

IDSWrapper: a Linked Data interface to the Institute for Development Studies' data

Christophe Guéret^a, Victor de Boer^a, Duncan Edwards^b, Timothy G. Davies^c

^a *The Network Institute, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands*

E-mail: {c.d.m.gueret, v.de.boer}@vu.nl

^b *Institute of Development Studies, Brighton, UK*

^c *Web Science Doctoral Training Centre, University of Southampton, UK*

Abstract. This short paper provides a description of the IDS Wrapper used to expose the data from the Institute for Development Studies' Knowledge Services as Linked Open Data. The IDS Wrapper provides Linked Data access to 35,000 research documents on development research as well as its metadata. The IDS Wrapper links this metadata to a number of external sources: DBpedia, GeoNames, Lexvo and the IATI Linked Data set. We expect that the IDS data will play a central role in the larger web of Linked Data for global development.

Keywords: international development, global change, knowledge sharing, data wrapper

1. Introduction

It has been recognized (*c.f.* [2]) that development knowledge is a global public good that belongs to everyone, and from which everyone should benefit. Knowledge sharing is therefore an important issue in the field of international development.

The Institute of Development Studies (IDS) is a leading global charity for international development research, teaching and communications. IDS was founded in 1966 and enjoys an international reputation based on the quality of its work and its commitment to applying academic skills to real world challenges. Its purpose is to understand and explain the world, and to try to change it - to influence as well as to inform.

IDS is a pioneer in development communications hosting a range of innovative and highly regarded knowledge services - including Eldis, BRIDGE, and the British Library for Development Studies¹. These services seek to mobilise knowledge to support more

informed decision-making by those in a position to influence change. This is based on the belief that decision-making is strengthened when it is underpinned by timely and relevant information that reflects a diversity of viewpoints. The Eldis and Bridge services put an emphasis on sourcing research from "Southern" organisations to attempt to balance the domination of research originating from "Northern" organisations to provide this diversity of perspectives on development issues.

This short paper provides a description of the IDS Wrapper used to expose the data from the IDS Knowledge Services as Linked Open Data [1]. We first introduce the data provided by the Knowledge Services Open API and then describe how it is exposed as Linked Data. Then, we locate this data set within the bigger picture of Linked Data for development.

2. Data from the Knowledge Services Open API

The Knowledge Services Open API provides easy programmatic access to over 35,000 of thematically organised research documents that are freely available online, and over 8,400 metadata records about organ-

¹The portal for the IDS knowledge services can be found at <http://www.ids.ac.uk/go/ids-knowledge-services>

isations working in development (*c.f.* Table 1 for an overview of the data served). This data is published under the Creative Commons BY Attribution licence and is therefore free to distribute, remix, and utilise to build applications and services.

The data is accessed through an API that offers a variety of search and identifier based queries. One can look for an entity in particular knowing its unique identifier or perform a topical search around one or more keywords. The Knowledge Services API is available at <http://api.ids.ac.uk/> and is free to use by anyone having registered on the portal.

To expose this API data as Linked Data, we decided to make use of the identifier based services. This allows for a straightforward mapping between an identifier based URI scheme and the relevant API call (see Section 3.1 for more information on the URI scheme we adopted). These API calls are issued by a wrapper which is a service developed separately from the Knowledge Services API and the associated browsing interface. This separation is a design choice allowing for more flexibility in the development and deployment of the Linked Data service.

In the following, we describe the technical details about this wrapper and the content of the exposed Linked Open Data.

3. IDS Wrapper

To expose the IDS data as Linked Open Data, we created a wrapper based on Java restlet technology². The data available from the API is divided into different type of entities grouped in collections. The wrapper defines URIs that directly map a resource to its collection, its type (see Table 1) and its identifier. The wrapper exposes these different entities on demand. Upon a HTTP call to a IDS URI, the wrapper fetches information from multiple sources:

1. First, the wrapper calls the IDS REST API to fetch all known information about the resource (document, organisation, theme, etc.). The JSON key-value pairs are translated into RDF predicates and objects that combined with the resource URI form the triples. The mapping between JSON keys to predicates is based on a hand-written RDF schema embedded with the wrapper code. This schema also lists the `rdf:range` of

the predicate, allowing specific values to be converted to typed literals or resources.

2. Secondly, the wrapper enriches the translated data through a number of custom *linkers*. These linkers take a key-value pair and based on those to establish links to internal or external resources. We describe the linkers used in Section 3.3.

The wrapper is currently deployed on Google AppEngine at <http://idswrapper.appspot.com>. Its code is freely available on GitHub at <https://github.com/cgueret/LinkedIDS>.

In the following, we describe the URI scheme used and how it is mapped to the queries sent to the data API. We also explain the vocabulary used and how connections are established within the data set and with external data sets.

