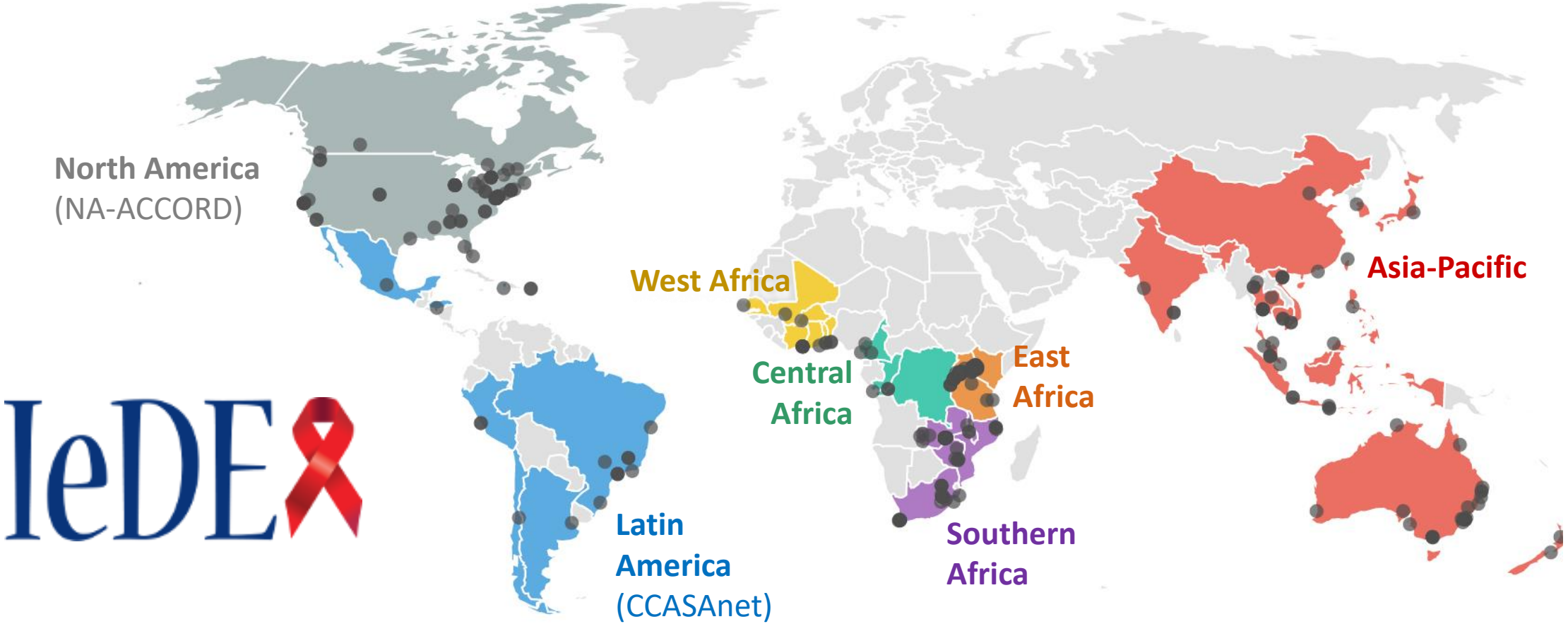
Streamlining Data Sharing in a Global HIV Research Consortium

Stephany Duda PhD and Judy Lewis PhD

Vanderbilt University Medical Center

October 1, 2021

IeDEA – A global HIV research consortium



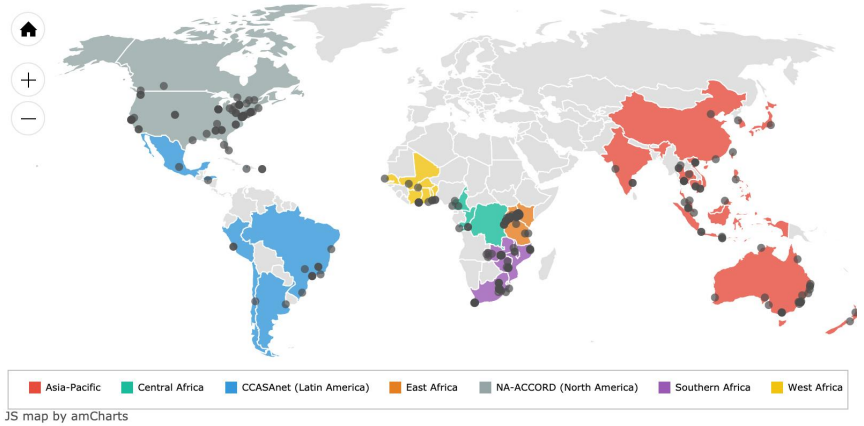
IeDEA

International epidemiology Databases to Evaluate AIDS

A global, NIAID-funded collaboration of seven regional HIV observational research networks with combined data on nearly two million persons living with HIV (PLWH)

