readable form[4]. The DC subschema for rights and licenses is incorporated in the *DCAT* standard of the W3C for the representation of the catalogue meta-level[5]. The *HyperCat* specification follows a similar notion, however it enforces the use of URIs for values and contemplates machine-readable content as a possible form to which they dereference[6].

The Creative Commons consortium publishes guidelines for describing permissions, jurisdictions and requirements on works in general[7]. Specifically for data, the Open Data Institute has proposed the *ODRS* vocabulary[8], which addresses license compatibility and introduced the separation between data and content in the application of licenses. The *ODRL Policy Language* made the leap from licenses to policies, by introducing policy inheritance and profile[9]. Coupled with these are the online repositories of licenses expressed in RDF, including *LicenseDB*[10], which uses a mostly in-house vocabulary, and the Linked Data license repository of the Universidad Politécnica de Madrid[11], which uses ODRL. Under the reasonable assumption that the policies used to assess exploitability are formulated with the expressivity of ODRL, policy reasoning is reduced to a problem of policy compatibility, which is extensively studied in literature [5]. A form of policy reasoning is called *policy propagation*; in it, Policy Propagation Rules (PPR) are defined as Horn clauses on top of ODRL. We refer to this study as the reference method to manage a database of PPRs, in which the evolution of the requirements is tackled with an iterative process to compress the rule base and refine the ontological description of the actions involved [3].

The many apects of provenance, or *lineage*, were summarised in the W7 ontological model [8], although most of the existing work addresses provenance as the description of data origins. Our reference model for the representation of provenance is the W3C Recommendation *Prov-O*[12].

This paper adopts the notion of Supply Chain Management [6]. By lifting the metaphor with data as the materials and metadata as the information, we abstract from the complexity of sub-problems like *data integration*, *metadata storage* or *policy management*. We have gathered from this abstraction and the above survey that, to the best of our knowledge, there is no end-to-end solution for exploitablity assessment today.

## IV. Data Cataloguing as a Metadata Supply Chain

In our proposal, such a solution is implemented within a *data cataloguing system* as an essential element of the Data Hub. We propose here a methodology to develop such an end-to-end solution, whose role is to clarify: a) what is the general

[2]Dataverse, http://dataverse.org
[3]Socrata, http://www.socrata.com
[4]Example at the time of writing: https://opendata.camden.gov.uk/api/views/6ikd-ep2e.json
[5]DCAT, http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/
[6]HyperCat specification, http://www.hypercat.io/standard.html
[7]Creative Commons rights language, https://creativecommons.org/ns
[8]Open Data Rights Statement Vocabulary, http://schema.theodi.org/odrs
[9]W3C ODRL community, https://www.w3.org/community/odrl/
[10]LicenseDB, http://licensedb.org
[11]Linked Data licenses, http://oeg-dev.dia.fi.upm.es/licensius/rdflicense/
[12]Prov-O, http://www.w3.org/TR/2013/REC-prov-o-20130430/

life-cycle of the data within a Smart City data hub; b) what are the actors involved in such a process; c) what are their goals and tasks; d) what resources are needed, when and how they can be acquired and managed; and e) what operations have to be supported, in order for the exploitability assessment to be performed; finally, f) what are the requirements for the methodology to be applied.

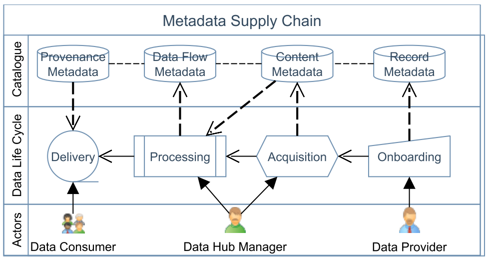


Fig. 2: Metadata Supply Chain: overview.

The methodology that is introduced in this Section supports what we call "Metadata Supply Chain Management" (MSCM), and is based on a *Data Catalogue*.