3.1. URI scheme

Every entity within the data set has a unique identifier assigned to it. The wrapper uses this identifier in a generic URI scheme based on the name of the collection, the type of the entity and the value for the identifier:

```
ids:<collection>/resource/<type>/<id>
```

In this URI pattern, “ids” is used for the hostname of the server for hosting the wrapper³.

For instance, the following resource points to the description of “Gambia”, which is a country within the collection Eldis:

```
ids:eldis/resource/country/A1078
```

3.2. Vocabulary used

The data exposed by the data API is made of a collection of fields/values. By default, all the fields are turned into URIs which are accessible under the namespace `ids:vocabulary`. We established mappings for some of them to common vocabularies, for example to Dublin Core for document metadata, FOAF for personal data and SKOS for thesaurus descriptions. We decided to keep the denomination from the data API to allow for a smoother transition for users currently using the API and willing to use the RDF data. These users still see the terms they are used to.

²<http://www.restlet.org/>

³As of May 2012, <http://idswrapper.appspot.com/>

Type	Number	Description
Documents	+35,000	Describe research publications, including title, description, publisher information, thematic, region, and country categorisation, and a Url to the full text of the document
Organisations	+8,400	Describe development organisations including, name, Url, description, thematic, regional, and country categorisation - types of organisations include publishers, funding bodies, NGOs, INGOS, UN agencies
Themes	+1,050	Thematic categories - hierarchical category
Countries	244	Details of countries, including Name, ISO codes - used to describe research country of focus, country of publication, organisation location
Regions	10	Details of Regions - used to describe research region of focus

Table 1

List of entities documents with the IDS data base

However, some fields are not used and are directly replaced by terms from popular vocabularies. This is the case for the term `ids:vocabulary/object_type` which we replace by `rdf:type`. Not doing so would have had a significant negative impact on the usability of the data set, as everyone expects resources to have an `rdf:type`.

3.3. Links to external resources

When building the RDF description of an entity, the wrapper calls a set of linkage services to find external and internal entities to re-use.

3.3.1. Lexvo

The search API provided by Lexvo is used to get a pointer to a page. This page is de-referenced as RDF and the name of the entity to link to is search for within the description. In this way, a document description is enriched with Lexvo language information. Example in Figure 1

<code>rdf:type</code>	<code>ids:Document</code>
<code>ids:country_focus_objects</code>	A1225 Country Uganda UG
<code>dcterms:language</code>	http://lexvo.org/id/iso639-3/eng
<code>ids:title_sort</code>	Assessing Climate Change Vulnerability use of CARE's Climate Change Vulnerability (CVCA) Methodology within the Global V

Fig. 1. Example of link with Lexvo, the literal “English” has been replaced by the corresponding resource

3.3.2. DBpedia

This linker makes use of the public SPARQL end point provided by DBpedia. The full name of the entity is sought and its type, manually mapped to DBpedia equivalent, is indicated to better filter the results. An example of link found for the theme “Food Security” is shown in Figure 2

<code>ids:metadata_url</code>	http://api.ids.ac.uk/openapi/eldis/get/theme/
<code>ids:object_id</code>	C444
<code>ids:children_url</code>	http://api.ids.ac.uk/openapi/eldis/get/children/
<code>owl:sameAs</code>	http://dbpedia.org/resource/Food_security

Fig. 2. Example of a link with DBpedia, the theme being described is linked to the equivalent entity in DBpedia

3.3.3. IATI

The International Aid Transparency Initiative (IATI)⁴ aims to make information about aid spending easier to find, use and compare. Its dataset has been published as Linked Open Data on the Kasabi web platform at <http://kasabi.com/dataset/iati>. It is accessible through HTTP access and SPARQL queries. The IATI linker also uses the full name of the theme and uses a SPARQL query to find a matching Sector, the IATI equivalent of a theme. If a matching sector is found, an `owl:sameAs` relation to the sector URI is added to the theme description. These links between the IATI and IDS datasets can be exploited to match aid programs to research studies. Figure 3 shows an example of such link.

<code>ids:name</code>	higher education
<code>ids:metadata_url</code>	http://api.ids.ac.uk/openapi/eldis/get/themes/C282/full/higher-education/
<code>ids:object_id</code>	C282
<code>owl:sameAs</code>	http://data.kasabi.com/dataset/iati/codelist/IATI/Sector/114/11420

Fig. 3. Example of a link with IATI. The IDS theme C282 (“Higher education”) being described is linked to the equivalent sector in IATI

3.3.4. Geonames

The search API provided by Geonames is used to fetch the identifier of the resource to link to. For this query, the two letters code of the country found in the

⁴<http://www.aidtransparency.net/>

IDS data is used together with the full name. The result is turned into the matching RDF resource on Geonames by using applying their URI scheme <http://sws.geonames.org/<id>>. Figure 4 shows an example of such link.

ids:iso_three_letter_code	GMB
ids:metadata_url	http://api.ids.ac.uk/openapi/eldis/get/cou
ids:object_id	A1078
owl:sameAs	http://sws.geonames.org/2413451

Fig. 4. Example of link with Geonames, the country being described is linked to the equivalent entity in Geonames

3.3.5. Internal links

The data API provides relations between the documents/organisations and the themes/countries/regions they relate to. For instance, a given document will contain material about a particular region. We use the information provided by the API to generate internal links within the RDF data set. An example of such link is given in Figure 5

(CVCA) Methodology within the Global Water Initiative East Africa P
http://idswrapper.appspot.com/eldis/resource/region/C21
2012-03-29 09:07:06

Fig. 5. Example of internal link connecting the document to its region

3.4. Known limitations

There are some limitations directly related to the usage of a wrapper to expose the data.

