Geographical Service: a compass for the Web of Data.

Gianluca Correndo, Manuel Salvadores, Yang Yang, Nicholas Gibbins, Nigel Shadbolt
Intelligence, Agents, Multimedia (IAM) Group
School of Electronics and Computer Science
Southampton, UK
{gc3, ms8, yy1402, nmg, nrs}@ecs.soton.ac.uk

ABSTRACT

This paper describes a Linked Data service that supports the navigation and retrieval of geographical entities for the UK territory. Geographical entities, in the extent of this paper, are linked data resources that describe objects that have a geographical extension. The service presented in this paper allows the querying of resources that contain or are contained by a given entity URI. The recent publication of UK Public Sector Information (PSI) data sets has brought to the attention of the community the redundant presence of location based context. At the same time it stresses the inadequacy of current Linked Data services for exploiting the semantics of such contextual dimensions for easing entity retrieval and browsing. We present an approach for a geography based service that helps in querying qualitative spatial relations for the UK geography (proper containment so far). We also provide an exploitation scenario based on a backlinking service and PSI Open Linked Data, published within the EnAKTing project.

Categories and Subject Descriptors

H.3.4 [**Systems and Software**]: Distributed systems; H.5.4 [**Web**]: Navigation; H.3.5 [**Online Information Services**]: Web-based services

General Terms

Linked Data, geographical services

Keywords

Linked Data, geographical reasoning, Web of Data

1. INTRODUCTION

The Linked Data Initiative represents the first collaborative effort to create a *Web of Data* (WoD henceforth) of considerable scale, providing few, simple guidelines for publishing content using well established standards [3]. Such guidelines and standards are leading the way to a new paradigm of interaction between government and citizens in the UK. In order to pursue better access for citizens to information held by local as well as national public organisations, the UK government has recently launched[1] a public initiative for publishing Public Sector Information (PSI), adopting Linked Data tenets as future best practices. Data sets recently delivered to the public include: government expenses, NHS trusts' performances, public transportation, and a whole set of statistics about crime, mortality, census, environment, school and social indicators. Some of the data sets mentioned have been published already in Linked Data format, others have been translated within the EnAKTing project, and many others are waiting to be freed in the LOD cloud.

Such a prolific inflow of Linked Data poses new questions and challenges to the community of researchers and developers: how is it possible to integrate such different information into a meaningful schema? How is it possible to exploit the little semantics that goes a long way? How do we choreograph the publishing activity of separate organizations from the public sector? A common trait of PSI seems to be its locality: local and national public organisations are in fact mainly concerned with the collection of data about their territory, and the distribution of their resources.

In the WoD vision, links between resources from different publishers are particularly important since they are the ones that allow new data to be discovered and integrated into the current discourse. It is frequently the case that different URIs are used to refer to the same things, motivating the use of co-reference services for the resolution of instance equivalences. Knowledge of this type of relationship increases the potential for reuse since information from previously unknown sources is now accessible, and makes the problem of co-reference resolution of primary importance [9]. In any case, we can expect more and more of this linking data to be made available as the number of Linked Data publishers increases.

The publication of an authoritative geography of the UK, (its regions, counties, districts and their connections) by Ordnance Survey (the national mapping agency for Great Britain, OS henceforth) as Linked Data, has opened interesting scenarios for exploiting semantics in contextualising the information sources published on `data.gov.uk`. The geographical dimensions in PSI data sets are already represented, but their semantics may be lost if they are not exploited for creating new collections of data, browsing related resources, and making connections.

In this paper, we present a service for querying spatial relationships for the UK (extensible to other countries when authoritative knowledge bases are available). We start in Section 2 where the available knowledge bases are described along with an introduction of the qualitative spatial reasoning supported. Section 3 provides a rationale for the

[1]Public access to the site `http://data.gov.uk` has been granted the 19^{th} of January, 2010.

LDOW2010, April 27, 2010, Raleigh, USA.

developing of such service in support of Linked Data browsing and retrieval. In Section 4 the implementation of the geographical service and its APIs are described. The paper then concludes with a description of an evaluation of the presented service using public sector information from the UK government in Section 5 and some concluding notes in Section 6.

2. BACKGROUND

The World Wide Web and the WoD can both be understood as hypertext systems, where the general purpose of the hypertext system is for information discovery by navigation. Providing reasoning over hyperlinks for the purpose of navigation can benefit information discovery. In 1990, Nanard brought the concept of "semantic network" from Artificial Intelligence [16] into the hypertext field by creating a Conceptual Hypertext System [13], in which a hyperlink can be reasoned by using a domain model classification. In the above system, *typed links* and *typed chunks* are used to define relationship between *types* in order to incorporate knowledge into a hypertext. This domain model classification is used to classify the documents and documents that share metadata, and which are deemed to be similar in some way. The Conceptual Open Hypermedia Service (COHSE) project [4] later took this approach forward by providing ontological reasoning based on links of services to bridge the navigation gap between the Web and Linked Data, where the link services provided a mapping between concepts and the lexical labels on the web page.