Figure 2 gives an illustration of the elements of the methodology and their interaction. The primary requirement of our methodology is that a Data Catalogue exists, and is a shared resource on which all the different actors and phases rely. There are 3 types of actors involved in the methodology. A *Data Provider* aims to publish a new data source in order to provide value to the task of a given *Data Consumer*. A *Data Hub Manager* has the role to supervise the infrastructure in terms of configuration, maintenance and monitoring. Our methodology follows a *Data life-cycle*, which comprises four phases:

- **Onboarding**: data sources are registered with the Data Hub;
- **Acquisition**: data are imported in the Data Hub;
- **Processing**: data are processed, manipulated and analysed in order to generate a new dataset, targeted to support some data-relying task;
- **Delivery**: resulting data are delivered to an external system/end-user.

The Metadata Supply Chain Management (MSCM) activity follows the Data Life Cycle in parallel. In the following paragraphs we provide the details of each phase, focusing on:

- the *objectives* that need to be reached;
- the *roles* of the actors in this phase;
- the required *resources* to be managed;
- *operations* to be performed at the different stages;
- what are the *output* resources of each phase; and
- what are the *requirements* that need to be satisfied.

Tables I-IV detail each component of each phase in the methodology, and serve as a guide to its implementation in concrete use cases.

**Phase 1. Onboarding**. The Onboarding phase is dedicated to acquiring and managing information about data sources. When

TABLE I: Onboarding

| | |
|---|---|
| **Objectives** | Obtain information about a data source |
| **Roles** | A Data Provider and a Data Hub Manager |
| **Resources** | A Data Catalogue, including a Licenses Database, and a data source |
| **Operations** | Registration of the data source in the Data Catalogue. |
| **Output** | Structured information about the data source in the form of a Catalogue Record. |
| **Requirements**<br>*1.1*: The Data Provider associates a single License to the data source.<br>*1.2*: The License is granted to whoever exploits the given data source.<br>*1.3*: The License is described in the Licenses Database.<br>*1.4*: Policies are set of binary relations between a deontic component (permission, prohibition, requirement) and an action.<br>*1.5*: Policies are referenced by Policy Propagation Rules (PPRs), part of the Licenses Database. | |

TABLE II: Acquisition

| | |
|---|---|
| **Objectives** | Access the data source and collection of the related data. |
| **Roles** | The Data Hub Manager supervises and monitors the relevant procedures. |
| **Resources** | A Catalogue Record, containing information about how to access the data. |
| **Operations** | Collection of the data, inspection and eventually storage in a staging environment. |
| **Output** | Content Metadata, ready to be exploted by the required processes. |
| **Requirements**<br>*2.1*: The data source is accessible.<br>*2.2*: Acquisition is performed by respecting the data source License. | |

a *Data Provider* wishes to publish a new dataset, the Data Hub has to provide the required facility to do that. From the point of view of the *Data life-cycle*, in this phase the provider registers a new data source (or modifies an existing one) in the Data Catalogue, that is the space where dataset descriptions are managed. The Data Catalogue manages metadata about the data source as a *Catalogue Record*, following the W3C DCAT specification[13]. This description includes details about how the dataset will be populated, and more importantly includes information about ownership (`dc:creator`) and licensing (`dc:license`), as well as attribution statement. The onboarding process requires that the licensor states a single license (Requirement 1.1), applicable to whoever exploits the given data source (Requirement 1.2). The terms and conditions of the data sources are supposed to be in the set of the available licenses in a Licenses Database (Requirement 1.3). This includes the requirement that licenses be described as a set of ODRL policies [9], each one specified as a binary association between a deontic component and an action (Requirement 1.4), for example *requirement+attribution*, or *prohibition+commercial_use*. Existing policies are included in the set of Policies Propagation Rules (PPR) [3], also part of the Licenses Database.

