

IATI Linked Data: Discussion Paper

Discussion paper prepared by Tim Davies (tim@practicalparticipation.co.uk) for the IATI TAG meeting, May 2012.

Overview:

This paper sets out key considerations for making IATI data available as Linked Data. It proposes maintaining XML as the primary publishing format for IATI data, whilst agreeing an RDF Vocabulary for representing IATI as Linked Data.

It identifies a proposed decimal release modification to the standard which will support this, and highlights possible technical infrastructures which would help realise the maximum value from publishing IATI as Linked Data.

A. Why Linked Data?

Linked Data is an approach to publishing data on the web described by Tim Berners-Lee in the Linked Data Design Issues note: <http://www.w3.org/DesignIssues/LinkedData.html>

Making IATI data available as Linked Data supports activity identification, annotation of activities and integration of heterogenous data sources.

Identifying & annotating activities

Linked Data makes use of URIs as unique identifiers for entities. Establishing URIs for IATI activities would allow third-party data sources to refer to, and express additional information about, activities, and would support automatic discovery of annotations about an activity.

Data integration

Much of the value of IATI data on aid activities comes when the data is integrated with other open data sources. For example, when aid activity information can be combined with budget or geographical data, or when information within aid activities on organisations receiving funds can be combined with information on those organisations. The RDF data model used by Linked Data supports ad-hoc integration of heterogenous data sources.

• Case study: R4D and IATI



- The Research for Development portal records details of research papers funded by the UK Department for International Development (DFID). Many of these papers are funded by particular aid activities, details of which are available in the IATI data published by DFID.

- To explore connections between published research and IATI activity information, R4D converted details of their publications into RDF Linked Data, and also converted IATI files published by DFID into Linked Data. The resulting Linked

Data files were placed into a 'Triple Store' (Linked Data database) where they can be queried. By pointing from publications to their related projects, it becomes possible to identify research related to DFID projects. A widget has been created at <http://r4d.herokuapp.com> which draws upon this linked data to display related research for IATI activities. Another publisher of research who adopted the same RDF model for their data, and who articulated links to another IATI activity, could load this into the same data store, instantly making their publications available linked to IATI activities.

A. IATI as Linked Data

This paper proposes six steps towards providing IATI as Linked Data. These are proposed

as broadly cumulative, such that it is unlikely all six steps will be taken over the short term.

- B.1. Agreeing a vocabulary to convert IATI XML to Linked Data
- B.2. Agreeing a URI pattern for identifying IATI Activities and other key entities
- B.3. Updating the standard to allow publishers to articulate alternative URIs for their activities
- B.4. Publishing IATI code lists as RDF
- B.5. Providing a service to support dereferencing of URIs
- B.6. Providing a service to give access to published IATI data as RDF

B.1. Agreeing a vocabulary

A prototype conversion of IATI XML to Linked Data has been developed based upon a simple mapping of IATI XML entity names to RDF Properties. This is available as an XSLT at <https://github.com/aidinfo/IATI-XSLT/tree/master/templates/rdf>

However, RDF Vocabularies are more valuable when they re-use components from other vocabularies and provide additional structured information.

UNIT-4 consulting have proposed an RDF Vocabulary which will be available to review on the IATI Knowledge Hub. This vocabulary re-uses terms from:

- Dublin Core
- VCard
- SKOS

TAG members with an interest in Linked Data are encouraged to review and comment upon this draft vocabulary. A Linked Data Working Group may be created to drive this process.

The vocabulary should be presented for agreement to TAG during one of the upcoming standard update cycles, and once agreed should be posted as a supplementary standard on IATIStandard.org.

When the IATI Standard is updated then the Vocabulary should also be updated.

B.2. Agreeing a URI pattern

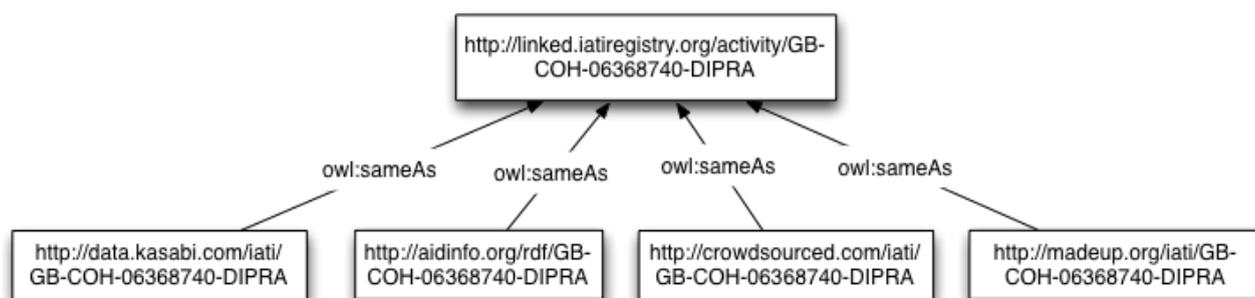
Just as IATI Identifiers should uniquely identify an activity, so URIs that can uniquely identify an activity are required. Anyone could host linked data about a given activity, and so in practice activity information might be available at multiple URIs. The owl:sameAs relationship can be used to indicate two different URIs are in fact talking about the same activity, and to support the integration of data from multiple parties about a single activity.