Many of the PSI data sets published so far can be plotted within a spatial and temporal dimension, in other words, all data can be linked together by its spatial and temporal indexes. Within this context, the need to provide services to reason the spatial and temporal aspects of the linked data is of key importance. This is unsurprising, the spatial and temporal reasoning have always been considered to be an important part of common-sense reasoning in Artificial Intelligence. In this section, we will mainly focus on qualitative spatial representation and reasoning. There are two major approaches to qualitative spatial representation - point based and region based [6]. Region based approaches, such as Topology [7] which describe relationships between spatial regions are more intuitive than point based approaches. The commonly known approaches for formalizing topological properties of spatial regions are based on work from Whitehead [17] and Clarke [5] who axiomatized mereotopologies (a theory that combines mereology and topology) using a single primitive relation and binary connectivity relationships. By using these primitive relations, other relations can be defined. The Region Connection Calculus (RCC8) proposed by Randell, Cui and Cohn[14] defines a set of jointly exhaustive and pairwise disjoint relations DC, EC, PO, EQ, TPP, NTPP, TPPi an NTPPi, as illustrated in Figure 1, and is the most well-known approach in the domain. Since the RCC Calculus is expressed in first-order predicate calculus, a wide range of theorem provers can be used for reasoning. For instance, Given a fixed vocabulary of relations, Ri, given R1(x,y) and R2(y,z), one can answer questions about the possible relations (from the set Ri) that can hold between x and z by looking up the composition table [8]. Although general 1st-order theorem proving is too inefficient to be useful for many purposes [11], it is relatively simple to implement and particularly useful in our case for reasoning are the geographic location relationships.

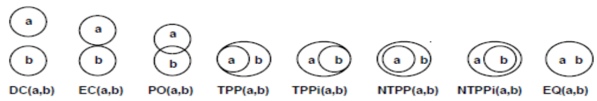

Figure 1: RCC Eight Jointly Exhaustive and Pariswise Disjoint Relations

Within the Linked Data context, there are several services providing resolvable URIs for geographic locations. Geo-names[2] for example, is a community based service that provides geographical representation of geographical entities covering all countries worldwide and manages eight million URIs for geographical resources. As a further example, the national mapping agency of Great Britain, Ordnance Survey, maintains a continuously updated database of the topography of Great Britain[3] and is responsible for surveying the boundaries of the administrative areas.

In this paper, we exploited the Administrative Geography ontology provided by Ordnance Survey as an authoritative knowledge base for querying the UK geographical structure [10]. Such ontology explicitly represents the mereological relationships within the administrative hierarchy, as well as topologically representing the boundary information between administrative units at the same hierarchical level. The following depicts the class hierarchy created in the administrative ontology from Ordnance Survey:

- CivilAdministrativeArea
  - EuropeanRegion
  - Country
  - UnitaryAuthority
  - MetropolitanDistrict
  - GreaterLondonAuthority
  - LondonBorough
  - District
  - CivilParish
  - Community
- Country

The topological relations adopted by this ontology were taken from the RCC8 and correspond to the properties NTPPi, TPPi, EC and EQ respectively. The topology of administrative geography of Great Britain contains no overlapping regions, therefore, the PO relation was not required. Later version of the ontology reported overlapping entities as well. The property of *spatial containment* used in the OS ontology (equivalent to the NTTP(i) and TTP(i) relations in Figure 1), implies a mereological relationship. For instance, if Hampshire *spatially contains* Fareham, then Fareham is a *part of* Hampshire.

[2]`http://www.geonames.org` last accessed 10/02/2010
[3]With the exception of Northern Ireland that is covered by a different agency, the Land and Property Services Northern Ireland.

Dereferenceable URIs adopted by the Linked Data community inherit the same properties of hyperlinks in the Web hypertext system, which is (among others) uni-directionality. The problem of such kind of links is that it is not possible to navigate back to the original resource by using dereferenciation mechanism only. This problem becomes even more relevant when URIs from previous authoritative data sets are reused in order to provide context and meaning to new data. It is in fact possible to browse from the new data to the old one, but not the other way around. The back-linking service[4] we have implemented for UK public sector information supports the discovery of back-links between datasets. The benefit of a back-linking service is that it enables users to discover, from a single dataset, other datasets which reference back to it, creating therefore data linkage opportunities between datasets, increasing the recall of valuable data sources, and doubling the network effect [15] that increases even more when co-reference systems are employed.

In this paper, we will mainly focus on exploring the possibility of exploiting semantics from authoritative knowledge bases to provide support for consuming Linked Data resources. The service provided will allow users to retrieve contained (and container) entity URIs from popular data sets by exploiting a co-reference service. Moreover, a back-linking service which we previously created in the EnAKTing project[5], will allow us to retrieve the information resources that addressed such URIs. Far from trying to provide a general purpose reasoner for geographical entities, the aim of the service described in the following sections is to exploit the semantically rich knowledge base for UK geography in order to ease users' navigation through the published PSI data sets. Similar capabilities were already provided by DBpedia Mobile [2], an application that retrieved DBpedia entries mashed up on a map based on users' geographical coordinates. The results provided by our service although are based on a spatial subdivision of the territory, subdivision that is already used by public sector organizations to classify their data (e.g. crime statistics are based on a police based subdivision of the territory, while MPs activities are related to the constituency they were voted in).

3. MOTIVATION

The Linked Data principles [3] promote a Web of Data whose architecture is inherently decentralised, relying on data already published (when available) in order to give semantics and context to new data. The growth the WoD has experienced over recent years relies on the simplicity of publishing and linking data. However, up to now a semantically coherent orchestration of data publishing is still a mirage. Nevertheless, relying purely on data linkage for the discovery and browsing of linked data resources would lead to a serious knot to untie in the near future. The use of ontologies and powerful ontology languages in publishing Linked Data will be an effort that must be justified against a scenario where such explicit semantics are rarely exploited.

In publishing UK Public Sector Information (UK PSI), we have identified an issue concerning data accessibility and navigability that addresses in particular the missing exploitation of semantics (in this case about qualitative spatial description of geographical entities). In this paper we present a solution to overcome such issue that soundly enhance data retrieval and browsing when geographical dimensions are involved.