**Phase 2. Acquisition**. After onboarding a new data source, the data need to be acquired by the data hub. "Acquiring" means that the data hub is given a means to control the delivery cycle of the data whose awareness was granted through the onboarding phase. It is the role of a *Data Hub Manager* to supervise this process and monitor the acquisition, including implementing the needed strategies for data update and quality

[13]DCAT, http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/

control. This activity can be rather complex as it may include automatic and supervised methods, and going into the details of it is out of scope for this article[14]. What is important for us is that this phase should provide a sufficient amount of *metadata* in order to support data processing. *Content Metadata* (see Figure 2) refers to topical and structural information that might be established by accessing the actual data[15], to support the configuration of integration strategies by the Data Hub Manager. This phase is based on the assumptions that the data source is actually accessible by the Data Hub (Requirement 2.1) and that acquisition is possible according to the data source license (Requirement 2.2).

**Phase 3. Processing**. In this phase the data are manipulated in order to fulfil a given task that relies upon them. This activity can be seen as supporting a traditional ETL[16] task. *Content Metadata* include information about the data sources in order to support the configuration of these processes, whether it is an automatic method or a process supervised by the *Data Hub Manager*. However, here we focus on the metadata that the data processing phase must produce in order for the Data Hub to support the user in the assessment of data exploitability. Metadata about possible *processes* should be collected and stored in the catalogue, in order to allow reasoning on policy propagation [3], and to attach the required policies to the resulting dataset. Processes can be described as relations between data objects (Requirement 3.1). This is the approach followed by Datanode [2]. Therefore, ETL pipelines can be annotated with *data flow descriptions* as representation of the processes using Datanode, allowing to execute Policy Propagation Rules (PPRs) and determine what policies can be attached to the output of each process [3]. In a general case, the *Data Hub Manager* is responsible for providing such information, as well as assessing that the processing itself is made respecting the policies of the data sources (Requirement 3.2). Moreover, these metadata should provide an abstract representation of the process so that, once combined with the actual input (a given data catalogue record and content metadata), it would be possible to generate the relevant policies. In other words, a given data flow description should be valid for all possible executions of a process (Requirement 3.3).

**Phase 4. Delivery**. In this phase data are delivered to the end user or application. The Data Catalogue provides the required metadata to be distributed alongside the process output. Delivered data should include provenance information such as ownership, attribution statement and policies (permissions, requirements, prohibitions). Delivered metadata should be included in the provenance information (Requirement 4.2), in order to support the user in assessing the data exploitability for the task at hand[17]. Once the metadata reach the end-user, the exploitability task is indeed reduced to the assessment of the compatibility between the actions performed by the user's application and the policies attached to the datasets, with an approach similar to the one presented in [5], for example using the SPIN-DLE reasoner[18].

TABLE III: Processing

| | |
|---|---|
| **Objectives** | Obtain a description of the ETL process suitable to reason on policies propagation. |
| **Roles** | The Data Hub Manager to configure the processes and produce descriptions of the data flows. |
| **Resources** | A Catalogue Record linked to Content Metadata. Processing will need to exploit the former or the latter, on a case by case basis. |
| **Operations** | Processes must be described as networks of data objects relying on the Datanode ontology. |
| **Output** | Data flow descriptions to be registered in the Data Catalogue. |
| **Requirements**<br>*3.1*: Processes can be described as data flows with Datanode.<br>*3.2*: ETL processes do not violate the License of the source.<br>*3.3*: Process executions do not influence policies propagation. | |

TABLE IV: Delivery

| | |
|---|---|
| **Objectives** | Deliver the set of policies associated with the data as part of the provenance information. |
| **Roles** | The Data Consumer. |
| **Resources** | Catalogue Record, Data flow metadata, Policy Propagation Rules base |
| **Operations** | Reason on PPRs given the data flow description and the rule base. |
| **Output** | Set of policies attached as part of the provenance information of the returned data. |
| **Requirements**<br>*4.1*: Data flow descriptions and License policies enable reasoning on Policy Propagation Rules.<br>*4.2*: End-user access method includes provenance information. | |

## V. Evaluation

Our hypothesis is that an *end-to-end* solution for exploitablity assessment can be developed by using state-of-the-art techniques through the implementation of the methodology introduced so far. We now validate this statement by describing the solution developed in the MK Data Hub, following the scenario introduced in Section II.