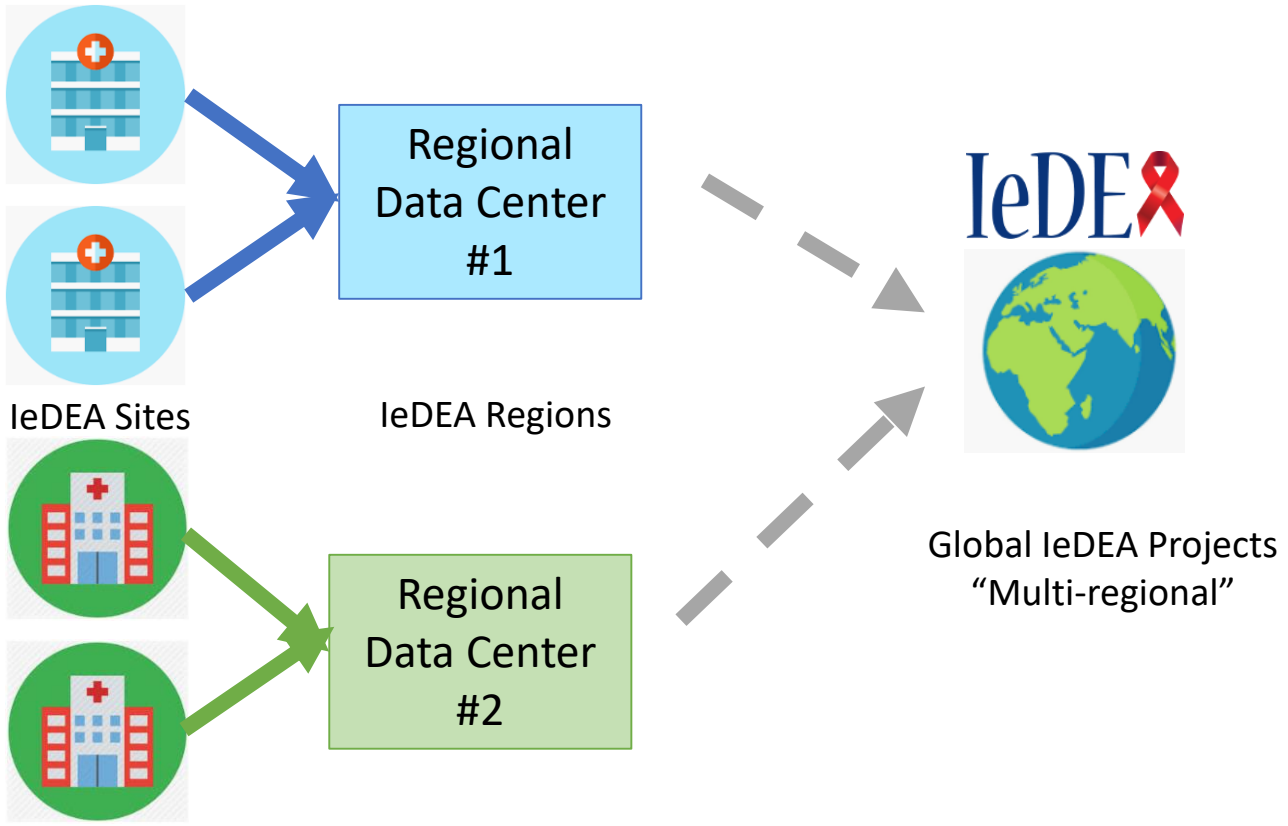
IeDEA is a collaboration, not a single network

- Each region operates independently, has its own coordinating center, and conducts regional research.
- Regions come together for multiregional research projects.
- There is no overall coordinating center for IeDEA (as of July 2016).
- There has never been an overall data coordinating center for IeDEA



(This turned out to be great)

Flow of IeDEA data

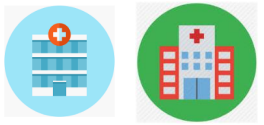


In IeDEA

- Sites generate the data.
- Regional Data Centers combine all the data from one region.
- Researchers can get data from multiple regions for a global IeDEA project.

Data Considerations

- Data from every clinic can be different.



- Data at every Regional Data Center can be different.

Regional Data Center #1

Regional Data Center #2

- Global IeDEA data are not stored centrally – subsets of the data are merged for specific projects.
- Sites and Regions have the ultimate say in whether their data is included for a specific project.

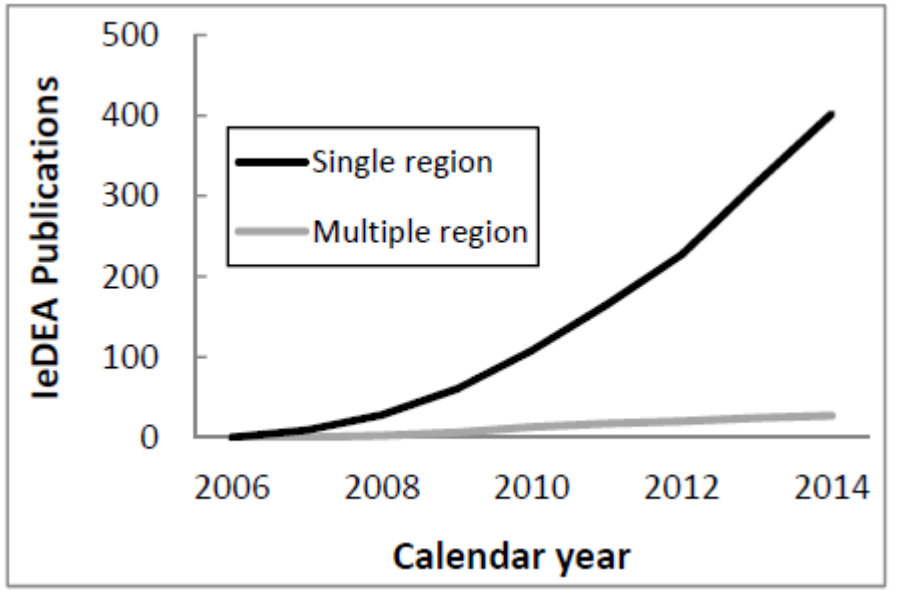
IeDEA

In the Early Days of IeDEA…

- We had no standardized way to share data for global projects.
- Multi-regional projects (projects with 3+ IeDEA regions) were very slow, in part because it was difficult to merge the data.

Cumulative number of IeDEA publications by publication year

(figure from Constantin Yiannoutsos)



The Harmonist Project

Researchers wanted to collaborate efficiently, but they needed the right tools to support them. We focused on three areas:

(1) Support for Data Model *IeDEA Data Exchange Standard*	→	Clear data standards
(2) Consortium Portal *Harmonist Hub*	→	Managed research workflow
(3) Data Quality and Sharing Tool *Harmonist Data Toolkit*	→	High quality data

Goal: address IeDEA's needs but always look for generalizable solutions that can benefit others.