The issue is about the usage of geographical entities for contextualising local information (i.e. information that are related to a particular geographical location, for example the population of a region, the MPs of a constituency, or various statistical data based on territory). In publishing this kind of information, we provided alignments of our data (at least for the geographical dimensions represented in the data) to authoritative knowledge bases using co-reference systems [9]. The problem we have to deal with originates with the fact that, since the public sector information published was originated by different sectors of UK government, the kind of spatial classifications used were highly heterogeneous, ranging from local parishes to counties and up to European regions (e.g. South East of England). The different granularities used to classify the data means, in Linked Data terms, that related information sources link to different URIs. Some data may be in fact relevant for constituencies, while others may use a different granularity (by county for example), and the URI of a county is obviously different from the set of URIs of all its constituencies. Available knowledge bases about the geographical or administrative subdivision of a territory can be exploited to cover such gap in data granularity.

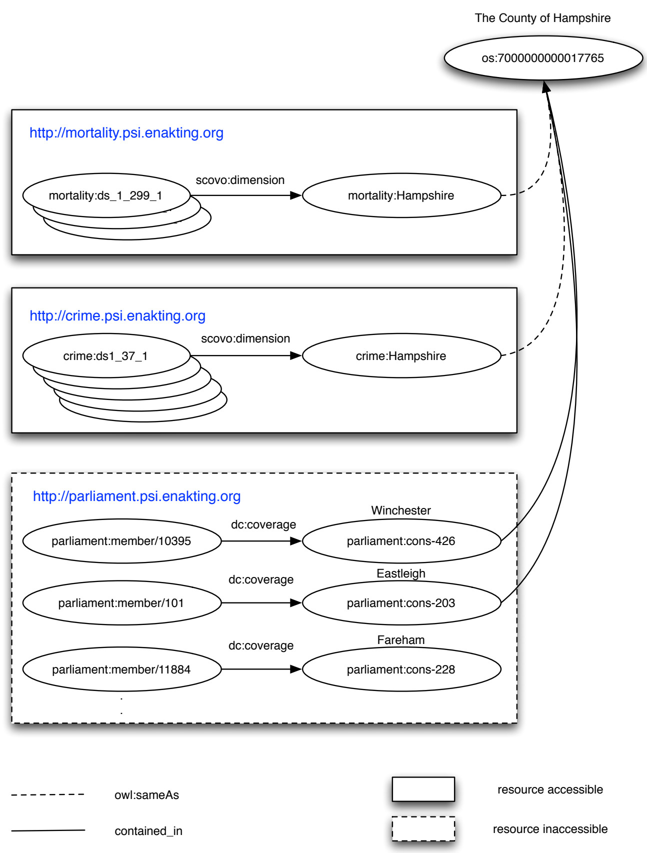

Figure 2: Resource irretrievable via geographical gap

Taking as an example the PSI data sets published re-

[4]http://backlinks.psi.enakting.org
[5]http://enakting.org

cently[6], we adopted the Ordnance Survey administrative ontology in order to provide context to our data items (i.e. SCOVO items instances[7] and local governmental data). The SCOVO ontology allows us to describe statistical data as a collection of *Items* where each item describes a statistical value (i.e. a single cell in a multidimensional table) along with all the dimensions that characterise it. In the case of UK PSI statistics, many data sets collected were related to geographical regions (counties, districts, etc.)

In this case, users who wished to discover useful information about their own region (e.g. the County of Hampshire, top Figure 2) would start their searching activity by browsing one of its available URIs. The OS URI for such geographical entity would be `os:7000000000017765`[8], but any equivalent URI provided by a co-reference system will provide the same results as will be described in the following. Using a backlinking service for resolving the entities linking to the given URI for Hampshire, we are able to retrieve links to mortality statistics (`mortality:ds1_299_[1...3]`[9]) and crime statistics (`crime:ds1_37_[1...11]`[10]). In Figure 2 those URIs are contained in boxes labelled as *"accessible"*, meaning that those URIs are retrievable following back already existent arcs. Those SCOVO data sets' items address in fact Hampshire county as one of their dimensions. What is missing is the further data collected that reports valuable information about regions contained in Hampshire. In particular, within the EnAKTing project, we published linked data about the singular constituencies too. In detail we published, for each of constituency, an historical record of the MP in charge for that constituency, his/her voting records and expenses. In Figure 2 those resources are contained in dotted boxes labelled as *"inaccessible"*, meaning that they cannot be retrieved with the existent knowledge.

Example URIs for such inaccessible resources are[11]:

```
parliament:cons-637 rdfs:label "Winchester"
parliament:cons-203 rdfs:label "Eastleigh"
parliament:cons-228 rdfs:label "Fareham"
```

The URIs for, respectively: Winchester, Eastleigh, and Fareham, are therefore not retrieved by the resolution of the Hampshire URI (obviously) or by the additional service provided from the backlinking service.

Despite the fact that an entity is still semantically different from the parts that compose it, the information relevant for all its constituting parts can still be relevant for the entity as a whole. Without covering such geographical gap it is not possible to access all the relevant sets of information, provide them to the user or process them in some way in order to summarise their content.

The aim of this research is to exploit authoritative knowledge bases in order to cover such gaps, allowing therefore citizens to retrieve information resources relevant to their region of interest. Moreover, there are many data sources that describes geographical resources, and all of those are already partially aligned. The integration of different knowledge bases could lead to the possible exploitation of such alignments in order to bridge data sets and reuse the available knowledge in more than one context.

4. GEOGRAPHICAL SERVICE FOR UK

To support the user's experience in browsing and discovery of new resources in the WoD, we have developed a geographical service for querying the UK territory structure. The decision to restrict the service to the UK territory is mainly due to the fact that the service is mainly used in order to support the discovery of UK PSI resources. Knowledge about geographical containment is exploited here to link information that is contextually related because of their spatial dimension.

For this use case we have implemented a service for querying the topological structure of UK (from the broader entity to the more particular and the other way around) that can be easily integrated into a web of linked data. The service, accessible at `http://geoservice.psi.enakting.org` is designed in order to be easily integrated both into web applications and in linked data resources and it follows few basic principles:

Lightweight Service : The service should be easy to use and resolve a specific problem. A geographical service is a component of the WoD that supports discovery when geographical entities are involved, it is not a general purpose reasoning engine.