Figure 3 illustrates the components and their role in the data and metadata life-cycle of the MK Data Hub. The Onboarding phase is the initial step of our methodology, and it is supported by providing an *input interface* to *Data Providers*, implemented as a Data Hub Portal page and a Web API. Following our sample use case, some data sources are registered in the Data Catalogue. They are *Air Quality and Moisture Sensors* in the Bletchley area, the *Flickr API* (including a number of images annotated with geocoordinates associated with the ward), the *UK Food Estanblishments Info and Ratings API*, as well as topographical information exposed by the *Ordnance Survey* and *statistics* from the Milton Keynes Council. Each one of these data sources have a single license associated (R 1.1), applicable to whoever makes use of the

[14] For example, data sources could be registered as web accessible resources (via HTTP or FTP), Web APIs, or uploaded files. Methods for acquisition can include collecting resources from external systems or requiring an ingestion API to be exposed.

[15] For example the types of the entities included in the content, the set of attributes, local and global identifiers (and their structure or format), relations and references to external datasets, as well as statistics about them.

[16] Extract, transform, load (ETL), https://en.wikipedia.org/wiki/Extract,_transform,_load.

[17] It is worth noting that the actual assessment of compatibility between the user's task and the policies of the output data is not part of this methodology, and is left to the end user.

[18] SPIN-DLE, http://spin.nicta.org.au/spindle/index.html

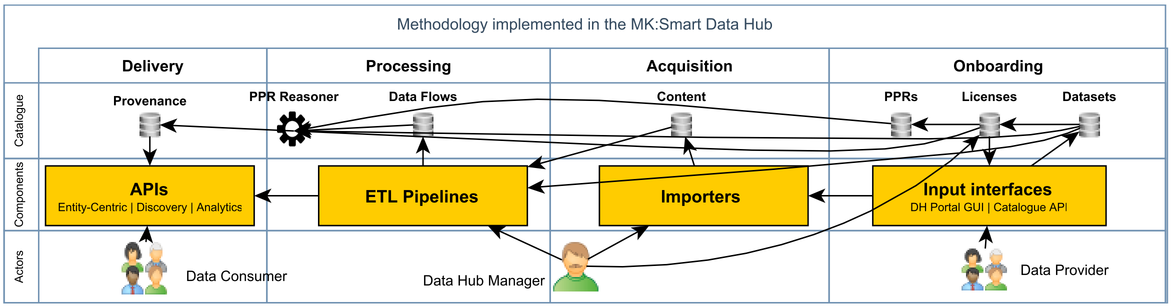


Fig. 3: MK Data Hub overview. The figure shows the phases of the methodologies and how they are supported by the MK Data Hub. The *Data Catalogue* is the component responsible for managing the *Metadata Supply Chain*, interacting with the other components of the system. On the right side of the image, *Data Providers* to register new data sources. A *Data Hub Manager* is responsible for the description of licenses, and supervises the activity of *importers* and ETL *pipelines*, including the curation of data flow descriptions (*Data Flows*) and policy propagation rules (*PPRs*). *Data Consumers* invoke APIs and associated *Provenance* information is provided from the *Data Catalogue*, exploiting a *PPR Reasoner* that relies on *Data Flows* descriptions and *PPRs*.

data (R 1.2), and described in RDF/ODRL in a *Licenses Database* (R 1.3, see also Figure 3). For example, the metadata about the *Water Level Gauge - Bletchley - E21505* data source is one of the relevant data sources for the area. Figure 1a shows the Data Catalogue record as presented in the MK Data Hub web portal. As shown in Listing 1, the related description includes a reference to the *Open Government License*, described following Requirement R 1.4 (see Listing 2). These policies have related PPRs in the Licenses Database (R 1.5). It is the role of the Data Hub Manager to provide the necessary descriptions in the License Database.

