H3ABioNet
Pan African Bioinformatics Network for H3Africa

Domain-Specific Minimum Information Data Reporting Guidelines

Recommendations for Use

Acknowledgments

All reporting guidelines were developed by the H3ABioNet Data Standards Work Package, in collaboration with additional interested parties including, H3Africa project members, SIREN and KidGen Australia.

Members include:

Judit Kumuthini
Melek Chaouch
Michael Thompson
Zahra Mungloo-Dilmohamud
Lyndon Zass
Dassen Sathan
Anisah W. Ghoorah
Faisal M. Fadlelmola
Mamana Mbiyavanga
Shakuntala Baichoo
Katherine Johnston
Sumir Panji
Chris Fields
Nicola Mulder

Table of Contents

Purpose

Minimum reporting requirements and (or) guidelines outline an essential list of primary data and metadata elements (fields) which are recommended to be captured to improve the interoperability of data, as well as the reproducibility and interpretability of experiments within a given research endeavour.

To aid the aforementioned purposes, the H3ABioNet Minimum Data Dictionaries project has aimed to develop domain-specific minimum reporting guidelines, which can be employed to promote comprehensive and harmonized data collection within specific domains e.g. kidney disease research, stroke research etc. These reporting guidelines are provided as data dictionaries, information glossaries which define all data elements in the reporting guideline for easier understanding, and associated XML schemas, which can be employed to streamline data collection/reporting of the elements within a given reporting guideline.

These data reporting guidelines were based on the H3Africa Standard CRF recommendations as produced by the Phenotype Harmonization Working Group and Health Informatics Work Package, which in turn was based on PhenX collection measures (https://original-phenxtoolkit.rti.org/index.php). Additionally, the reporting guidelines and associated data dictionaries were harmonized, consolidated and mapped to pre-existing ontologies, and reviewed by international domain-specific experts.

The purpose of the following report is to guide users on how to use the reporting guidelines and employ the associated data dictionaries and XML schemas in their current and future research endeavours. These guidelines are recommended to be used in conjunction with the Data Management Plans developed by H3ABioNet.

Overview

Domain-specific studies can be poorly articulated and indexed, raising the need for minimum reporting requirements and (or) guidelines to fully understand the context, methods, data and conclusions that pertain to a study and/or experiment. The primary aim of a guideline is to provide broad guidance for a specific study's data collection and management. It is not intended as a script for conducting clinical assessment, or the application of care. Reporting guidelines are an essential part of quality experimental practice and various guidelines targeting specific domains have previously been created e.g. guidelines for study protocols, such as animal intervention studies and randomised controlled trials.

The H3ABioNet Minimum Data Dictionaries project has developed domain-specific minimum reporting guidelines. These data reporting guidelines are subdivided into three sections based on varying data sources:

1. **Participant-Level Information**: Contains phenotype/domain-specific information related to the patient and(or) participant. Elements included under this section include demographics, anthropometrics and disease-specific information.
2. **Study-Level Information**: Contains specific information related to the developed study including study design and aim(s).
3. **Experiment-Level Information**: Contains data relating to the experiment(s) within the specific study, including, experimental protocol, instrumentation employed, quality control procedures, and tertiary data analysis.

The reporting guidelines are provided as data dictionaries which are subdivided into 8 columns, describing each data element in the reporting guideline. These 8 columns are described below:

- **Element**: The first two columns describe the category and name of a given data element within the reporting guideline.
- **Definition**: Defines a given element based on the ontology to which the element is mapped.
- **Data Type**: Describes the particular kind of data element, as defined by the values it can take.
- **Accepted Value**: Describes, where applicable, the values which can be inputted with a specific element, as well as value restrictions.
- **Importance**: Specifies an element as either essential (E) or optional (O). Essential elements are crucial to the interpretability and reusability of experiments and data within the domain. Optional elements aid the aforementioned factors, but are not required for these purposes.
- **Ontology ID**: Specifies the ontology ID to which an element is mapped to. An ontology encompasses a representation, formal naming, and definition of the categories, properties, and relations between data elements (entities) that substantiate one, or many domains.
- **Concordant Ontologies**: Describes other ontologies which include similar data elements.
- **Concordant Standards**: Describes other reporting and(or) collection standards which include the data elements.

The reporting guidelines can be easily and freely employed in order to guide study planning, as well as data collection and management strategies. It can also be used to guide users as to which ontologies and reporting guidelines to consult, and are easily adjustable and customizable for researchers to employ with ease.

XML Schemas

Overview

XML schemas have been developed as part of the reporting guidelines to ensure the preservation of data consistency and integrity, in structure and correctness, accounting for varying data collectors, generators, managers, sources and transmissions into software systems. The purpose of these schemas is to present a vendor-neutral and interoperable data interchange format for reporting clinical research data. XML stands for eXtensible *Markup Language*, a language for describing data. It can be used to describe any data by using markup *tags* (or *elements*), snippets of information that you add to your data to provide qualification and description. The developed XMLs conform to a schema which defines validation rules as well as a set of allowed tags, ensuring mechanical correctness and conformity to set standards.