Linked Data Compatible : The geographical service should be used as a resolvable URI like any other resource, in order to be used in linked data content as a useful provider of relevant URIs. Moreover the service should provide the results in a number of different formats that will be decided using content negotiation and `HTTP 303` redirection.

Co-reference Support : The service should exploit the already available knowledge about instance equivalence provided by co-reference systems[12] in order to return results useful in a number of different data sets.

4.1 Data collection and normalisation

OS provides an ontology[13] and an RDF dump about spatial relations between UK regions. The triples from OS have been parsed and only the relation of physical containments have been retained, normalised and completed with the inverse relations in a separate knowledge base. The service presented here, for the sake of simplicity and efficiency[14], manages only the NTPP, and the relative inverse, the NTPPi relations. The knowledge extracted from the OS data set has been then normalised in terms of an internal ontology that represent qualitative spatial relations.

The normalisation step has been introduced in order to allow the service to integrate further geographical hierarchies in the future (e.g. geonames provides containment of

[6] `http://browser.psi.enakting.org`
[7] `http://purl.org/NET/scovo`
[8] `PREFIX os:<http://data.ordnancesurvey.co.uk>`
[9] `PREFIX mortality:<http://mortality.psi.enakting.org/id/>`
[10] `PREFIX crime:<http://crime.psi.enakting.org/id/>`
[11] `PREFIX parliament:<http://parliament.psi.enakting.org/id/>`
[12] Like `http://sameas.org`
[13] `http://www.ordnancesurvey.co.uk/oswebsite/ontology/SpatialRelations/v0.2/SpatialRelations.owl`
[14] In this way the data to manage has been drastically reduced in order to provide a very focused service.

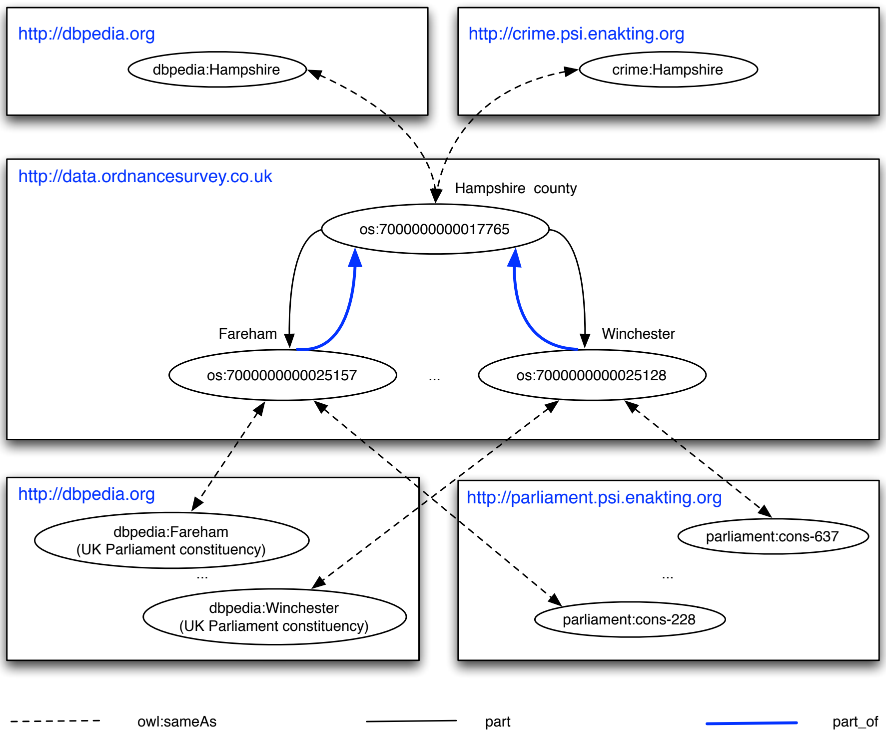

Figure 3: Coupling of co-reference and Ordnance Survey geographic ontology

geographical features). The future integration of qualitative spatial knowledge bases is devised in order to extend the service outside the borders of UK and for providing an assessment of co-references between geographical entities.

A simple example of how the normalised triples from OS ontology are used in coupling with a co-reference service for bridging the navigational gap for different data sets is depicted in Figure 3; in the figure it is possible to see that a single statement from OS describing the fact that the County of Hampshire **contains** Fareham and Winchester[15]:

```
os:7...17765 os:contains os:7...25157.
os:7...17765 os:contains os:7...25128.
```

has been translated into an internal representation containing both relations: part, and part_of; like the following:

```
os:7...17765 geoservice:part    os:7...25157.
os:7...25157 geoservice:part_of os:7...17765 .
os:7...17765 geoservice:part    os:7...25128.
os:7...25128 geoservice:part_of os:7...17765 .
```

The containment relations so normalised (see central part of Figure 3) are then internally stored in the system and queried for serving users requests.

The normalised containment relations are integrated with the information provided by the co-reference system that allows to bridge different data sources both in the input phase (i.e. where the input URI must be translated in the OS equivalent, see top part of Figure 3) and the output phase (i.e. when the results must be translated into a target data set provided by the user, see bottom part of Figure 3). The co-reference service used in this paper is the `http://sameas.org` service from Glaser et al. [9]. The relevant bundles have been retrieved from the service and cached for performance. It is important to note that, in order to chose the wanted quality of service, one could opt for using one co-reference service instead of another. The functionality provided is transparent from the provenance of the co-reference bundles.

[15]OS URIs are shortened, the trail of '0' are replaced by '. . . '.

