**CONSIDER Statement:**

**Consolidated Recommendations for Sharing Individual Participant Data from Human Clinical Studies**

This document provides a checklist and a set of recommendations to guide principal investigators (PIs), study team members and data sharing platform representatives in optimal sharing of individual participant data (IPD) from human clinical studies. We use the term study to include interventional trials and observational studies.

Recommendations are structured by several domains. Checklist below represents a brief summary of recommendations. The details link leads to elaborated descriptions of each item.

We welcome feedback to any checklist item at craig.mayer2 ‘at’ nih.gov

The acronym is loosely derived from letters contained in the title: **CONS**olidated **RE**commendations for sharing **I**ndividual participant **D**ata). Letters E and R are re-ordered to create a more memorable acronym.

# Data Format

## Share person table in CDISC or OMOP format

[details](#_Details_2), [positive-example](#_Positive_Example)

## For improved integration of research and routine healthcare data (e.g., from Electronic Health Record system or healthcare billing data), group relevant data elements into relevant data domains (e.g., medication history, laboratory results history, medical procedure history)

[details](#_Details_Emergence_of)

## Use consistent relative time (start counting at day 1, not 0)

[details](#_Details_3)

## Utilize ClinicalTrials.gov fields for uploading study protocol, empty case report forms, statistical analysis plan and study URL link

[details](#_Details_4), [positive-example](#_Positive_Example_4)

## Provide basic summary results using results registry component of Clinicaltrials.gov

[positive-example](#_Positive_Example_5)

## Share Case Report Forms in non-PDF, machine-readable format.

## [details](#_Details_5), [positive-example](#_Positive_Example_3), [challenge-example](#_Challenge_Example_6)

## If you considered formally defined research Common Data Elements at study design (more common for studies initiated after 2015), provide a spreadsheet file that lists all CDEs utilized by your study. Include unique CDE identifiers (e.g., PhenX VariableID).

### Description

### Positive Example

Data elements in AllofUs study are linked to LOINC codes. Elements are listed [here](http://athena.ohdsi.org/search-terms/terms?vocabulary=PPI&conceptClass=Clinical+Observation&page=1&pageSize=15&query=) and example of mapping can be seen for this [weight data element](http://athena.ohdsi.org/search-terms/terms/903121) (mapped to LOINC CDE [29463-7](https://loinc.org/29463-7/)).

### Challenge example

AllofUs study provides the source for each data element ([here](http://athena.ohdsi.org/search-terms/terms?vocabulary=PPI&conceptClass=Clinical+Observation&page=1&pageSize=50&query=)). Identifiers for individual CDE are not provided. Although, in many cases, the instrument does not have a formal identifier that could be listed.

## Use data formats that can be natively loaded (without add-ons) into multiple statistical platforms

[Detail](#_The_preferred_file), [positive example](#_Positive_Example_6), [challenge example](#_Challenge_Example_5)

# Data Sharing

## Use ClinicalTrials.gov registry to document your study

[positive-example](#_Positive_Example_1)

## Do not limit study metadata to the legally required elements. Also populate optional elements (such as data sharing metadata)

## On ClinicalTrials.gov, if you answer Yes to share\_ipd\_data, do not leave the data\_sharing\_plan text empty

[challenge-example](#_Challenge_Example_2)

## If Individual Participant Data is shared on a data sharing platform, update the ClinicalTrials.gov record with the URL link to the data.

### Details

At ClinicalTrials.gov, use the following infrastructure:

"AvailIPDList": {

"AvailIPD": [

{

"AvailIPDId": "NLMS",

"AvailIPDType": "Individual Participant Data Set",

"AvailIPDURL": "YOUR ULR GOES HERE",

"AvailIPDComment": "INSTRUCTIONS"

}

### Positive example

Study <https://clinicaltrials.gov/ct2/show/study/NCT00005159> provides link to IPD at Biolincc repository.

"AvailIPDList": {

                "AvailIPD": [

                  {

                    "AvailIPDId": "NLMS",

                    "AvailIPDType": "Individual Participant Data Set",

                    "AvailIPDURL": "<http://biolincc.nhlbi.nih.gov/studies/nlms/>",

                    "AvailIPDComment": "Public-use IPD through BioLINCC. Registration is required for access."