Listing 1: Dataset: Water Level Gauge - Bletchley - E21505: RDF description.

```
:water-level-gauge-bletchley-e21505
 a dcat:Dataset ;
 dc:title "Water Level Gauge - Bletchley - E21505" ;
 mks:owner "Environment Agency" ;
 mks:policy policy:open-government-license ;
...
```

Listing 2: Open Government License: policy set

```
:open-government-license
 odrl:permission [ a odrl:Permission;
  odrl:action odrl:derive ,odrl:distribute ,
    ldr:extraction ,odrl:reproduce ,odrl:read ,ldr:reutilization;
  odrl:duty [odrl:action odrl:attachPolicy ,odrl:attribute]] .
```

Data sources like the *Flickr API* come with peculiar terms and conditions[19] (Listing 3). Some of them refer to the usage of the API, others to the assets the data are describing (like Flickr images). In these cases the Data Hub Manager limits the descriptions to the policies that are applicable to the accessed data, and describe them in the Licenses database. The description always include a reference to the document from which the policies have been extracted.

Listing 3: Flickr TOS

```
mks:flickrtos odrl:prohibition [ a odrl:Prohibition;
    odrl:action odrl:sell ,odrl:sublicense ,cc:CommercialUse]
  odrl:duty [ odrl:action odrl:attribute ];
  mks:attributionStatement "This product uses the Flickr
       API but is not endorsed or certified by Flickr." ;
  dct:source <https://www.flickr.com/services/api/tos/>
```

[19]Flickr: https://www.flickr.com/services/api/tos/

This applies also to other sources taken into consideration, like the *UK Food Estanblishments Info and Ratings*[20] as well as statistics from the MK Council, that come with an Open Government License. The result of this phase is a set of *Catalogue Records* supporting all the requirements of the initial phase of our methodology.

The Acquisition phase is the stage of the methodology that covers the execution of the processes required to populate the dataset from the sources, assuming they are accessible (R 2.1). This can be achieved in different ways in the MK Data Hub. For each type of source the data cataloguing system implements a dedicated *metadata extractor* with the objective to complement the *Dataset Record* with more metadata for supporting the data processing. For example, air quality and soil moisture sensors push regular streams of data in the Data Hub. The Flickr API is invoked on demand and information stored at query time in temporary datasets. During these processes, metadata about the geolocalisation of the related items are extracted and stored in the Data Catalogue. *Content Metadata* include the location of the Flickr images, while geocoordinates of the sensors are part of the *Dataset Record*. It is the responsibility of the Data Hub Manager to verify that the acquisition of the data is possible without violating the sources' terms of use (R 2.2). In the Processing phase, data are extracted, transformed and loaded (ETL) in datasets using dedicated *pipelines*. Each pipeline performs a number of operations on the data sources in order to select the relevant information and transform it in a format suitable for the task at hand. Listing 4 shows the description of the processing pipeline of a file data source from Milton Keynes Council, as configured by the Data Hub Manager. The file is downloaded from the remote location and a copy is stored locally in a staging area (see also Figure 3). The content is then transformed into RDF using the CSV2RDF approach[21]. After that, a SPARQL query remodels the data applying the W3C

[20]The dataset includes a snapshot of the food hygiene rating data published at http://www.food.gov.uk/ratings.

[21]http://www.w3.org/TR/csv2rdf/

Datacube Vocabulary[22] data model. These data are accessed by a SPARQL query, which selects a relevant portion of the data for the task at hand.

Listing 4: Processing pipeline for a CSV file.

```
:input a dn:Datanode; mks:format mks:csv; dn:hasCopy [
  dn:refactoredInto [
   mks:format mks:rdf; dn:usesSchema csvOntology:; dn:
       remodelledInto [
    dn:usesSchema qb:; dn:hasSelection :output ]]] .
```

The descriptions of the data flows are provided a single time by the Data Hub Manager following Requirement R 3.1. In this activity, the Data Hub Manager verifies that the ETL process is compliant with the source License (R 3.2). This model represents the process in an abstract way, and it is agnostic with respect to the actual input (R 3.3).