Exploiting co-reference services and OS ontology, it is therefore possible to infer containment relation between resources from different data sets. For example:

```
dbpedia:Hampshire owl:sameAs os:7...17765
AND
os:7...17765 geoservice:part os:7...25128
AND
os:7...25128 owl:sameAs dbpedia:Winchester
⟹
dbpedia:Hampshire geoservice:part dbpedia:Winchester
```

4.2 RESTful API

The service is accessed via HTTP GET requests and provide two essential information: the list of entities **contained** the input URI, and the list of entities that **contains** the input URI. The interface is then accessible via the following URIs:

```
http://geoservice.psi.enakting.org/{command}/
  {dictionary}/{format}/{URI}
```

In the above API description, the parameters are enclosed in brackets and their meaning is the following:

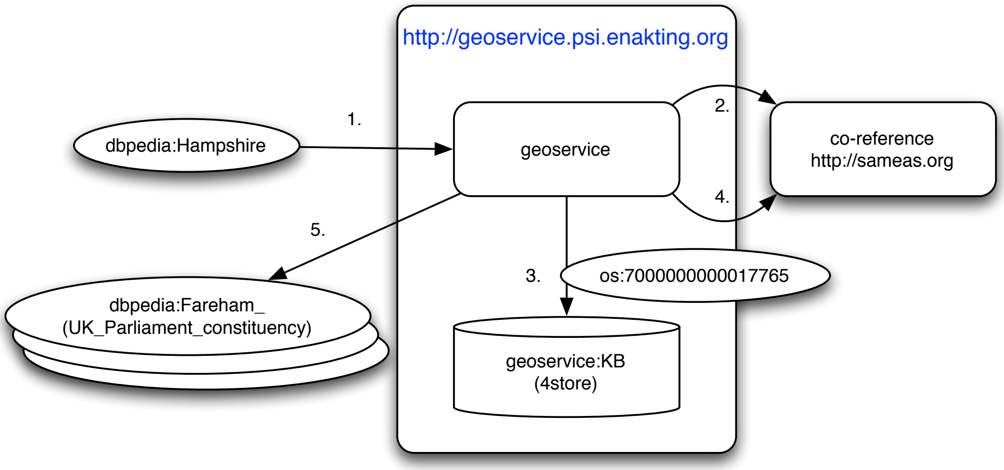

Figure 4: Overall architecture and interaction with co-reference system

command: can be either **contains** or **container**: in the first case it returns the URIs of the entities contained by the input URI; in the second case it returns the URIs of the entities that contains the input URI.

dictionary: can be one of the followings (**dbpedia**, **os**, **statistics**, **geonames**, **enakting**, **opencyc**, **openlylocal**, or **none**) and instructs the service to use the co-reference system in order to retrieve the equivalent URIs in the respective data sets (i.e. DBpedia [1], Ordnance Survey [10], UK National Statistics[16], Geonames[17], PSI enAKTing[18], OpenCYC [12], Openly Local project[19]). The value **none** is used for not applying any filter. In this case the URIs returned will be the ones from the Ordnance Survey plus the ones returned from the co-reference service.

format: the format parameter is optional and can be one of the followings (**rdf**, **text**, **ttl**, or **json**). The value of the **format** parameter decide then the format of the returned content: RDF/XML for **rdf**; list of URIs separated by new lines for **text**; RDF/Turtle for **turtle**; and finally JSON[20] for **json**. If the parameter is not given the right content is decided using the 303 HTTP redirection. Even for the content requests `Accept:text/html` done using the browser, the client is redirected to the HTML page of the service initialised with the input URI.

URI: is the URI of the input entity to query using the service. The service uses a co-reference system in order to find the equivalent URI for the Ordnance Survey and the Geonames data set. This means that the user can use one of the data set of preference (e.g. DBpedia or Geonames) and ask for contained, or container, entities in one of the desired target data set (e.g. again DBpedia, Geonames, or enAKTing published information).

[16]`http://statistics.data.gov.uk` last accessed 10/02/10
[17]`http://geonames.org` last accessed 10/02/10
[18]`http://browser.psi.enakting.org` last accessed 10/02/10
[19]Community devoted to provide linked data access for local government data, see `http://openlylocal.com` last accessed 10/02/10
[20]`http://json.org` last accessed 10/02/10

The service returns a list of URIs if the content type is **text** or **json**. The RDF content, for both **rdf** and **turtle**, describes the containment relations between the input URI and the resulting resources. In both cases the returned URIs are translated into the desired address space.

The procedure followed by the service, and an overall architecture, is depicted in Figure 4, and can be describe as follows:

1. user generated request (HTTP GET request)
2. normalisation of the input URI to OS
3. computation of the property closure (i.e. **part** or **part-of**) over the normalised URI
4. optional phase of translation and filtering of the resulting URIs to the target URI space
5. formatted content, as per user request, returned to the user (HTTP Response)

As an example, consider the case of a software client who needs to know all the geographical entities contained in the **Hampshire**, the request can adopt as an input one of many available URIs describing the **Hampshire** county, a popular choice could be the DBpedia URI (i.e. `URI = http://dbpedia.org/resource/Hampshire`). The agent can then explicit the desired target data set, for example the DBpedia data set itself (i.e. `dictionary = dbpedia`), and instruct the server to return the JSON format of the document (i.e. HTTP header contains `Accept:application/json`). The HTTP request will be then the following:

```
GET /contains/dbpedia/http://dbpedia.org/
     resource/Hampshire
Host: geoservice.psi.enakting.org
```

```
Accept: application/json
```

And the service will return a response redirecting the client to the right URL:

```
HTTP/1.1 302 Found
Location: http://geoservice.psi.enakting.org/
contains/dbpedia/json/http://dbpedia.org/
resource/Hampshire
```

That, once resolved, will finally return the desired content, a JSON array of strings that represents the URI of the DBpedia resource describing entities contained in Hampshire:

```
HTTP/1.1 200 OK
Content-Type: application/json

["http://dbpedia.org/resource/North_East
_Hampshire_%28UK_Parliament_constituency%29",
 "http://dbpedia.org/resource/East_Hampshire
 _%28UK_Parliament_constituency%29", ...
```

The client agent can obviously immediately refer to the right URL and retrieve the content in the right format straight away. A useful way to exploit such service can be seen when data sets other than OS one are queried. Not every data set in fact provides a clear semantic representation about mereological relations. This is due to the fact that the focus of many data set is to provide information about a particular region: encyclopaedic information from DBpedia, statistics information from the UK National Statistics, geographical features from Geonames[21], conceptual description from Open CYC, local government information from Openly Local, and UK PSI from EnAKTing.

