



IATI INTERNATIONAL
AID
TRANSPARENCY
INITIATIVE

IATI

DRAFT Technical proposals for how IATI will be implemented

An overview of the detailed technical proposals

Produced by IATI Technical Advisory Group

March 2010

Background

This note is an overview of a more detailed technical paper being discussed by the TAG technical subgroup. It covers emerging thinking, areas for discussion and some early recommendations for the technical proposals for implementing IATI, attempting to set out HOW data providers will publish data and in what format, and HOW users of information can find, access and use the data.

Main considerations for IATI technical design

Openly licensed — any third party must be allowed to use any published IATI data under consistent open terms, without requiring explicit permission from the donor who provided the data.

Machine-readable — it must be possible for computer programs to extract useful information from IATI data without manual intervention.

Easily accessible — users must be able to obtain IATI information automatically and anonymously. IATI should use well-known and well-supported open networking formats and protocols.

Decentralised — IATI must be capable of continuing to function without a centralised administrator or computer system

Comparable — while not all donors will supply all data specified by IATI, whenever two donors *do* supply the same data, it must be possible for users to compare those data in a meaningful way.

Flexible — data formats and publishing schedules must allow donors the flexibility to omit information that is not relevant or available at different points in the information's life cycle.

Extensible — it must be possible for donors to supply additional information not covered by IATI.

Vendor- and platform-independent — IATI information publishing must not depend on software from a specific vendor, or on a specific hardware or software computing platform.

Multilingual — for human-readable text, IATI must support all major world languages, and it must support multiple versions of the same text in different languages.

Overview of draft proposals

The initial recommendation is that IATI adopts a decentralised, web-based architecture where:

Donors will publish machine-readable aid- data files in a standard XML format freely available on public websites.

An **IATI website** will create a **registry** of *links* to these aid-data files (*not* store the actual data).

Donors must make **information about their published aid data files** available to the IATI registry.

In addition to the donor ID, **globally unique IATI identifiers** for projects should be established to allow common referencing for co-financed projects and to enable traceability.

There will be a **notification mechanism** to let users know about new or modified aid-data files.

The **IATI registry** could take on additional functions such as authenticating sources of data, providing amalgamated notification feeds, services and tools to help donors create and host aid data files

How data will be published and accessed

How will data users access information?

There are two categories of users of aid information we need to consider: *specialist users with technical expertise* such as application developers, owners of aid databases such as Aid Information Management Systems, statisticians, researchers, and analysts; and *non-technical end users* such as parliamentarians, policy staff, and citizens.

IATI aims to provide non-technical end users with access to better quality data and better quality services, websites and tools to help access it. IATI is designed to do this by serving the specialist users directly with access to the raw data in a technical format (XML) that enables them to become *Infomediaries*, and relatively easily and cheaply repackage the information into new websites, applications and other tools for the non-technical end users.

The infomediaries are a crucial part of the IATI initiative as they will act as the main *retailers* of the data, and serve most users. IATI's role is as the *wholesaler* for aid data, providing and distributing more, better quality data directly to these *retailers*. As a result those retailers will be more numerous and better stocked, with fresher, cheaper products.

These specialist users will be able to access the data both manually by searching directly on the registry website, or automatically by establishing systems to retrieve data directly. While we have striven to keep the architecture and data formats as simple as possible – and it *will* be possible for non-technical end users to find and access the raw data - it is likely to require a certain amount of technical knowledge and effort to use the data effectively. However, it would be relatively simple to create simple translation tools for both specialist and non-specialist users to open this data in desktop office applications such as Excel.

Additionally potential users of data will be able to *subscribe to notifications* when new data are available. Notifications remove the need for manual work, constant visits to websites, or complex work-arounds to determine if or when a donor's aid information has changed and needs to be re-loaded. There are well established standards for this known as "RSS" feeds.

How should data providers publish information?

The emerging thinking is that donors should use existing web and internet infrastructure as their channel to publish aid data in **machine-readable files**, in an agreed XML format. Donors can publish those files on a new or existing website as they would any other resource, such as a graphics file. Each file will have its own unique web address (URL).

They will then inform the registry of new or updated data by providing some basic information - such as the web address (URL) and date- relating to the data files. This can be done through an automated *push* notification procedure, or manually through the IATI registry website.

The registry

As described, we are proposing to create a central registry which will not hold the actual data, but will be a website application that operates as a catalogue that provides links to and information about data that are available elsewhere.

There are already examples of this decentralised approach for publishing and storing data with a central registry being used elsewhere: such as CKAN¹ and data.gov.uk². The main difference and advantage of the IATI registry over these will be that the data available via the IATI registry will be consistent and comparable.

Additional value offered by the registry

The core function of IATI registry is to catalogue links to the data files. However, the web application that hosts the IATI registry could also provide additional value added functions such as:

- the ability to authenticate sources of data
- amalgamated RSS notification feeds from multiple donors
- hosting services and software tools to help donors create and host aid data files and notifications

Global identifiers

We are proposing that each project has its own unique identifier in the form of a web based location address e.g. <http://example.org/acme/activities/12345/>. This will enable projects to be linked to each other to enable traceability or to identify where projects are co-financed.

In addition IATI should establish URL based identifiers for other main elements such as country, sector and organisation. This will lay the foundations for providing information at those locations in the future³. For example, "http://aidtransparency.net/countries/HT/" could contain an XML summary file of *all* aid activities in Haiti

¹ Comprehensive Knowledge Archive network <http://www.ckan.net/>

² Data.gov.uk is the UK Government registry for its public data

³ As well as enabling a future semantic web / linked data approach <http://www.w3.org/DesignIssues/LinkedData.html>

Licensing of data

Why is licensing important?

To enable specialist users to access the data and make it usable for others, simply publishing information on the internet is not enough. Users and developers need to be given a green light to build on, republish and create derivative works from documents and datasets, not just to be able to access them. An explicit legal statement granting users permission to reuse and redistribute digital material is vital in this respect.

What are the main considerations?

- There should be no constraints for using data, including for commercial use
- Do providers want attribution of the source of data when it is used?
- Do we want to require that anyone using the material should license any derivative works under the same (or similar) license (share-alike)?
- License policy must be easily understood by users - one license model for all IATI data would be preferable from a user perspective
- Data and documents both need to be licensed and have different requirements.

Proposed licensing model

We recommend one of three models is adopted⁴:

Licensing option A (attribution): Creative Commons Attribution for content and Open Data Commons Attribution for data.

Pros: Requires that users acknowledge publishers, if this is required. This can be beneficial to the provider of data.

Cons: Requires users to attribute, which is more onerous than if the material were in the public domain. Without share-alike, users can take the data and not share the results.

Licensing option B (attribution + share-alike): Creative Commons Attribution Share-alike for content and Open Data Commons Open Database License (ODbL) for data.

Pros: Requires users to share back the fruits of their labour. In particular this might be an attractive option to compel commercial organisations to contribute back to the community, and is beneficial for the open aid data movement as a whole.

Cons: Requires users to share-alike, which is more onerous than if the material required attribution only or were in the public domain.

Licensing option C (public domain): Open Data Commons Public Domain Dedication and License (PDDL) or Creative Commons (CC0).

Pros: Imposes no restrictions on users, so most attractive option to them

Cons: No requirement to attribute publishers or put works back in public domain

⁴ See <http://www.opendefinition.org/licenses/> for information about the different licenses

Example of IATI registry architecture

