

## Section II: Schedule of Requirements

### IATI Data Store

#### Background:

The International Aid Transparency Initiative provides a standard publishing format for data on development cooperation and humanitarian aid to be shared and compared by all. Using transparent, good quality data helps efforts to achieve sustainable development. IATI sets the standard for the way information is reported so it is easier to plan, track and compare the progress and outcomes of different projects and activities. IATI data is 'open data', so it can be used by anyone for any purpose.

Accessing and using the best available data is crucial for informing effective decisions to promote development and provide essential support in the face of humanitarian crises. Access to IATI data is an important planning aid for developing and donor country governments. It enables academics, politicians, campaigners, journalists, and members of the public to understand how development and humanitarian resources are used. It also gives them the information they need to hold to account those who provide or use aid funding.

This is an open tender for the building of the IATI DataStore - a robust online data service providing timely, standardised access to all data published using the IATI standard. The tender includes the development of the product as well as its hosting for at least the first year of operation.

#### Submission:

Please note that no cost associated with the required services should be disclosed in the technical proposal. The costing should only be disclosed in the financial proposal.

The bidder should provide:

- Evidence of their knowledge and experience of the IATI standard
- Details of the team they have at their disposal to carry out this work
- A detailed technical proposal that takes into account, but is not bound by, the terms provided in this document.
- Proposed timelines for completing both phases of development (as outlined below)
- A financial proposal containing costs for hosting and maintaining the DataStore.

## **Objective:**

The IATI DataStore (DS) is a repository containing all usable data published to both the IATI activity and organisation standards. It provides a robust, timely and comprehensive data service to the following user groups:

- Developers of management information systems and information products
- Researchers and data scientists seeking to analyse the data
- Publishers seeking to confirm that their data has been successfully loaded into the DS

The datastore is NOT an information product of benefit to non-technical users. It allows technical users to access the data they require through selections of filters, serialisations and data formats and delivers data through both API and downloadable file endpoints. Its design should encompass the following user stories.

## **Outputs:**

The DS should be delivered in three phases that run back-to-back.

- Phase One will allow developers to access their selected activities in XML and JSON format via API.
- Phase Two will provide analyst and publisher access to a range of serialisations and reports.

### **Phase One**

- **Extract, Transform and Load**
  - DS should poll the IATI Registry continuously for updated or new files. Metadata should be stored and made available on the results of each poll (the list of files to be fetched) and each attempted fetch.
  - DS should access Registry validation logs for each file. (There is no need for DS to perform validation.) Failed files should be ignored. Activities in files with non-critical validation errors should be processed. Only files are fetched which have a current IATI registry validation log.
  - Data should be transformed to the latest version of the standard (e.g. Version 2 standardisation of codes, narrative elements, location elements, vocabulary attributes, etc.) [In future, after the next major upgrade, there will be a need for data to be accessed in the two latest versions of the standard.]
  - Monetary values should be stored in the original currency and US Dollars (calculated using historical IMF rates for the relevant value dates).
  - A copy of the most recent original xml should be stored for each activity.
  - Processing and error handling during ETL needs to handle known complexities, including the reporting of the same activity twice, temporarily missing activities, etc

- All logic employed in transformation processes must be clearly and transparently documented for both developers and end users.
- Metadata should be maintained that records all validation errors (via Registry) and transformation processes for each activity.
- Data should be stored in a manner that maximises speed and flexibility of query and access.
- Filters
  - Querying should be possible using the following filters:
    - Last updated date and time
    - All elements of the standard
    - Free text search across title, description or entire activity
    - Pre-processed aggregations and calculations fields, including: US Dollar monetary values; CRS Sectors; Years (from dates)
- Developer Output
  - Developer queries should result in the appropriately filtered delivery of entire standardised activities in xml or json format, or original (untransformed) xml. The standardised output should follow the standard IATI XML as close as possible.
- API
  - Any products making use of the DS, including the DS query interface itself, may only communicate with the datastore via the API.
  - The API must be RESTful, must be versioned, must use HTTPS
  - Results should be deliverable in full or paginated, depending on the choice of the DS user.

## **Phase Two**

- CSV/XLSX Serialisations
  - CSV and XLSX outputs should be available in the following serialisations - defined by the unit represented on each row
    - Organisation
    - Activity
    - Transaction
    - Activity Budget
    - Organisation Budget
    - Result
    - Location
    - Document links
  - XLSX output should allow for multiple serialisations to be selected to build a multi-tab workbook.
  - Transaction and Budget serialisations should also allow for data rows exploded by recipient country/region and/or sector where multiple values for country/region and/or sector exist
  - Each serialisation should be built with a default set of columns [tbc] but users should have the option to make their own selection of columns to display.
- Analyst Interface
  - A functional user interface should allow analysts to build complex queries on both activities and organisations that

- Result in csv or xlsx file downloads
  - Utilise the filters outlined above
  - Choose from a number of serialisations outlined above
  - Select the columns to display within each serialisation
- Publisher Interface
  - Provide a report that allows publishers to check the status of their data in both the Registry and Data store

### **Phase Three: Maintenance/Hosting services**

- The bidder must undertake to host the DataStore for a period until 31 December 2019. In addition, the bidder should confirm availability to provide hosting/maintenance services beyond this date for a period of one year, providing the estimated annual cost, as part of your financial proposal.

### **System Requirements**

- Coding Standards
  - To be built using python 3, using lxml for XML processing.
  - Code must be written to a style guide (for example adherence to PEP8/docstring conventions, with linters added to continuous integration checks)
  - Any third-party packages that are used should be well documented and well supported
  - To be built under TDD principles:
    - Robust repeatable feature testing
    - Custom functionality must be unit tested (including examples of where there are edge cases in data, including mocking of HTTP requests)
  - Must be easy to debug
  - A number of functions and processes used by the DS are applicable to other IATI tools. The DS should where possible, reuse existing IATI components.
  - The code of the DS will be published on Github with an GNU 3.0 open source licence.
- Agile Management
  - The proposal should adopt an agile approach to development which allows for reasonable modifications to the original specification based on feedback from the IATI secretariat and user testing
- Documentation
  - Developer Manual
  - Analyst User Manual
  - Data Dictionary

## **Timing:**

The bidder should provide clear timelines for the proposed delivery of all phases of development and maintenance/hosting. While time is of the essence, the evaluation will seek to assess this against the quality of the product being offered, so the shortest timeline is not necessarily the most competitive.

Indicative timing phase:

Phase 1: 2-4 months

Phase 2: 2-4 months

Phase 3: period up to December 2019.

## **Reporting and Testing:**

The bidder should provide the following

- A technical specification (based on these Terms of Reference) – to be agreed with IATI before development work starts. Thereafter this will be a living document that can be subject to change by mutual agreement between IATI and the bidder.
- A project management plan that includes milestones at agreed points in the development which allow IATI to review progress
- A communications channel (eg. Slack) for the development team and IATI to exchange queries and comments throughout the development cycle.