Using the service presented in this paper is easy to exploit the OS administrative ontology in order to retrieve geographically relevant information regardless from the starting data set. As an example, let us consider the case where a user may want to retrieve information about local government of its own city, for example about Southampton, UK. The easiest thing to do is to start from a recognizable URI such as the DBpedia ones:

`http://dbpedia.org/resource/Southampton`

From this URI the user can retrieve general information about the city, even the names about some of the city *leaders*. No further information is available on the Southampton DBpedia page about local government information. Asking the geographical service to return the contained entities from the Openly Local site we can then retrieve more resources:

```
http://openlylocal.com/id/wards/4925
http://openlylocal.com/id/wards/4929
...
http://openlylocal.com/id/wards/4938
```

Those URIs are the ones published for each one of the wards present in the city of Southampton and provides not only the names of the local councillors but also some other statistics about the ward (i.e. demographics and religious statistics). Moreover, asking again the service for the DBpedia URIs we are able to retrieve the followings:

```
http://dbpedia.org/resource/Southampton_Test_
    %28UK_Parliament_constituency%29
http://dbpedia.org/resource/Southampton_Itchen_
    %28UK_Parliament_constituency%29
```

From those URIs we are then able to check then the identities of the MPs in charge (in the Southampton page from DBpedia their are mentioned both as leaders of the city whereas an MP is actually in charge only to its constituency where s/he has been elected. Asking then for the URIs from the data sets provided by the EnAKTing project we would be able to retrieve the followings:

```
http://parliament.psi.enakting.org/id/cons-536
http://parliament.psi.enakting.org/id/cons-535
```

Following such links the user would be able then to retrieve other information about the MPs from each constituency (even retrieving an historical record of them) and further information about their political activity.

5. EVALUATION

We have evaluated our geographical service from two different perspectives. The first one looks at the direct benefit that our backlinking service for Public Sector Information[22] would gain from expanding its navigability through geographic containments (see Section 5.1). The second evaluation is more analytic and looks at the new knowledge generated as part of the translation process from an authoritative geographic closure to the covered vocabularies (see Section 5.2) .

5.1 Backlinking Service Integration

This section studies the navigability improvement that our backlinking service for the PSI in the UK would experiment by plugging the containments from a wide range of vocabularies.The PSI Backlinking Service provides an access point to retrieve backlinks from Foreign URIs. Foreign URIs make data discovery difficult because it is not possible to navigate the RDF documents of the WoD bidirectionally. `http://backlinks.psi.enakting.org` provides an API to retrieve collections of backlinks for a given URI. The study of the covered knowledge bases[23] in the UK PSI Backlinking Service made explicit that one of the most highly connected data sets in the PSI WoD are the ones representing some type of geographic information.

In this evaluation we have used the Backlinking Service as a client of the Geographical Service in order to expand the backlinks that we can get from geographic resources. We have kept the decentralization nature of the Backlinking and Geographical services and basically the Backlinking Service performs HTTP requests to get the geographic containments (see Figure 5). When the geography extension is enabled the backlinking service gets the list of contained entities for the input URI and returns the backlinks connected to any URI part of containments. The request to the Geography Service is performed using "contains" as *command* JSON as *format* and "none" as *dictionary*. The selected dictionary is "none" because the Backlinking Service doesn't know before

[21]Geonames provides a containment relation that does not however reflect any administrative subdivision

[22]`http://backlinks.psi.enakting.org` last accessed 10/02/10

[23]`http://backlinks.psi.enakting.org#KBs` last accessed 10/02/10

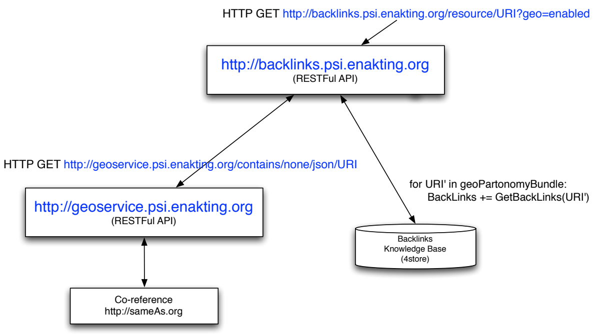

Figure 5: Interaction of the backlinking and geographical services

hand what type of URIs will be the source of backlinks for a certain geographic region. So as to improve the coverage we aim to get all the possible containments from all the dictionaries supported in the geographical service.

There is a natural outcome from this integration and it can be shown using how the systems works when asking for backlinks connected the URI *dbpedia:Hampshire*. Prior to the use of the geographical extension a request to retrieve backlinks for *dbpedia:Hampshire* would just give back 14 URIs related the UK region of *dbpedia:Hampshire* or any equivalent URI part of the same co-reference bundle in *sameAs.org* (see Figure 6). This same request when the geographical service is integrated returns the following additional backlinks:

- 6 010 resources that represent schools from `http://education.data.gov.uk`. These RDF documents represents the totality of education entities in the region of Hampshire.
- 42 mortality statistical resources from `http://mortality.psi.enakting.org`. This statistics are segmented by geography and gender.
- 981 CO_2 emission measurements from `http://co2emission.psi.enakting.org`. These resources represent the CO_2 emissions for the region of Hampshire between 2005 and 2007.
- 300 resources with information of energy consumption from `http://energy.psi.enakting.org`. This data sets publishes the energy consumption in the UK in respect to fuel in the road network between 2005 and 2007. These results represent all the RDF documents linked to geographical regions contained in Hampshire.
- 4 788 population census information segmented by age and sex from `http://population.psi.enakting.org`.
- 224 parliamentary identities from `http://parliament.psi.enakting.org`. These represent mandates for different members of the UK Parliament and House of Commons.