                  }

# Study design

## At study design time and when resources allow, adopt previously defined applicable common data elements

### Details

Common Data Elements initiatives in various domains (e.g., [PhenX](https://www.phenxtoolkit.org), PROMIS) aim to standardize data collection.

# Case Report Forms

## Share all Case Report Forms used in a study

### Details

Interpretation of data is easier if full conte

### Positive example

UK Biobank provides screenshot of the form (see <http://biobank.ctsu.ox.ac.uk/showcase/refer.cgi?id=158595>) for data element <http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=22501>

This allows researchers understand why value -1: never went to school appears in the data since the form shows a check box near the question that allows research participant to indicate that they never went to school.

ClinicalTrials.gov allows upload of resource type: annotated CRFs.

# Data dictionary

## Provide data dictionary documentation separate from de-identified individual participant data. Since it contains no participant level data, do not require local ethical approval as a condition of releasing the data dictionary (avoid a requestwall for data dictionary).

[details](#_Details_If_DD), [positive-example](#_Positive_Example_9)

## Share a data dictionary as soon as possible. Do not wait until the data collection is complete.

### Details

Data dictionary has scientific value for other studies in the field and can help such studies speed up their study design phase. For registry studies (without a fixed end date), early sharing is equally important.

### Positive example

AllofUs research study shared their Case Report Forms at the start of the study (see <https://www.researchallofus.org/data-sources/survey-explorer>).

## Provide data dictionary in a single, machine-readable file.

[details](#_Details_This_simplifies), [positive-example](#_Positive_Example_8), [challenge-example](#_Challenge_Example_3)

## For each data element, provide a data type (such as numeric, date, string, categorical)

[Details](#_Details_Specifying_data), [positive-example,](#_Positive_Example_7) [challenge-example](#_Challenge_example_4)

## For categorical data elements, provide a list of permissible values and distinguish when numerical code or string code is a code for a permissible value (versus actual number or string)

[Details](#_Details_1), [positive example](#_Positive_example_10), [challenge example](#_Challenge_example)

## Link data elements (including individual permissible values) to applicable routine healthcare terminologies or research common data elements

[Details](#_Details_This_recommendation), [positive example](#_Positive_example_AllOfUs), [Challenge Example](#_challenge_example_1)

## Provide complete data dictionary (all elements in data are listed in a dictionary) and all types of applicable dictionaries (date elements, forms [or groupings], and permissible values)

### Details

Note: NIDA trial had 5 data files that were missing a data dictionary file.

### Challenge example

## Include sufficient description for data elements

### Details

In some cases, a descriptive name can be sufficient to define a data element and interpreting a data element is straightforward. However, if a study has two distinct data elements with identical name, data can be hard to interpret.

### Positive example

All elements in UK Biobank have description. See <http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=22501> for example for data element titled ‘Year ended full time education’) has the following detailed description: ‘Some values have special meanings defined by Data-Coding 100306. Units of measurement are calendar year.’

### Challenge example

In study NCT01751646: ‘Vitamin D Absorption in HIV Infected Young Adults Being Treated With Tenofovir Containing cART’: a description for forms C100 and B100 both state ‘Specimen Tracking Form’. This makes it to interpret if the data represent the same or two distinct specimen. Avoid identical descriptions for 2 separate items.

## Use identifiers (unique where applicable) for data element, forms and permissible values.

### Details

Permissible value data dictionary should be linked to data element data dictionary. Provide annonated case report forms

# Data de-identification

## Provide data de-identification notes

positive-example, challenge-example, details

# Choice of a Data Sharing platform

## Choose a platform that allows download of all studies available on the platform

### Positive example

ClinicalStudyDataRequest.com platform provides a spreadsheet download with a list of all studies included on the platform. (available at <https://www.clinicalstudydatarequest.com/Documents/All%20Sponsor-Funder%20Studies.xls>)

## Choose a platform that supports batch request (ability to request multiple studies with one request)

# Acknowledgement

This document is part of training materials objective of a project funded by NIH Office of AIDS Research titled: Identification of Research Common Data Elements in HIV/AIDS using data science methods.