The Harmonist Project

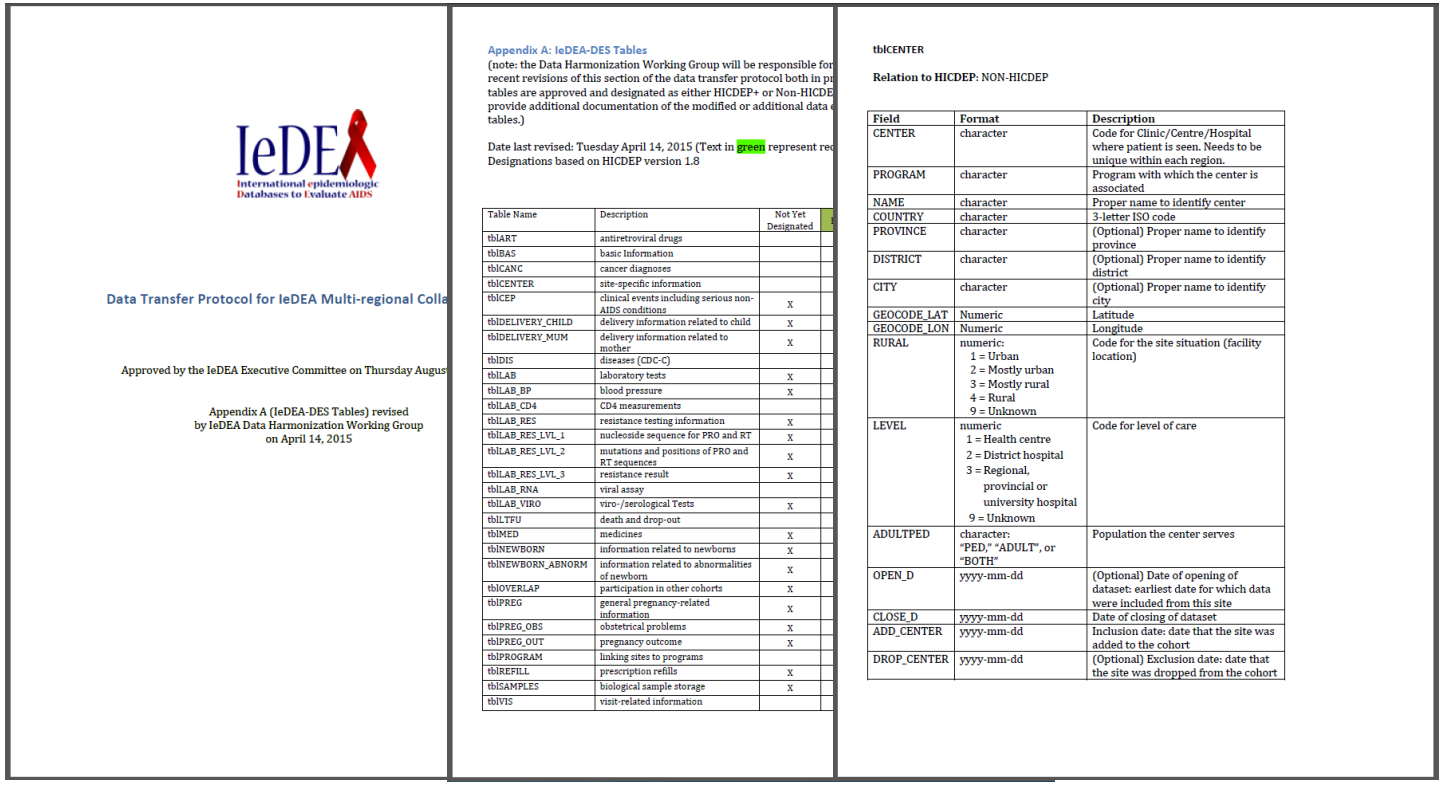
(1) Support for Data Model
IeDEA Data Exchange Standard → Clear data standards

(2) Consortium Portal
Harmonist Hub → Managed research workflow

(3) Data Quality and Sharing Tool
Harmonist Data Toolkit → High quality data

(1) IeDEA Data Exchange Standard (DES)

The **IeDEA Data Exchange Standard (DES)** is a common data model for HIV data sharing within IeDEA. The document defines variable names, variable definitions, and code lists.



IeDEA
International epidemiologic Databases to Evaluate AIDS

Data Transfer Protocol for IeDEA Multi-regional Colla

Approved by the IeDEA Executive Committee on Thursday August

Appendix A (IeDEA-DES Tables) revised
by IeDEA Data Harmonization Working Group
on April 14, 2015

Appendix A: IeDEA-DES Tables

(note: the Data Harmonization Working Group will be responsible for
recent revisions of this section of the data transfer protocol both in pr
tables are approved and designated as either HICDEP+ or Non-HICDE
provide additional documentation of the modified or additional data
tables.)

Date last revised: Tuesday April 14, 2015 (Text in green represent rec
Designations based on HICDEP version 1.8

Table Name	Description	Not Yet Designated
tblART	antiretroviral drugs	
tblBAS	basic Information	
tblCANC	cancer diagnoses	
tblCENTER	site-specific information	
tblCEP	clinical events including serious non-AIDS conditions	X
tblDELIVERY_CHILD	delivery information related to child	X
tblDELIVERY_MUM	delivery information related to mother	X
tblDIS	diseases (CDC-C)	
tblLAB	laboratory tests	X
tblLAB_BP	blood pressure	X
tblLAB_CD4	CD4 measurements	
tblLAB_RES	resistance testing information	X
tblLAB_RES_LVL_1	nucleoside sequence for PRO and RT	X
tblLAB_RES_LVL_2	mutations and positions of PRO and RT sequences	X
tblLAB_RES_LVL_3	resistance result	X
tblLAB_RNA	viral assay	
tblLAB_VIRO	viro-/serological Tests	X
tblLTFU	death and drop-out	
tblMED	medicines	X
tblNEWBORN	information related to newborns	X
tblNEWBORN_ABNORM	information related to abnormalities of newborn	X
tblOVERLAP	participation in other cohorts	X
tblPREG	general pregnancy-related information	X
tblPREG_OBS	obstetrical problems	X
tblPREG_OUT	pregnancy outcome	X
tblPROGRAM	linking sites to programs	
tblREFILL	prescription refills	X
tblSAMPLES	biological sample storage	X
tblVIS	visit-related information	

tblCENTER

Relation to HICDEP: NON-HICDEP

Field	Format	Description
CENTER	character	Code for Clinic/Centre/Hospital where patient is seen. Needs to be unique within each region.
PROGRAM	character	Program with which the center is associated
NAME	character	Proper name to identify center
COUNTRY	character	3-letter ISO code
PROVINCE	character	(Optional) Proper name to identify province
DISTRICT	character	(Optional) Proper name to identify district
CITY	character	(Optional) Proper name to identify city
GEOCODE_LAT	Numeric	Latitude
GEOCODE_LON	Numeric	Longitude
RURAL	numeric: 1 = Urban 2 = Mostly urban 3 = Mostly rural 4 = Rural 9 = Unknown	Code for the site situation (facility location)
LEVEL	numeric 1 = Health centre 2 = District hospital 3 = Regional, provincial or university hospital 9 = Unknown	Code for level of care
ADULTPED	character: "PED," "ADULT", or "BOTH"	Population the center serves
OPEN_D	yyyy-mm-dd	(Optional) Date of opening of dataset: earliest date for which data were included from this site
CLOSE_D	yyyy-mm-dd	Date of closing of dataset
ADD_CENTER	yyyy-mm-dd	Inclusion date: date that the site was added to the cohort
DROP_CENTER	yyyy-mm-dd	(Optional) Exclusion date: date that the site was dropped from the cohort