3.4.1. No search

There is no search functionality available, no SPARQL and no dump of the data set. Someone willing to get the RDF description of something has to know the name of the collection, the type and the identifier of the sought entity. As a temporary fix, users of this RDF data set rely on the original search API provided by IDS to search for the entities. Note that this also restricts us from providing statistics on the internal and external links.

3.4.2. Warmup time

The linkage with external entities is done on demand and cached. Therefore, when a given entity is dereferenced for the first time in the wrapper, some seconds are spent querying the external services for links.

4. Example of usage

We are currently developing a client application that exploits the IDS Linked Data for explorative document browsing. In this application, a user can browse through the IDS documents and is shown the IDS metadata (theme, organisation, etc.). In addition to this, snippets and links to documents as well as images from external sources, that are linked to the document are presented to the user. These resources are for example linked via the theme of the object.

An example of such an enrichment path is a document on “Assessing Climate Change Vulnerability in East Africa” (with the URI [ids:eldis/resource/document/A60737](http://ids.eldis/resource/document/A60737)) which is linked to the term “Climate Change” (available at [ids:eldis/resource/theme/C308](http://ids.eldis/resource/theme/C308)). The DBpedia linker establishes a relation to [dbpedia:Climate_change](http://dbpedia.org/ontology/Climate_change), which in turn links to a number of people involved in this field (as well as to pictures of these people), publications written about the subject as well as other resources. The application will display a selection of these resources related by the DBpedia or IATI linker. At the same time, the information linked via the GeoNames linker will be used to -for example- plot IDS documents or organisations on a map.

5. The bigger picture

Mobilising knowledge is vital for effective international development. The IDS Eldis database is just one of a number of well curated publication repositories focussed on development, and increasingly the owners of these repositories are looking to expose their catalogues as linked data. For example, the Research for Development (R4D) catalogue of research funded by the UK Department for International Development (DFID) (<http://www.dfid.co.uk/r4d/>) has added an RDF export profile to its Open Archive Initiative (OAI) interface, and is making a regular RDF dump available at <http://data.kasabi.com/dataset/r4d-aid-data> using a model based on the Food and Agriculture Organisation’s (FAO) LOD-BD guidance (<http://aims.fao.org/lode/bd>). Even though most publication repositories use different taxonomies in their metadata, the common use of links to DBpedia when publishing linked data provides increased opportunities to provide search and retrieval across datasets, or to enhance searches on one dataset with information from another. For de-

velopment information specifically relating to agriculture, the extensive FAO AGROVOC multilingual thesaurus, available as linked data (and already mapped to a range of other sources, including DBpedia) provides additional connections that can be exploited by applications to provide more advanced search and retrieval of relevant information.

Making knowledge accessible to development practitioners is not just about providing better search: it can also involve proactively integrating knowledge resources into practice environments. IATI has developed an XML standard for the publication of information on development activities and aid flows, and a linked data vocabulary, drawing on DC, SKOS and VCARD is currently in development, with proposed linkages to GeoNames, and the development of a project to map a number of key development project taxonomies together. A pilot project by Research for Development has linked R4D publication records to the aid project that funded them, exposing this in a web widget at <http://r4d.herokuapp.com>. Extending this concept, so that all research relevant to a particular project (particularly early stage projects - as IATI seeks publication of planned as well as in progress project information) based on geographical and thematic linkages, as well as funding, could be displayed has the potential to increase the use of past evidence in project planning and evaluation.

6. Future Work and Conclusions

In this paper we provided a description of the IDS Wrapper used to expose the data from the Institute for Development Studies Knowledge Services API. This service gives access to Linked Data descriptions of 35,000 research documents on development research. The IDS Wrapper currently links these RDF descriptions to a number of external sources: DBpedia, GeoNames, Lexvo and the IATI Linked Data set.

We expect that the IDS data will play a central role in the larger web of Linked Data for global development. Future work include the creation of links to more external sources, in particular other development related data sets. We are also busy with a number of client applications that are either planned or being actively developed at the moment.

References

- [1] Tim Berners-Lee. Linked Data, 2006. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [2] Worldbank. Sharing knowledge to achieve development goals. Précis report 234, Worldbank, 2003. <http://bit.ly/JrUos1>.