

Organisational Identifiers: Proposed principles, documentation and modifications

Prepared by Tim Davies (tim@practicalparticipation.co.uk) for the IATI TAG meeting in May 2012.

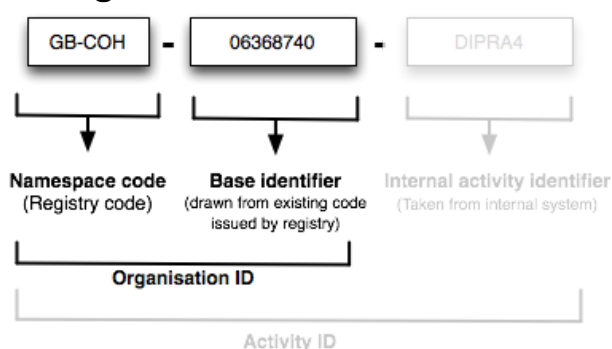
Summary:

This paper sets out principles for selection of organisation identifiers. It is primarily based on describing and formalising current practice. However, it also sets out a number of emerging issues which should be addressed in either decimal, or version number upgrades of the standard.

TAG opinions are sought on:

- The proposed organisational identifiers guidance (A)
- The proposed changes (B)
- The additional issues (C)

A. Organisational Identifiers Guidance



Organisational identifiers have two uses in IATI.

- They are used to identify organisations reporting activities, receiving funds, providing funds, or otherwise involved in activities
- They form the basis of Activity Identifiers, who are composed of the reporting organisation Organisation ID, and an Activity ID.

The IATI Organisational Identifier scheme is based on re-using existing known identifiers for an organisation. Identifier strings consist of two components:

- 1) **The namespace code** (a code that identifies a given organisational identifier registry)
- 2) **The base Identifier** - re-use of an **existing identifier** from that namespace/registry

(1) and (2) are separated by a dash '-'.

The namespace code (1) may contain a dash '-'.

Worked example

GB-COH is the namespace code given by IATI Support to the UK Companies Registrar which issues numerical company identifiers to legally constituted limited companies in the United Kingdom of Great Britain and Northern Ireland. 06368740 is the company number (base identifier) issued to Development Initiatives Poverty Research Limited. Combining these we get an IATI Organisational Identifier of: GB-COH-06368740

Sub-units

Where a registration scheme includes or allows separate identifiers for sub-units (e.g. is able to refer to departments within an organisation), then these may be included following a second '-' delimiter. (e.g. GB-COH-06368740-AidInfo)

Replacement considerations: Where an existing base identifier contains '-' (dash) then this should be replaced with '_' (underscore).

A.1. Guidance on selecting the namespace or registration agency to use:

An organisational identifier should:

1. **Use an existing registration number/identifier** that has been given to an organisation by an official registry of organisations (for example, a national company registrar, or drawn from a government register of charities). Registries that provide machine readable access to the data about organisations are generally preferred over those that do not (subject to (3) below).
2. **Use a namespace/registry code from the existing IATI list of namespaces and registration agency codes.** This is currently available at <http://bit.ly/iati-org-reg>. If no existing code is available then you should contact [IATI Support](#) with details of the registration agency you wish to re-use a code from in order to have a code set up. This list is currently maintained by the IATI secretariat over the short-term, with the hope that the wider open data community will adopt and develop a shared list of registration agency namespace codes in future.
3. **Where an organisation has multiple registration numbers, prefer those identifiers which uniquely pick out the organisation itself, rather than properties of the organisation.** For example, a VAT/Tax registration number describes that an entity is registered for tax, but may not map directly onto a single legal entity - as - for example - groups of companies may share a VAT registration number. For this reason, an identifier such as GB-COH-123456 would be preferred to GB-VAT-987765. If you are unsure of which registration agency to prefer, please consult IATI Support. If no authoritative registration agency information is currently collected (e.g. internal identifiers and supplier numbers are recorded, but no external identifiers) IATI publishers are encouraged to identify opportunities to collect an authoritative public identifier.
4. **Where an organisation is not identified in an official national registry, but it is identified in one of the recognised IATI Bi-lateral, Multi-lateral or INGO code lists, these codes should be used.** Multi-lateral and INGO numerical codes from these lists should be prefixed with DAC- The bi-lateral list (e.g. GB-1) will be deprecated in a future update of the standard and replaced with an official DAC provided code-list. In the meantime, non-prefixed bi-lateral code list values from IATIStandard.org can be used.
5. **Where none of the above provide an appropriate identifier, third-parties may provide appropriate lists of organisations.** Consult with IATI Support to identify appropriate identifier lists.
6. **Where no other identifiers are held or available, agree a registry code under ALT- with IATI Support for your organisation and use IDs within this namespace.** For example, if Development Initiatives were to only have internal supplier database identifiers for organisations, it could use codes of the form MISC-DEVINIT-123456. It would register MISC-DEVINIT with IATI Support, and ideally would provide some sort of reconciliation service to return the names and any additional information held that relates to each organisation ID.
7. **Where no identifier is held, make use of the @type attribute** to indicate the type of organisation (using the Organisation Type code list) and do not provide a @ref value.