Maintaining the IeDEA DES

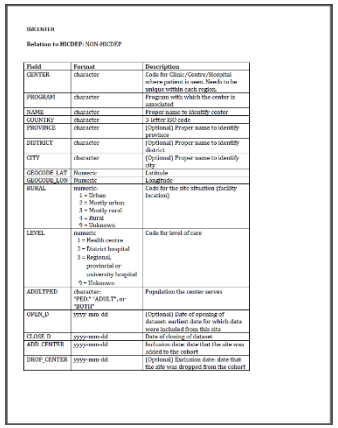
- Challenges with MS Word documents
  - Multiple versions, potentially conflicting edits
  - Hard to find latest version in files, email
  - Single copy is not group editable
  - Not machine-readable

IeDEA

Data Transfer Protocol for IeDEA Multi-regional Collaborations

Approved by the IeDEA Executive Committee on Thursday August 16, 2012

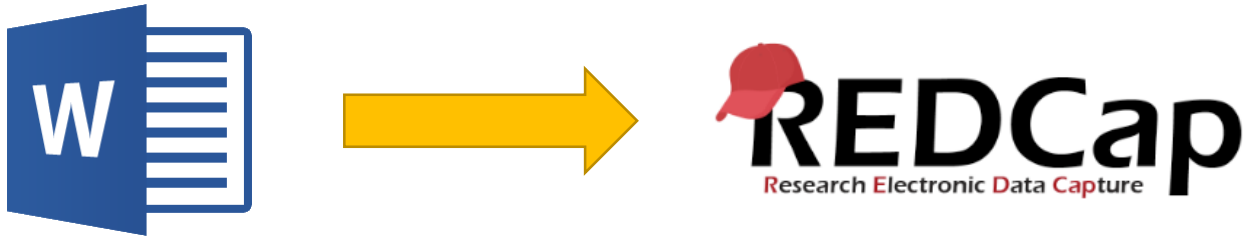
Appendix A (IeDEA-DES Tables) revised
by IeDEA Data Harmonization Working Group
on April 14, 2015


tblCENTER

Relation to HICDEP: NON-HICDEP

Field	Format	Description
CENTER	character	Code for Clinic/Centre/Hospital where patient is seen. Needs to be unique within each region.
PROGRAM	character	Program with which the center is associated
NAME	character	Proper name to identify center
COUNTRY	character	3 letter ISO code
PROVINCE	character	(Optional) Proper name to identify province
DISTRICT	character	(Optional) Proper name to identify district
CITY	character	(Optional) Proper name to identify city
GEOCODE_LAT	Numeric	Latitude
GEOCODE_LON	Numeric	Longitude
RURAL	numeric: 1 = Urban 2 = Mostly urban 3 = Mostly rural 4 = Rural 9 = Unknown	Code for the site situation (fertility location)
LEVEL	numeric 1 = Health centre 2 = District hospital 3 = Regional, provincial or university hospital 9 = Unknown	Code for level of care
ADULTPED	character: "PED","ADULT", or "BOTH"	Population the center serves
OPEN_D	yyyy mm dd	(Optional) Date of opening of dataset: earliest date for which data were included from this site
CLOSE_D	yyyy-mm-dd	Date of closing of dataset
ADD_CENTER	yyyy-mm-dd	Inclusion data: date that the site was added to the cohort
DROP_CENTER	yyyy-mm-dd	(Optional) Exclusion date: date that the site was dropped from the cohort

- Needed a machine-readable solution that was easy to edit and didn't require technical training.
- **Solution:** Use REDCap to create human-readable forms that produce machine-readable structures



What is REDCap ?

Research Electronic Data Capture

~5440 sites
142 countries
1.3M users

projectredcap.org

Representing the IeDEA DES in REDCap

To represent the DES in REDCap, we designed three data entry forms:

1. Information about Tables (e.g, demographics, visits, labs, meds)
2. Information about Variables
3. Information about Code Lists

Example: Tables

Harmonist 0A: Data Model (IeDEA-DES)

Actions: Modify instrument | Download PDF of instrument(s)

Table Metadata

Editing existing Record ID **3** tblBAS

Record ID

3

To rename the record, see the record action drop-down at top of the Record Home Page.

Table Definition

Table name

tblBAS

Table type

- (•) One row per patient
- ( ) Multiple rows per patient
- ( ) N/A (e.g., tblCENTER)

reset

Table definition
(brief text)

Basic information

Display Settings

Display this table in human-readable documents and forms?

- (•) Yes
- ( ) No

reset

Table display name
(optional, if different title is needed for human-readable documents)