SameAs identifiers can be articulated horizontally between any number of URIs. However, this requires each publisher to be aware of other possible representations of an activity, limiting the possibility of automatically discovering data. An application which had only been able to look up the data at the left and right of the diagram below would not know they are about the same activity.



By contrast, if we establish a convention for IATI Activity URIs, then even if those URIs

cannot be dereferenced, then all files using them can know they are talking about the same activity, as in the model below.



Therefore, this proposal suggests the use of <http://linked.iatiregistry.org/> as the URI base for either primary URI, or sameAs URIs in RDF representations of IATI data. All files should at least use [http://linked.iatiregistry.org/activity/\[iati-identifier\]](http://linked.iatiregistry.org/activity/[iati-identifier]) but may also choose to use [http://linked.iatiregistry.org/organisation/\[organisation-id\]](http://linked.iatiregistry.org/organisation/[organisation-id]) to identify organisations.

B.5 and B.6 sets out how these URIs could be made dereferencable. However, even if no server is running at <http://linked.iatiregistry.org> agreeing on these URI patterns supports third-parties to discover and integrate data about the same activities.

B.3. Standard modifications

IATI publishers will not be obliged to publish their own Linked Data. However, if a publisher choose to publish linked data about their IATI activities then allowing them to declare where this data is published would support discovery of it, and any additional information they may choose to publish as Linked Data alongside it.

For this reason, it would be beneficial to introduce two new optional properties into the IATI standard.

- **`iati-activities/@linked-data-uri-base`** - a URI path upon which an activity identifier can be appended to get a dereferenceable URI for any activity contained within a file.
- **`iati-activity/@linked-data-uri`** - a URI for a given activity (over-rides `iati-activities/@linked-data-uri-base` if set)

Where a publisher declares using one of these properties that authoritative linked data is accessible for an activity then consuming applications that are generating Linked Data from an IATI XML file should assert an `owl:sameAs` relationship to the relevant URI.

They should still use, or `owl:sameAs`, the URI specified in B.2. in addition.

B.4. Publishing IATI code lists as RDF IATI code lists should be made available as RDF at [http://linked.iatistandard.org/codes/\[CODELISTNAME\]/\[CODE\]](http://linked.iatistandard.org/codes/[CODELISTNAME]/[CODE]) using SKOS so that these URIs can be used in creating IATI Linked Data.

B.5. Providing a dereferencing service

A server could be maintained at <http://linked.iatistandard.org> to provide a minimum set of services to support discovery of IATI linked data. This service could:

- Resolve requests to <http://linked.iatistandard.org/codes/> and provide the relevant code lists or codes as SKOS data (either by directly serving this data, or redirecting to the

appropriate location).

- Maintain a cached list of known sameAs relationships for IATI Activities (and possibly, organisation IDs), so that requests to [http://linked.iatistandard.org/activity/\[iati-identifier\]](http://linked.iatistandard.org/activity/[iati-identifier]) are resolved with RDF data that asserts relevant sameAs relationships. These sameAs relationships may be discovered (a) by reading IATI XML files to look for `iati-activities/@linked-data-uri-base` and `iati-activity/@linked-data-uri` attributes; (b) accepting third party data dumps with known relationships; (c) running regular searches for relationships.

This would allow <http://linked.iatistandard.org> to act as a hub connecting up linked data about IATI from across the web, whilst minimising the server load or requirements on this server.

This is likely to require the development of some custom code, and occasional maintenance to accept third-party data dumps. It may be possible to draw upon the <http://sameas.org/> service or code in order to operate this service.

B.6. Providing an IATI data store

The greatest value from IATI Linked Data will be realised when it is possible to query across the sum total of facts expressed about IATI Activities from heterogeneous services across the web.

In order to support such queries it would be advantageous to establish a single triple store where IATI data can be stored as linked data. This would consist of:

- A regularly converted copy of the IATI XML data as Linked Data
- Aggregated data from third parties

Each different source of data could be stored in it's own named graph, to allow querying across sub-sets, or the total collection, of available data.

To provide an IATI triple store would require:

- Provision of a suitably powered server infrastructure
- Creation of tools to synchronise data from the IATI registry into this triple store
- Tools to crawl and aggregate third party data
- A query interface onto the data, which supports querying by named graph
- Development of an accessible front-end to the query interface to support a diverse range of users

I. Related projects

There are a range of projects which development of IATI Linked Data should reference:

- The Humanitarian Exchange Language (HXL) project are exploring a linked data approach to store data on humanitarian resources;
- LinkedGov are providing a platform to generate linked data from UK Government Data;
- Data.gov.uk and Data.gov both have programmes involving the creation of linked data;
- Reegle and other environment actors have been creating extensive environmental data linked datasets.