Replacement patterns

Where an existing base identifier contains '-' (dash) then this should be replaced with '_' (underscore).

Once determined, the organisational identity can be used in the @ref attribute of reporting-org and participating-org, and should be used in @ref attribute of transaction/provider-org and transaction/recipient-org.

The string value name within any organisational identifier XML tag is provided for convenience of consuming applications only, and should not be treated as definitive.

A.2. Interpreting identifiers

The above principles are designed to increase the chance that different IATI publishers will use the same identifier for the same organisations, whilst avoiding the creation of a central registry of organisations. However, it is possible that the same organisation might be identified by multiple organisational IDs across the whole corpus of IATI data, and so any attempt to find all the activities involving a particular organisation will need to perform additional data processing and research.

To split the namespace, and base identifier, users can draw on the list of Registration Codes, and split out anything following one of the recognised prefixes, such that a parsing process can check if the string begins with a known **namespace code**, and if so, can strip of this portion, and split the resulting string on any dashes to get the **original base identifier**.

Using the **namespace code** it may be possible to locate services which can be used to look up further information on the **base identifier**.

A.2.1. Identifier reconciliation

The market or community may provide tools which support the reconciliation of organisation identifiers into further details about that organisation. This might require consuming applications to select the appropriate reconciliation tool depending on the namespace of an identifier, or might provide a service to reconcile any known identifier. No standard for the exchange of organisation information is yet established, but the IATI TAG continue to monitor the development of wider community standards on this matter, and if a widely accepted standard emerges, will recommend the use of this.

Example 1: Open Corporates

OpenCorporates.com is building an open dataset drawn from Company Registries across the world. You can check which registries is currently includes data for at <http://www.opencorporates.com>. A consuming application could make use of the OpenCorporates.com URL structure for company information to replace GB-COH- in any identifier with '<http://opencorporates.com/companies/gb/>' and append .xml or .json to the resulting string to get a URL which, when accessed, will return data on the organisation in question. For example, the identifier GB-COH-06368740 could be programmatically mapped to the URL <http://opencorporates.com/companies/gb/06368740.json> which provides a machine-readable name for the organisation in question.

Example 2: Open Circe

OpenCirce is a prototype project to resolve organisational identifiers, including IATI Organisational Identifiers. It is aware of a number of registration agencies and third-party data sources, and given an identifier appended to the URL <http://opencirce.org/org/code/> will attempt to resolve information about it. For example, the URL <http://opencirce.org/org/code/GB-COH-06368740> provides the organisation name and

onward links to further sources of information about a given company. Append .xml onto any OpenCirce URL to get structured data.

OpenCirce also supports codes from the DAC namespace, and includes crowd-sourced information such as organisation websites. For example, <http://opencirce.org/org/code/DAC-41119> provides details on UNFPA.

Details of other known services for resolving Organisational Identifiers and source code for carrying out this task can be shared on the Knowledge Base.

A.2.2. Asserting equivalence of identifiers

At present the IATI standard only allows one identifier (@ref) to be specified for an organisation. If you know multiple identifiers for an organisation then you may wish to independently host a supplementary file indicating identifiers that describe the same organisation. As this situation has not yet arisen, please liaise with IATI Support when planning to publish such a file so that appropriate standards for this can be identified.

A. Changes required to implement this as an agreed standard

The IATI Organisation Standard has not been documented fully up until this point. This paper presents the proposed documentation under A.