Display order for this table

1

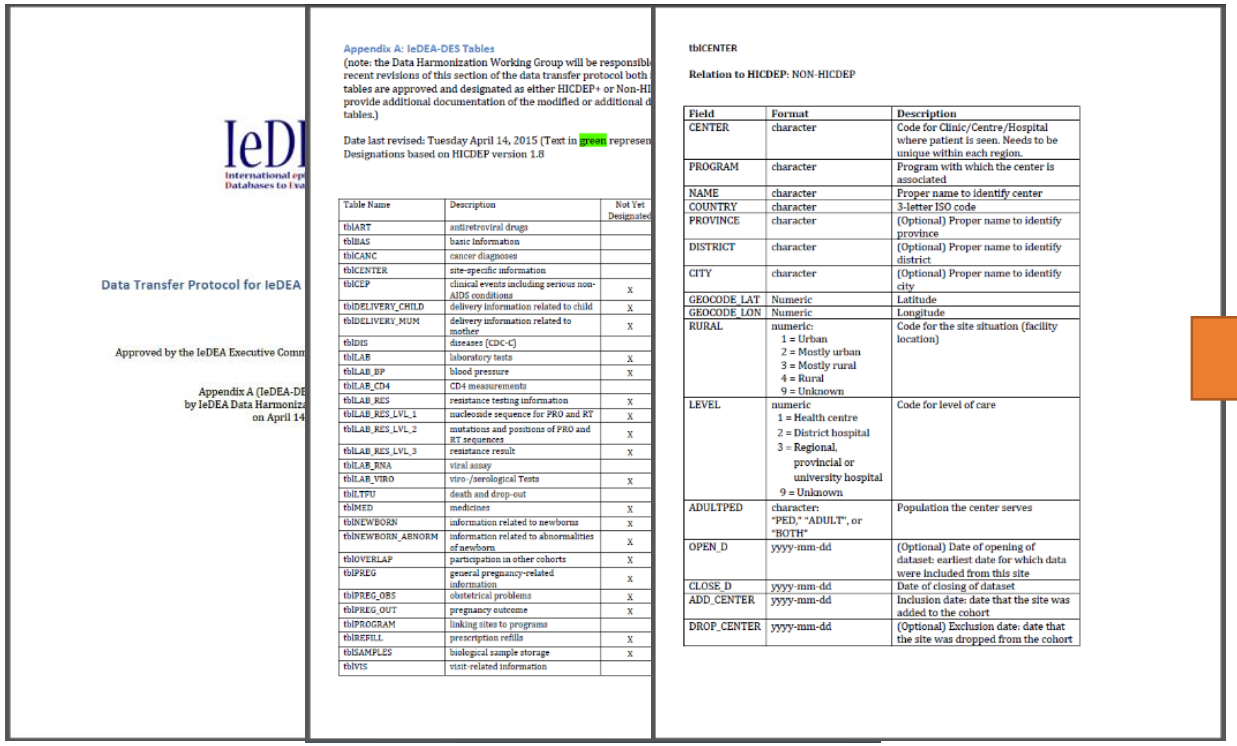
(e.g., use 1.5 to place a table between tables 1 and 2)

Text (HTML-formatted) to display before the table definiton

Every Patient ID must have one and only one entry in tblBAS.

Expand

Initial REDCap Mapping



Data Transfer Protocol for IeDEA

Approved by the IeDEA Executive Comm

Appendix A (IeDEA-DE
by IeDEA Data Harmoniz
on April 14

Appendix A: IeDEA-DES Tables

(note: the Data Harmonization Working Group will be responsibl
recent revisions of this section of the data transfer protocol both
tables are approved and designated as either HICDEP+ or Non-Hi
provide additional documentation of the modified or additional d
tables.)

Date last revised: Tuesday April 14, 2015 (Text in green represen
Designations based on HICDEP version 1.8

Table Name	Description	Not Yet Designated
tblART	antiretroviral drugs	
tblBAS	basic information	
tblCANC	cancer diagnoses	
tblCENTER	site-specific information	
tblCEP	clinical events including serious non-AIDS conditions	X
tblDELIVERY_CHILD	delivery information related to child	X
tblDELIVERY_MUM	delivery information related to mother	X
tblDIS	diseases (CDC-C)	
tblLAB	laboratory tests	X
tblLAB_BP	blood pressure	X
tblLAB_CD4	CD4 measurements	
tblLAB_RES	resistance testing information	X
tblLAB_RES_LVL_1	nucleoside sequence for PRO and RT	X
tblLAB_RES_LVL_2	mutations and positions of PRO and RT sequences	X
tblLAB_RES_LVL_3	resistance result	X
tblLAB_RNA	viral assay	
tblLAB_VIRO	viro-/serological Tests	X
tblLTFU	death and drop-out	
tblMED	medicines	X
tblNEWBORN	information related to newborns	X
tblNEWBORN_ABNORM	information related to abnormalities of newborn	X
tblOVERLAP	participation in other cohorts	X
tblPREG	general pregnancy-related information	X
tblPREG_OBS	obstetrical problems	X
tblPREG_OUT	pregnancy outcome	X
tblPROGRAM	linking sites to programs	
tblREFILL	prescription refills	X
tblSAMPLES	biological sample storage	X
tblVIS	visit-related information	

tblCENTER

Relation to HICDEP: NON-HICDEP

Field	Format	Description
CENTER	character	Code for Clinic/Centre/Hospital where patient is seen. Needs to be unique within each region.
PROGRAM	character	Program with which the center is associated
NAME	character	Proper name to identify center
COUNTRY	character	3-letter ISO code
PROVINCE	character	(Optional) Proper name to identify province
DISTRICT	character	(Optional) Proper name to identify district
CITY	character	(Optional) Proper name to identify city
GEOCODE_LAT	Numeric	Latitude
GEOCODE_LON	Numeric	Longitude
RURAL	numeric: 1 = Urban 2 = Mostly urban 3 = Mostly rural 4 = Rural 9 = Unknown	Code for the site situation (facility location)
LEVEL	numeric 1 = Health centre 2 = District hospital 3 = Regional, provincial or university hospital 9 = Unknown	Code for level of care
ADULTPED	character: "PED," "ADULT", or "BOTH"	Population the center serves
OPEN_D	yyyy-mm-dd	(Optional) Date of opening of dataset: earliest date for which data were included from this site
CLOSE_D	yyyy-mm-dd	Date of closing of dataset
ADD_CENTER	yyyy-mm-dd	Inclusion date: date that the site was added to the cohort
DROP_CENTER	yyyy-mm-dd	(Optional) Exclusion date: date that the site was dropped from the cohort

Word version of the IeDEA DES

Record ID	Table Metadata	Variable Metadata
1 tblART		
2 tblART_MUM		
3 tblBAS		
4 tblCANC		
5 tblCENTER		
6 tblDELIVERY_CHILD		
7 tblDELIVERY_MUM		
8 tblDIS		
9 tblLAB		
10 tblLAB_BP		
11 tblLAB_CD4		
12 tblLAB_RES		
13 tblLAB_RES_LVL_2		
14 tblLAB_RES_LVL_3		
15 tblLAB_RNA		
16 tblLAB_VIRO		
17 tblLTFU		
18 tblMED		
19 tblNEWBORN		
20 tblNEWBORN_ABNOR		
21 tblPREG		