# Abbreviations

DD = data dictionary

IPD=Individual Participant Data

# Details

## Data format

### Share Person table

#### Details

OMOP person table specification can be found at <https://github.com/OHDSI/CommonDataModel/wiki/PERSON>

#### Positive Example

Trial <http://clinicaltrials.gov/ct2/show/NCT01612169> IPV data posted on NIDA Data Share (at <https://datashare.nida.nih.gov/study/nidactn0049>) provide file dem.csv provides one row per person with person\_id and basic demographic data. (following CDISC standard).

### For improved integration of research and routine healthcare data (e.g., from Electronic Health Record system or healthcare billing data), group relevant data elements into relevant data domains (e.g., medication history, laboratory results history, medical procedure history)

#### Details

#### Emergence of several common data models (CDMs) for healthcare data shows that there is a common way to organize clinical data. For example, rather than considering numerical value as test result and unit in which the value is expressed as two separate data elements, a review of several CDMs shows that these are typically grouped into a single data row.

### Use consistent relative time (start counting at day 1, not 0)

#### Details

#### When using relative time, specify index event with datetime granularity. Refer to the first day as relative day 1. Do not use day 0 as a relative date. For example if the index event is signing of informed consent and it was signed at 10:31am on March 10, 2011, the index date-time is midnight of March 10, 2011. In relative time, of an event on the next day (on March 11) at 11:15am, would have relative time of Day 2, 11:15am. (for analogous discussion in astronomical data see <https://en.wikipedia.org/wiki/Sol_(day_on_Mars)#Usage_in_Mars_landers)>

### Utilize ClinicalTrials.gov fields for uploading study protocol, empty case report forms, statistical analysis plan and link to

#### Details

ClinicalTrials.gov allows for the upload of relevant documents related to the study page including protocols, analysis plans and other related documents while also allowing for the providing of links to relevant materials such as data dictionaries and IPD.

#### Positive Example

NCT02755818 provides documents, such as protocol, informed consent form and results as both links to external sites and uploads to ClinicalTrials.gov.

### Provide basic summary results using results registry component of Clinicaltrials.gov

#### Positive Example

NCT00962780 has basic summary results of the study posted on ClinicalTrials.gov using the results registry component

### Share Case Report Forms in non-PDF, machine-readable format.

#### Details

Machine-readable Case Report Forms allow for easier integration and use os the information provided from the empty Case Report Forms

#### Positive Example

#### For example, REDCap, OpenClinica and several other Electronic Data Capture (EDC) systems allow export into CDISC Operational Data Model (ODM) format for forms. If no cross-platform standard is supported by your EDC, provide CRFs in the platform-specific format.

#### Challenge Example

NCT00005273 provides case report forms as PDFS of photocopied forms making any automated integration or machine reading of the information not plausible.

### If you considered formally defined research Common Data Elements at study design (more common for studies initiated after 2015), provide a spreadsheet file that lists all CDEs utilized by your study. Include unique CDE identifies (e.g., PhenX VariableID).

### Use data formats that can be natively loaded (without add-ons) into multiple statistical platforms

#### Details

#### The preferred file taps are comma/tab separated values (.CSV) filesinstead of SAS XPT, XLS/XSLX), which require add ons or conversions to be read in and used in differet statistical platforms

#### Positive Example

#### NCT01751646 provides IPD in CSV files easily usable in a multitude of statistical platforms. Trial <https://clinicaltrials.gov/ct2/show/study/NCT00005159> provides 3 formats for study data: .CSV, .SAS7BDAT, and .dat (text file with fixed width)

#### Challenge Example

NCT00951249 provides IPD as SAS XPT files which require processing and conversion to use in any non-SAS platform for view and analysis

## Data Sharing

### Use ClinicalTrials.gov registry to document your study

#### Positive Example

#### Trial titled “RC-HIVMAB060-00-AB (VRC01) in People With Chronic HIV Infection Undergoing Analytical Treatment Interruption” is registered at ClinicalTrials.gov under NCT02471326 <https://clinicaltrials.gov/ct2/show/study/NCT02471326>. It allows retrieval of study metadata that is unified across various platform.

### Missing sharing plan description

#### Challenge Example

#### Out of 69 reviewed HIV trials IPD sharing plans on ClinicalTrials.gov, 2 (NCT02756208 and NCT03275701) left the plan description blank.

## Data dictionary

### Provide data dictionary documentation separate from de-identified individual participant data. Since it contains no participant level data, do not require local ethical approval as a condition of releasing the data dictionary (avoid a request-wall).