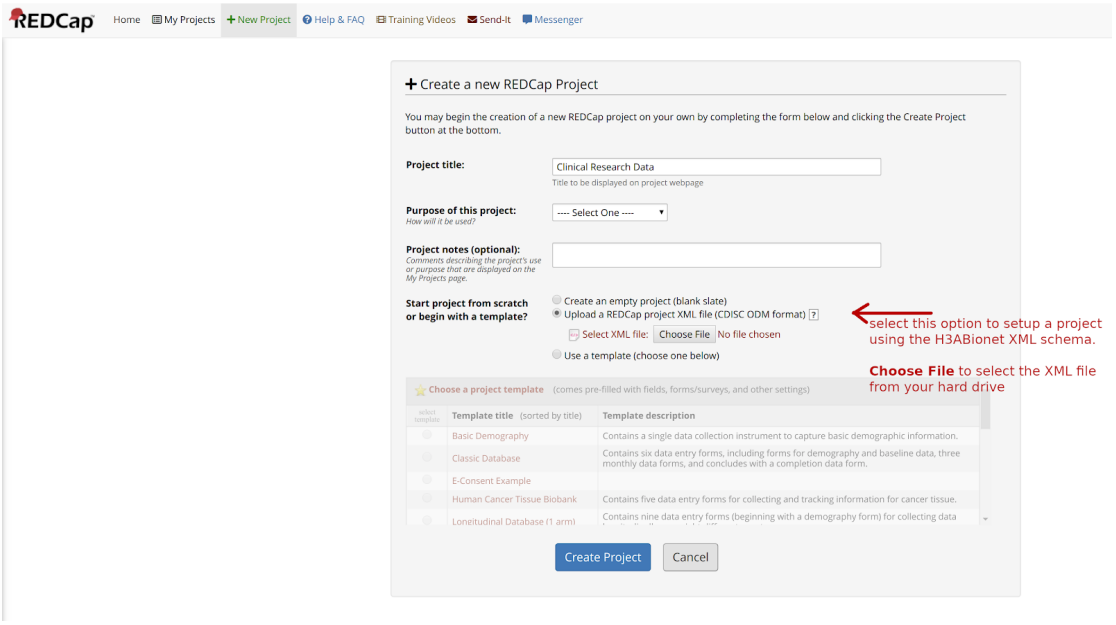
XML Implementation in REDCap

The XML schemas provided can be imported directly into REDCap to set up a data capturing project. Once imported, it will expose all the data capturing instruments and its associated documentation, just as would had it been done directly in REDCap.

The following steps outline the setup of a project using the reporting guidelines' XML schemas on REDCap:

1. Download the preferred XML schema from the H3ABioNet website to your hard drive.
2. Login to REDCap and create a new project

3. Complete the required fields on the first page. On the same page, choose the option to “Upload a REDCap project XML file (CDISC ODM format)” as shown below.



4. Select “Choose File”, browse to the location of the downloaded XML on your hard drive and select the file.
5. After completing step 4, click on “Create Project” and wait while the file is uploaded. REDCap should launch your new project when complete.

Frequently Asked Questions (FAQ)

The following section contains additional information as well as frequently asked questions relevant to the reporting guidelines and their use.

1. ***What is a data dictionary?***

   A data dictionary is a centralized repository of information describing the contents, format, and structure of a database and the relationship between its elements.

2. ***What is in the data dictionary?***

   The data dictionary describes each element included in the 3 sections of the domain-specific reporting guidelines, including its data type, accepted values and concordant ontology. See 'Overview' for more information.

3. ***Where can the reporting guidelines, and associated data dictionaries be found?***

   The reporting guidelines and associated data dictionaries are hosted on H3ABioNet's website (www.h3abionet.org/data-standards/datastds), but will also be submitted to the *FAIRsharing* portal. These reporting guidelines and data dictionaries may undergo updates, in which case it will be reflected on both websites.

4. ***How can the reporting guidelines and(or) data dictionaries be implemented at my institution?***

   To enable and promote the usability and accessibility of the reporting guideline, an XML schema was created. This XML schema can be implemented in *RedCap* (https://www.project-redcap.org/). See guidelines under 'XML Schemas'. The data dictionary is also presented in a user-friendly excel file, which separates sections into

sheets and can be used as guideline when creating data management plans, collection forms and(or) unique databases.

5. When to modify the data dictionary?

During its development, we aimed to make the reporting guidelines and associated data dictionaries as comprehensive as possible while also making it widely applicable. As such, it may still lack research-specific elements which are not covered by the reporting guideline. However, the guideline is meant to guide users to better produce and manage harmonized and standardised data, and can easily be adjusted for research-specific questions.

6. How to manage the research data before, during and after reporting?

The data reporting guidelines are employable during the development and initiation of research protocols. As the data load increases, it may require a dedicated data manager. The *Data Management Plans* (www.h3abionet.org/data-standards/datamanagement), as developed by H3ABioNet should be consulted in order to successfully manage data following the reporting/collection.