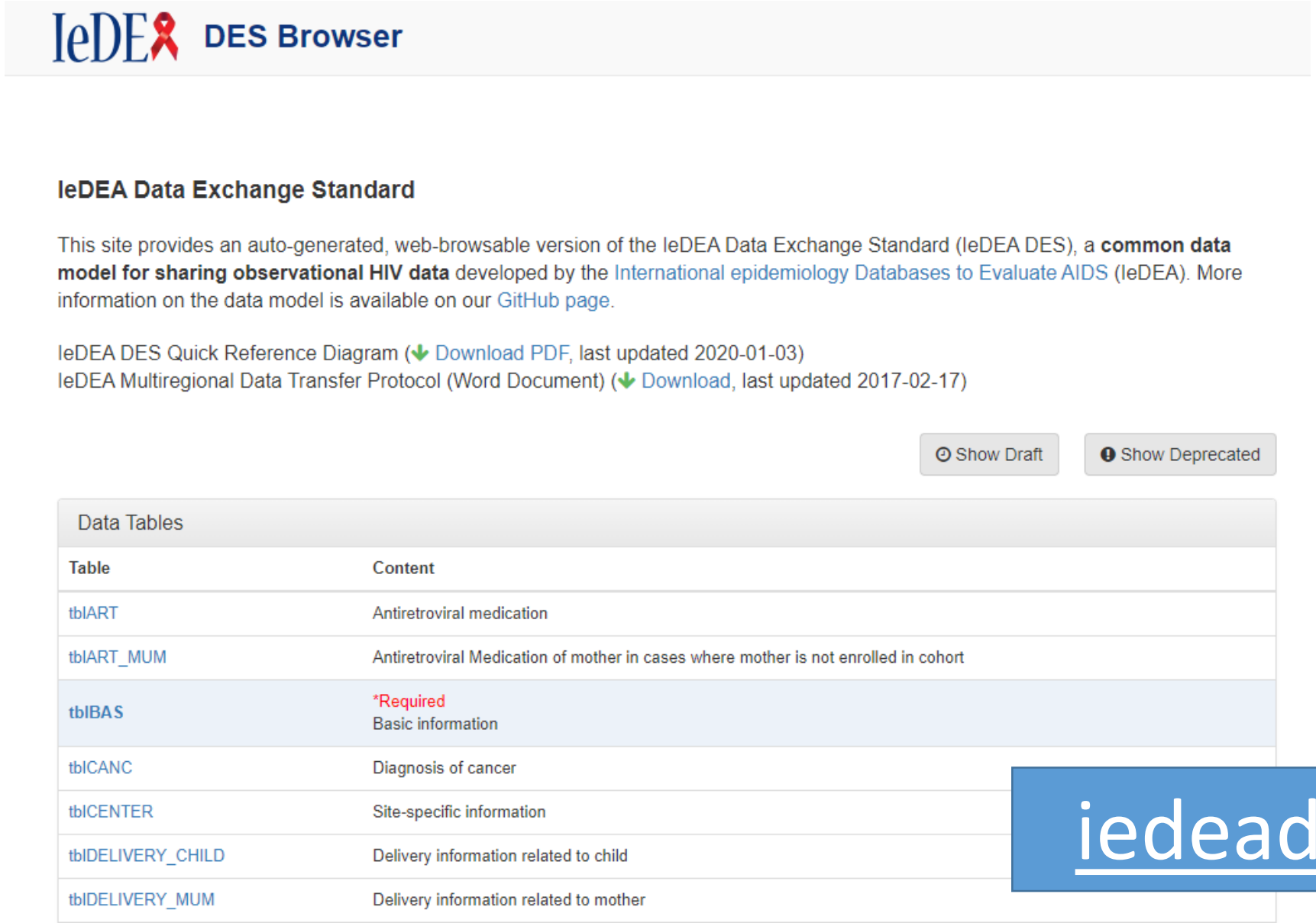
REDCap
Research Electronic Data Capture

Part of Machine-Readable DES

(JSON formatted export via REDCap API)


```
[{"record_id":"1","redcap_repeat_instrument":"","redcap_repeat_instance":"","table_name":"tblART","table_format"
:"2","table_definition":"antiretroviral
medication","table_display_y":"1","table_display_name":"","table_order":"2","text_top":"","text_bottom":"","tabl
e_deprecated___1":"0","table_deprecated_d":"","table_metadata_complete":"2","table_link":"","variable_name":"","
data_format":"","description":"","variable_order":"","code_text":"","variable_required___1":"","variable_depreca
ted___1":"","variable_deprecated_d":"","has_codes":"","code_format":"","code_list":"","code_file":"","codes_prin
t":"","variable_metadata_complete":""},{"record_id":"1","redcap_repeat_instrument":"variable_metadata","redcap_r
epeat_instance":1,"table_name":"","table_format":"","table_definition":"","table_display_y":"","table_display_na
me":"","table_order":"","text_top":"","text_bottom":"","table_deprecated___1":"","table_deprecated_d":"","table_
metadata_complete":"","table_link":"1","variable_name":"PATIENT","data_format":"1","description":"Code to
identify patient (Cohort Patient
ID)","variable_order":"1","code_text":"","variable_required___1":"1","variable_deprecated___1":"0","variable_dep
recated_d":"","has_codes":"0","code_format":"","code_list":"","code_file":"","codes_print":"","variable_metadata
_complete":"2"},{"record_id":"1","redcap_repeat_instrument":"variable_metadata","redcap_repeat_instance":2,"tabl
e_name":"","table_format":"","table_definition":"","table_display_y":"","table_display_name":"","table_order":""
,"text_top":"","text_bottom":"","table_deprecated___1":"","table_deprecated_d":"","table_metadata_complete":"","
table_link":"1","variable_name":"ART_ID","data_format":"1","description":"ATC code for the patient's ARV
medication","variable_order":"2","code_text":"Codes are drawn from the Anatomical Therapeutic Chemical (ATC)
classification system. Use ART code list maintained by HICDEP:
http://www.hicdep.org/wiki/Hicdep_1.100/TableArt/FieldArtId#CodingTable A searchable index is available at
http://www.whocc.no/atc_ddd_index/","variable_required___1":"1","variable_deprecated___1":"0","variable_deprecat
ed_d":"","has_codes":"1","code_format":"3","code_list":"","code_file":"","codes_print":"0","variable_metadata_co
mplete":"0"},{"record_id":"1","redcap_repeat_instrument":"variable_metadata","redcap_repeat_instance":3,"table_n
ame":"","table_format":"","table_definition":"","table_display_y":"","table_display_name":"","table_order":"","t
ext_top":"","text_bottom":"","table_deprecated___1":"","table_deprecated_d":"","table_metadata_complete":"","tab
le_link":"1","variable_name":"ART_SD","data_format":"4","description":"Date of initiation of
treatment","variable_order":"3","code_text":"","variable_required___1":"1","variable_deprecated___1":"0","variab
le_deprecated_d":"","has_codes":"0","code_format":"","code_list":"","code_file":"","codes_print":"","variable_me
tadata_complete":"2"},{"record_id":"1","redcap_repeat_instrument":"variable_metadata","redcap_repeat_instance":4
,"table_name":"","table_format":"","table_definition":"","table_display_y":"","table_display_name":"","table_ord
er":"","text_top":"","text_bottom":"","table_deprecated___1":"","table_deprecated_d":"","table_metadata_complete
":"","table_link":"1","variable_name":"ART_SD_A","data_format":"1","description":"Date approximation for
ART_SD","variable_order":"4","code_text":"","variable_required___1":"0","variable_deprecated___1":"0","variable_
deprecated_d":"","has_codes":"1","code_format":"1","code_list":"< = Before this date | D = Exact to the date | M
= Exact to the month | Y = Exact to the year | > = After the date | U =
Unknown","code_file":"","codes_print":"0","variable_metadata_complete":"2"},{"record_id":"1","redcap_repeat_inst
rument":"variable_metadata","redcap_repeat_instance":5,"table_name":"","table_format":"","table_definition":"","
table_display_y":"","table_display_name":"","table_order":"","text_top":"","text_bottom":"","table_deprecated___
1":"","table_deprecated_d":"","table_metadata_complete":"","table_link":"1","variable_name":"ART_ED","data_forma
t":"4","description":"Date of stopping of
treatment","variable_order":"5","code_text":"","variable_required___1":"0","variable_deprecated___1":"0","variab
```