The Data Hub exposes a number of APIs to access the data in various forms. For example, sensor data can be extracted as streams by providing temporal constraints. The Entity Centric API is a specialised service for data discovery, that aggregates information summaries from several datasets about a given entity. In our running examples, a *Data Consumer* requests information about a location in Milton Keynes, in the form of geocoordinates: `51.998,-0.7436`. The PPR Reasoner will be queried providing the actual input as a specific dataset in the catalogue, according to the user's query (R 4.1).

Listing 5: Policy Propagation Rules.

```
propagates(dn:remodelledTo,duty cc:ShareAlike)
propagates(dn:hasSelection,duty cc:ShareAlike)
propagates(dn:hasCopy,duty cc:ShareAlike)
```

The dataflow description will be complemented by the related dataset record metadata and associated policies from the licenses database. Listing 5 shows a subset of the rules that are activated in relation to the dataflow (Listing 4) and policies set (Listing 2). The propagated policies are displayed in Listing 6.

Listing 6: Policies associated with the returned data processed from the original Milton Keynes council CSV file.

```
[] a dn:Datanode ;
    odrl:duty [odrl:action odrl:attachPolicy , odrl:
        attribute]
```

The output includes an aggregated view of items related to that geolocation as well as provenance information for each one of them, including the policies relevant to assess the exploitability of each item (R 4.2).

## VI. Conclusions

In this paper, we proposed a methodology for supporting data exploitability as the assessment of the compatibility of data policies with the task at hand, and validated this methodology through its implementation within the MK Data Hub platform. Future work includes the support of multiple licenses by enabling "scopes" of use as additional metadata, user profiling in order to add more contextual information to the reasoning process, and expanding the data flow descriptions phase to also support articulate processes by adding process execution traces as part of the description. In a complex environment like the one of a Smart City Data Hub, there might be other research questions related to policies and constraints with respect to the data sources, data flow and output, respectively. We plan to explore these questions further in an expanding framework for computationally handling data usage policies, of which the presented methodology is the foundation.

[22]http://www.w3.org/TR/vocab-data-cube/

## References


[1] A. Adamou and M. d'Aquin. On requirements for federated data integration as a compilation process. In *Proceedings of 2nd International Workshop on Dataset PROFIling and fEderated Search for Linked Data (PROFILES).*, pages 75–80, 2015.

[2] E. Daga, M. d'Aquin, A. Gangemi, and E. Motta. Describing semantic web applications through relations between data nodes. Technical Report kmi-14-05, Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, 2014.

[3] E. Daga, M. d'Aquin, A. Gangemi, and E. Motta. Propagation of policies in rich data flows. In *Proceedings of the 8th International Conference on Knowledge Capture*, page 5. ACM, 2015.

[4] M. d'Aquin, J. Davies, and E. Motta. Smart cities' data: Challenges and opportunities for semantic technologies. *Internet Computing, IEEE*, 19(6):66–70, 2015.

[5] G. Governatori, A. Rotolo, S. Villata, and F. Gandon. One license to compose them all. In *The Semantic Web–ISWC 2013*, pages 151–166. Springer, 2013.

[6] R. B. Handfield and E. L. Nichols. *Introduction to supply chain management*, volume 1. prentice Hall Upper Saddle River, NJ, 1999.

[7] T. Nam and T. A. Pardo. Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, pages 282–291. ACM, 2011.

[8] S. Ram and J. Liu. A new perspective on semantics of data provenance. In J. Freire, P. Missier, and S. S. Sahoo, editors, *Proceedings of the First International Workshop on the role of Semantic Web in Provenance Management (SWPM 2009)*, volume 526 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.

[9] S. Steyskal and A. Polleres. Defining expressive access policies for linked data using the ODRL ontology 2.0. In H. Sack, A. Filipowska, J. Lehmann, and S. Hellmann, editors, *Proceedings of the 10th International Conference on Semantic Systems (SEMANTiCS 2014)*, pages 20–23. ACM, 2014.

[10] A. M. Townsend. *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company, 2013.