B.1. A new replacement rule to replace '-' (dash) within base identifiers with '_' (underscore) should be introduced

- *Change B.1 may require updating some existing identifiers*

B.2. The @type attribute needs to be available for all XML elements where an organisational @ref is allowed. This includes transaction/transaction-provider and transaction/transaction-receiver

- *Change B.2 may require some publishers to update their approach to outputting data, but has minimal impact.*

B.3. The list of recognised namespace and registration agency codes should be updated to reflect all namespaces currently in use

- *Change B.3 is administrative only and has minimal impact on existing publishers.*

B.4. The IATI Multi-lateral and INGO code-lists for organisations provided by IATI should be updated to include a DAC- prefix, so that, for example, 41110 becomes DAC-41110.

B.4. has implications for many of the organisations and activity identifiers currently used, and would mean many iati-identifiers change. Changing in the next three months before a large number of tools are relying upon stable iati-identifier values would be valuable (it may also be appropriate to consider Additional Issue 1 at the same time).

Applications that rely on the iati-identifier in order to check for updates to any given activity consider how this change impacts on their import/synchronization process.

C. Additional issues to address:

The following issues are presented for comment from the TAG. The TAG Secretariat will need to decide whether and when these are put forward for consideration as later decimal or version upgrades within the IATI Standard.

C.1. Replacement of current IATI Bi-lateral list

The current IATI Bi-lateral organisation code-list is based upon DAC codes, but with ISO codes replacing internal DAC numerical codes for countries. To align the IATI Standard better with the DAC CRS, this list should revert to the DAC numerical and non-ISO code format, and should be moved to the DAC namespace code.

For example, the identifier for DFID, 'GB-1', would become 'DAC-12-1'.

See http://iatistandard.org/codelists/organisation_identifier_bilateral for reference.

C.2. Locating a global government identification list

The DAC Bi-lateral list only contains some country government departments. We need a shared list for identifying government departments and units across the world, as these are rarely registered in other registration systems.

Wider consultation to locate appropriate lists is required.

C.3. iati-identifier replacement pattern

Activity identifiers (in iati-identifier) might include '-' (dash) within the identifier. The standard should be updated to suggest this is replaced with '_' to

C.3. Alternative organisation identifiers

Where a publisher is aware of more than one identifier for a given organisation it would be useful for them to provide this information to assist mapping between different organisational identifier.

There are a number of possible approaches to address this:

1. Adding an *@alternative-ref* attribute to any element which can take an organisational identifier *@ref* attribute. This has the advantage of creating the minimum impact on the standard, but has the limitation of only allowing one alternative identifier to be specified.
2. Adding an *@alternative-refs* attribute to any element which can take an organisation identifier *@ref* attribute, and allowing multiple comma separate organisation identifiers to be provided. This has the advantage of creating a minimum impact on the standard, but breaks with XML principles and adds an overhead for parsing tools - preventing simple access to organisational identifiers using XPATH standards.
3. Replacing any element which can contain an organisation identifier *@ref* with a complex element which can contain multiple 'code' and 'name' elements. Name elements would have language tags, and code elements would have a *@status* attribute indicating whether they are the 'preferred' ID, a 'secondary' ID, or a 'loose_match' ID. For example, `<reporting-org ref="GB-COH-06368740" type="22">Development Initiatives Poverty Research</reporting-org>` would become `<reporting-org type="22"> <code @status="preferred" @ref="GB-COH-06368740"/> <code @status="alternative" @ref=""/> <name xml:lang="en">Development Initiatives Poverty Research</name> </reporting-org>` This has the advantage of being the most expressive option, and supporting simple queries to bring together all the activities relating to some organisation whether or not a code has been given as the primary code. However, it would require a major update to the standard.
4. Creating some mechanism by which publishers can provide a standardised mapping between organisational identifiers either within their IATI files, or in supplementary files. This has the advantage of having no impact on the standard, but requires re-users to carry out considerable processing in order to map together identifiers.

This issue requires detailed discussion and a proposal to be put forward.

C.4. Standards for organisational information interchange

The TAG should continue to monitor wider standardisation efforts for the exchange of information about organisations, and should participate in discussions where appropriate to ensure the needs of the IATI community are met.