Data Model Browser



Website is auto-generated based on the content of the three REDCap projects.

iedeades.org

Why this matters

- One “source of truth” for the data model
- One place for users to visit
- Easy to add new variables and codes
- Machine-readable data model can inform additional software tools

- REDCap templates and module code are shared on GitHub
  - https://github.com/vanderbilt-redcap/data-model-browser
  - Currently developing an “External Module” version (one-click installation) to share via the REDCap Repository of External Modules

The Harmonist Project

(1) Support for Data Model
IeDEA Data Exchange Standard → Clear data standards

(2) Consortium Portal
Harmonist Hub → Managed research workflow

(3) Data Quality and Sharing Tool
Harmonist Data Toolkit → High quality data

The Harmonist Project

(1) Support for Data Model
IeDEA Data Exchange Standard → Clear data standards

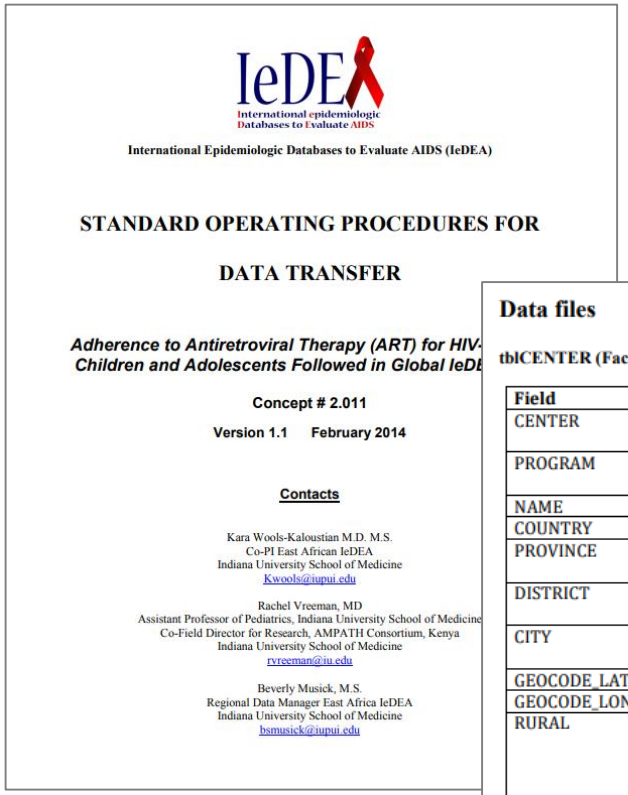
(2) Consortium Portal
Harmonist Hub → Managed research workflow

How can IeDEA investigators request data for projects?

(3) Data Quality
Harmonist Data Toolkit ata

Before…

- IeDEA investigators with an approved concept sheet developed a “Data Request” (Word document).
- Data request documents and due dates were circulated via email.
- A review of the first 4 data requests of 2016 (by different people) found that none aligned with the IeDEA DES.



International Epidemiologic Databases to Evaluate AIDS (IeDEA)

STANDARD OPERATING PROCEDURES FOR

DATA TRANSFER

Adherence to Antiretroviral Therapy (ART) for HIV-
Children and Adolescents Followed in Global IeDE

Concept # 2.011

Version 1.1 February 2014

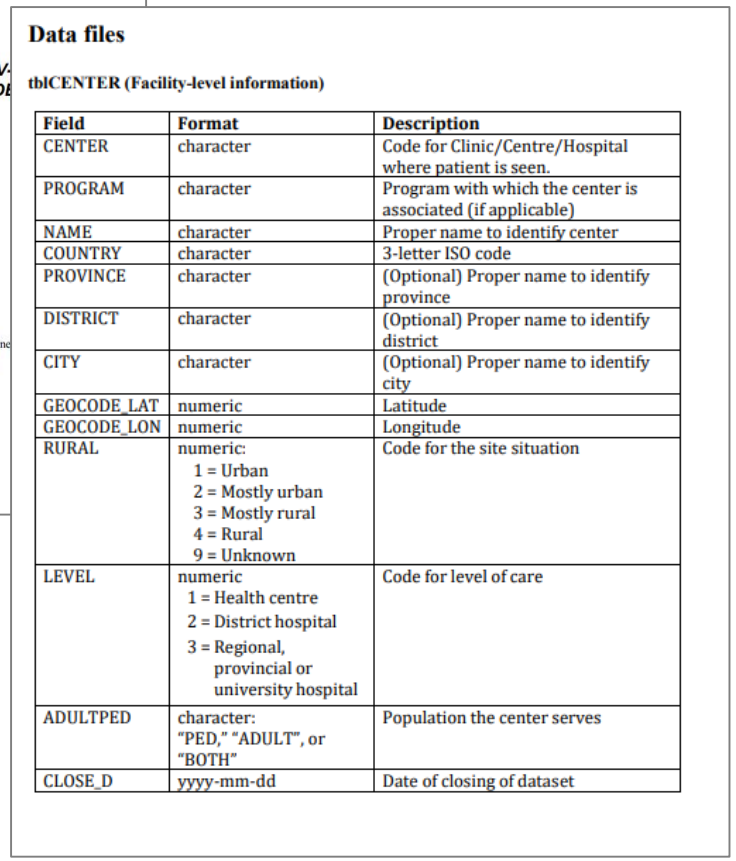
Contacts

Kara Wools-Kaloustian M.D. M.S.
Co-PI East African IeDEA
Indiana University School of Medicine
Kwools@iupui.edu