Figure 6 shows the output of the backlinking service with and without geographical extensions in the Backlinking Service. All the resources enumerated above are not specifically linked to `dbpedia:Hampshire` or equivalent URIs but to geographic containments of it in at least one of the data sets covered by the Geographical Service.

This scenario has shown one of possible scenarios where the exploitation of explicit semantic can improved the accessibility of the resources in the Web of Data. In esence the backlinking service is improving its graph connectivity by being aware of the new layer of Linked Data that the Geographical Service publishes via its RESTFul API. This case study also shows how different Linked Data RESTFul services (such as co-reference, backlinking and geographical services) can cooperate in a layer built on top of current Web of Data to improve its navigability.

5.2 Vocabulary Closure Coverage

The geographical service can be seen as an extra layer of linked data based on an initial geographic closure provided by Ordnance Survey and its extensions to other data sets via co-references. This extra layer of linked data is obviously an added value to the Web of Data. This section analyses the interlinking improvement between the data sets by means of number of triples produced by the Geographical Service.

Table 1 represents the amount of triples generated by our service in terms of number of triples that contain where the predicate is `geoservice:part` or `geoservice:part_of`. This table shows the numbers of triples linking every pair of data set in the system. For instance our Geographical Service has produced 30995 geographic containments between *dbpedia* and *mortality.psi.enakting.org*.

Of particular interest are the results from the *geonames* data set. In fact, the number of containment relations within such dataset is quite small compared to the number of containment relations provided by *geonames* itself (a rough estimate done by the authors counts about 9K relations). Such additional source of spatial knowledge open a scenario where the two knowledge bases can be compared and integrated for providing a better recall for the service. An important aspect to take into account in such a scenario would be the quality of the results computed by the integration.

The data seed that triggered this new knowledge is the OS to OS containments, 60M of statements. The total number of triples generated are 223M and these are partially interlinking every pair of data sets. Partially because the completeness of every pair of datasets' closure relies on the accuracy of the co-reference bundles extracted from *sameAs.org*. As the number of co-references from *sameAs.org* grows and improves its accuracy the Geographical Service will reflect those changes automatically. This side effects is one of key aspects of the Web of Data and its decentralized nature.

6. CONCLUSIONS

We have presented in this paper a service that helps users in browsing geographical resources from different datasets (dbpedia, geonames, data.gov.uk. psi.enakting.org, ...) by exploiting an authoritative ontology for the UK territory (Ordnance Survey). One of the novel aspects of this research is the use of a co-reference system (`http://sameas.org`) to extend the containments from one geographic data set to others where such containments are not so rich or complete. Moreover, the added value of integrating such geographical service with a backlinking service has been shown with respect to demonstrate a possible exploitation scenario on Public Sector Information. Due to the particular na-

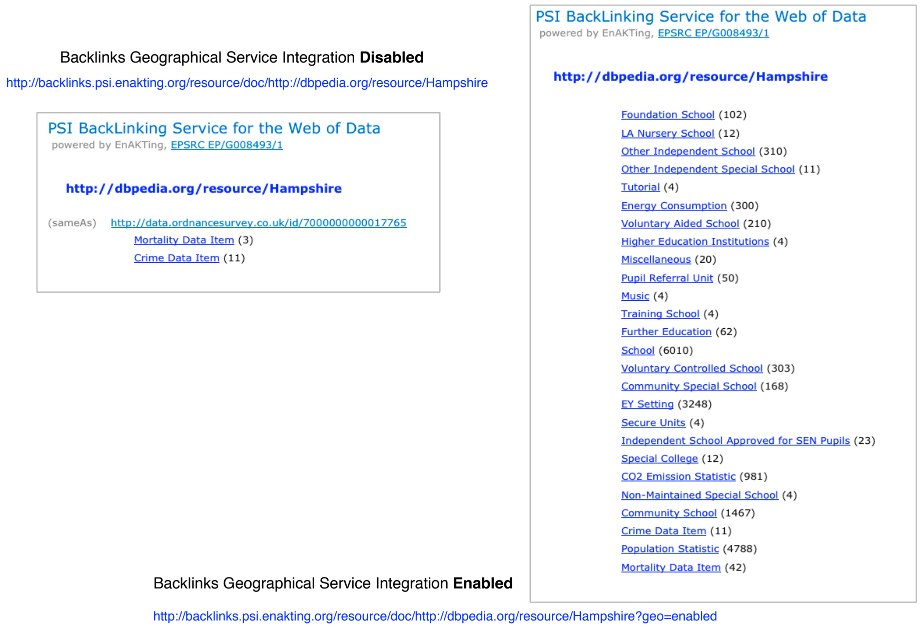

Figure 6: Output comparison for dbpedia:Hampshire with and without geographical partonomies