#### Details If DD does contain important intellectual property (IP), consider creating a smaller list of DEs that do not contain any IP and release this limited subset of DEs without employing a request-wall.

#### Positive Example

NCT01769456 has the data dictionary on the data sharing platform and is available for download without requiring any request, approval or the filling out of any documents. Another example for study [NCT00005159](https://clinicaltrials.gov/ct2/show/study/NCT00005159) is [here](https://biolincc.nhlbi.nih.gov/media/studies/nlms/Code_Manuals_and_Forms.pdf).

### Scattered data dictionary

#### Details This simplifies machine processing of available study data. Using a single file approach also ensures that each file (if scattered across multiple) uses the same structure (e.g., DE label, DE data type, DE permissible values [for categorical DEs])

#### Positive Example

NCT01772823 provides a single data dictionary document as a CSV containing all data elements and information pertaining to the data elements such as data type and description

#### Challenge Example

#### Trials NCT00005274 and NCT00005274 provided DD in several files. In order to relate those to CDEs, manual processing is required.

#### NCT01233531 has 17 data dictionary files, which includes documents in different formats and document types. Also includes identical file names that represent dictionaries for different data based on visit type and study population group. NIDA Data Share trial had 65 data files but only 63 data dictionary PDF files. Matching data files with data dictionary requires manual matching.

### For each data element, provide a data type

#### Details Specifying data type helps computers to process the information properly. Data type also helps with semantic matching to corresponding CDEs. For example, date of death data type is stated as date (not as character). Most studies collect categorical data types.

#### Positive Example

NCT00491556 provides the data type for each of the data elements in the DD

UK Biobank uses a comprehensive set of data types available at <http://biobank.ctsu.ox.ac.uk/showcase/help.cgi?cd=value_type>  
The UK Biobank description and listed types are in italic below

*The Value Type of a Data-Field describes the type of variable corresponding to it. There are 9 categories:*

* *Integer - whole numbers, for example the age of a participant on a particular date;*
* *Categorical (single) - a single answer selected from a coded list or tree of mutually exclusive options, for example a yes/no choice;*
* *Categorical (multiple) - sets of answers selected from a coded list or tree of options, for instance concurrent medications;*
* *Continuous - floating-point numbers, for example the height of a participant;*
* *Text - data composed of alphanumeric characters, for example the first line of an address;*
* *Date - a calendar date, for example 14th October 2010;*
* *Time - a time, for example 13:38:05 on 14th October 2010;*
* *Compound - a set of values required as a whole to describe some compound property, for example an ECG trace;*
* *Binary object - a complex dataset (blob), for example an image;*
* *Records - a summary showing the volume of records data available via the secure portal.*

#### Challenge example

#### NCT00046280 does not provide data type.

### For categorical data elements, provide a list of permissible values

#### Details

### For example, for educational level data element, it is important to know what possible values were considered during data collection. While it is possible to discover those permissible values from IPD, if some values were never applicable to any of the subjects, the reverse-engineered permissible value list will be incomplete. In terms of standards, CDISC ODM and REDCap provide a mechanism to list permissible values.

#### Positive example

### NCT00683579 provides permissible values and definitions for the values associated with categorical data elements.

#### Challenge example

### NCT00000590 does not provide permissible vales for categorical variables in the data dictionary. Another [example](https://biolincc.nhlbi.nih.gov/media/studies/nlms/Code_Manuals_and_Forms.pdf) is providing permissible value in the same document that describes the data elements, but in non-machine readable way.

### Link data elements (including individual permissible values) to applicable routine healthcare terminologies or research common data elements

#### Details

### This recommendation does not mean that all DEs (and permissible values) are linked to a terminology or applicable standard. Only where a relevant code exist (or is easy to find and reference) this link is recommended.

#### Positive example

### AllOfUs study links DEs to SNOMED CT and LOINC codes. For example, Body Mass Index data element is clearly linked to LOINC code of 39156-5 (for BMI)

#### Challenge example

### NCT01751646 has no link to any standardized vocabulary

### Distinguish string field with free text from string field (with a set of permissible values)

### Provide a complete data dictionary (all data files must be included)

Note: NIDA trial had 5 data files that were missing a data dictionary file.