Table 1: Datasets linkage improvement statistics

	OS	**dbpedia**	**statistics**	**mortality**	**parliament**	**crime**	**geonames**	**openlylocal**	**opencyc**
OS	**60469910**	1757760	45354078	1035901	1338214	235906	94072	18559453	1106900
dbpedia	1757760	59640	1393322	30995	46077	9570	3035	540619	35250
statistics	45354078	1393322	36179867	813660	1056892	206217	71232	14430773	819965
mortality	1035901	30995	813660	19109	23929	4607	1488	344436	17415
parliament	1338214	46077	1056892	23929	37631	7654	2410	436883	28070
crime	235906	9570	206217	4607	7654	2249	334	82559	4160
geonames	94072	3035	71232	1488	2410	334	224	26427	2475
openlylocal	18559453	540619	14430773	344436	436883	82559	26427	6498462	312120
opencyc	1106900	35250	819965	17415	28070	4160	2475	312120	27975

ture of the knowledge provided (i.e. closure of geographical containment properties), there is the possibility of overwhelming the user with information when asking about top level features (e.g. `England`). In order to cope with this eventuality, the service will be provided soon with the capability to limit the results by depth. Therefore, when asked about all the entities contained in the top level feature `England` at the first level of depth, the service will return only: `North East`, `North West`, `South East`, `Eastern`, `South West`, `East Midlands`, `West Midlands`, `Yorkshire & the Humber`, `Scotland`, `Wales`, `London` (different from `the City of London`).

Another important aspect not tackled in this work, and subject of future research, is the temporal extent of administrative divisions. The version of administrative geography of UK will change shortly and has changed frequently during the years (e.g. the number and borders of constituencies are reviewed every 10 or 15 years). New entities can be defined, old ones can be abolished or change status. For example Southampton, once part of Hampshire, became a *Unitary Authority* on the 1^{st} of April 1997. Since then, Southampton has been administratively detached from the county of Hampshire (i.e. not contained any more), although being still part of it as a *ceremonial county*. Versioning of information resources is an hot topic in Linked Data community and it is even more important when publishing Public Sector Information, whose content and validity must be put into context.

The research work reported here tackles an important aspect of Linked Data, the exploitation of explicit semantic content for enhancing resource retrieval and browsability. The choice to tackle geographical knowledge rather than some other data facet is mainly due to the analysis of the available data sources, their structure and the available knowledge exploitable for a better integration of the available information.

The use of co-reference systems allowed us to exploit the knowledge created in one organization (Ordnance Survey in this case) in different, and potentially novel, data collections, overlapping a qualitative spatial dimension that was not present before. Such reuse of knowledge is potentially innovative but poses many questions about the management of the quality of the knowledge and the entity alignments used. The presence, integration, and comparison of different geographical knowledge bases can be beneficial for the maintenance and discovery of entity alignments of good quality.

Another interesting aspect related to the use of co-reference services integrated with an additional knowledge source is the ability to exploit the data semantics in order to change the navigability of the datasets. Such change in the navigability is clear when new arcs are provided within the same data set (e.g. between *dbpedia* resource where they were not linked before) or between resources belonging to different data sets (see Table 1 for a complete account of the data sets connected).

7. ACKNOWLEDGEMENTS

This work was supported by the EnAKTing project funded by the Engineering and Physical Sciences Research Council under contract EP/G008493/1.

8. REFERENCES

[1] S. Auer, S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, and Z. Ives. Dbpedia: A nucleus for a web of open data. *in* 6^{th} *International Semantic Web Conference, Busan, Korea*, pages 11–15, 2007.

[2] C. Becker and C. Bizer. DBpedia mobile: A location-enabled linked data browser. In 1^{st} *Workshop about Linked Data on the Web (LDOW2008)*, April 2008.

[3] T. Berners-Lee. Design issues: Linked data. `http://www.w3.org/DesignIssues/LinkedData.html`, 2006.

[4] L. Carr, W. Hall, S. Bechhofer, and C. Goble. Conceptual linking: ontology-based open hypermedia. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 334–342, New York, NY, USA, 2001. ACM.

[5] B. L. Clarke. A calculus of individuals based on "connection". *Notre Dame J. Formal Logic*, 22(3):204–218, 1981.

[6] A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts. Qualitative spatial representation and reasoning with the region connection calculus. *Geoinformatica*, 1(3):275–316, 1997.

[7] M. J. Egenhofer. A formal definition of binary topological relationships. In 3^{rd} *International Conference, on Foundations of Data Organization and Algorithms (FODO)*, pages 457–472, New York, NY, USA, 1989. Springer-Verlag New York, Inc.

[8] C. Freksa. Temporal reasoning based on semi-intervals. *Artif. Intell.*, 54(1-2):199–227, 1992.

[9] H. Glaser, A. Jaffri, and I. Millard. Managing co-reference on the semantic web. In *WWW2009 Workshop: Linked Data on the Web (LDOW2009)*, April 2009.

[10] J. Goodwin, C. Dolbear, and G. Hart. Geographical linked data: The administrative geography of great britain on the semantic web. *Transaction in GIS*, 12(1):19–30, February 2009.

[11] Grzegorczyk. Undecidability of some topological theories. *Fundamenta Mathematicae*, 38:137–152, 1951.

[12] D. B. Lenat. Cyc: a large-scale investment in knowledge infrastructure. *Commun. ACM*, 38(11):33–38, 1995.

[13] J. Nanard and M. Nanard. Using structured types to incorporate knowledge in hypertext. In *HYPERTEXT '91: Proceedings of the third annual ACM conference on Hypertext*, pages 329–343, New York, NY, USA, 1991. ACM.

[14] A. A. Randell, Z. Cui, and A. G. Cohn. A spatial logic based on regions and connection. *In B. Nebel, W. Swartout, and C. Rich, editors, Principles of Knowledge Representation and Reasoning*, 1992.

[15] J. Rohlfs. A theory of interdependent demand for a communications service. *The Bell Journal of Economics and Management Science*, 5(1):16–37, 1974.

[16] J. Sowa and A. Borgida. *Principles of semantic networks : explorations in the representation of knowledge*. Morgan Kaufmann, 1991.

[17] A. N. Whitehead. *Process and Reality.* The MacMillan Company, New York, NY, USA, 1929.