Rachel Vreeman, MD
Assistant Professor of Pediatrics, Indiana University School of Medicine
Co-Field Director for Research, AMPATH Consortium, Kenya
Indiana University School of Medicine
rvreeman@iu.edu

Beverly Musick, M.S.
Regional Data Manager East Africa IeDEA
Indiana University School of Medicine
bsmusick@iupui.edu

Original data request (created by hand!)



Data files

tblCENTER (Facility-level information)

Field	Format	Description
CENTER	character	Code for Clinic/Centre/Hospital where patient is seen.
PROGRAM	character	Program with which the center is associated (if applicable)
NAME	character	Proper name to identify center
COUNTRY	character	3-letter ISO code
PROVINCE	character	(Optional) Proper name to identify province
DISTRICT	character	(Optional) Proper name to identify district
CITY	character	(Optional) Proper name to identify city
GEOCODE_LAT	numeric	Latitude
GEOCODE_LON	numeric	Longitude
RURAL	numeric: 1 = Urban 2 = Mostly urban 3 = Mostly rural 4 = Rural 9 = Unknown	Code for the site situation
LEVEL	numeric 1 = Health centre 2 = District hospital 3 = Regional, provincial or university hospital	Code for level of care
ADULTPED	character: “PED,” “ADULT”, or “BOTH”	Population the center serves
CLOSE_D	yyyy-mm-dd	Date of closing of dataset

We did have a private consortium website, but

- It only indexed concept sheets, not data requests or due dates
- We couldn't tell which projects had received data and which ones were pending
- Nobody used it because we could never remember our passwords…

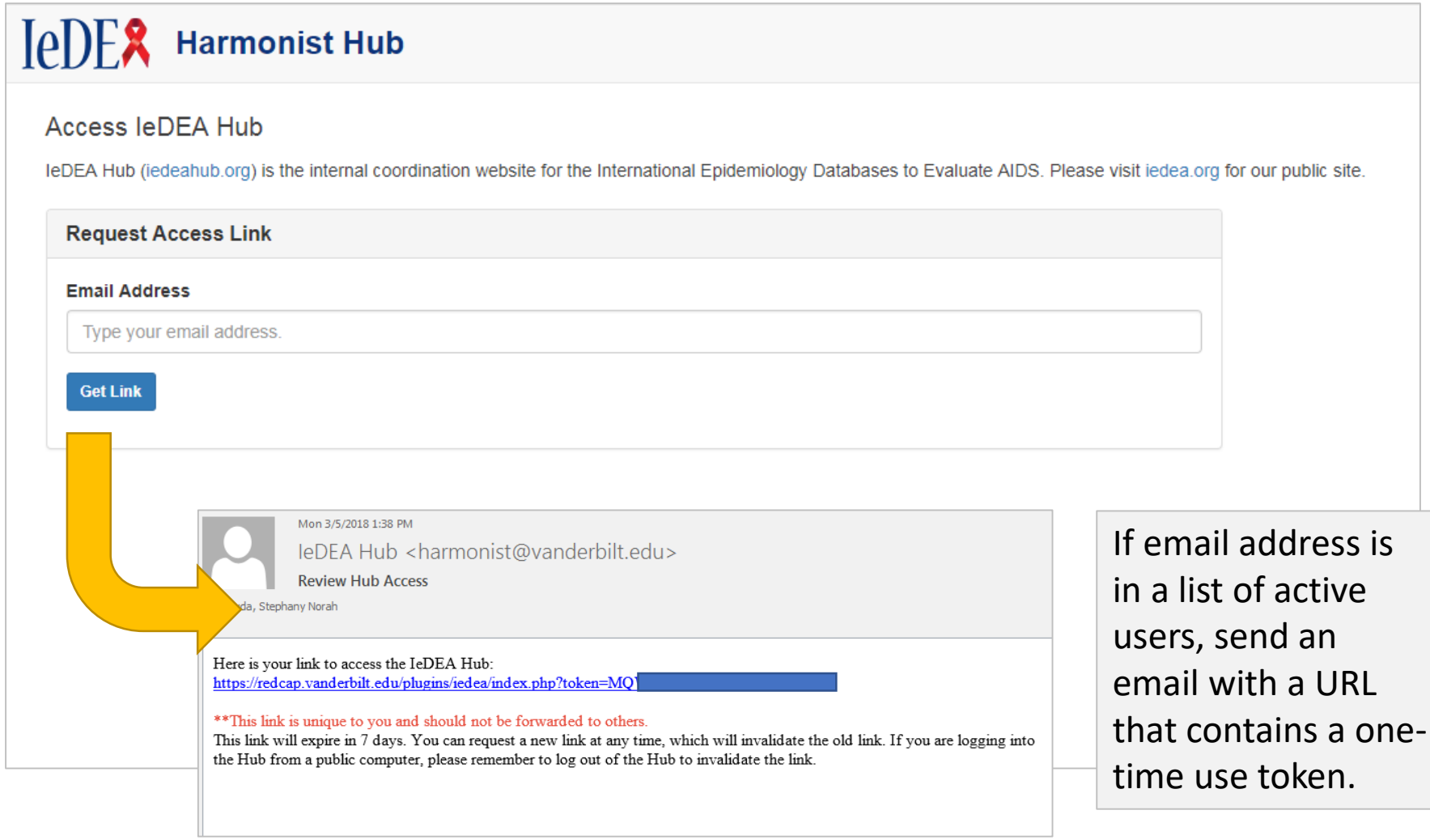
Designing the Harmonist Hub

We worked closely with IeDEA administrators, PIs, and investigators to design a web-based consortium management platform that could

- Index concept sheets and publications
- Coordinate Executive Committee review and voting for concept sheets, abstracts, manuscripts, etc.
- Implement a workflow with dashboards and automatic emails
- Generate metrics based on IeDEA's scientific portfolio

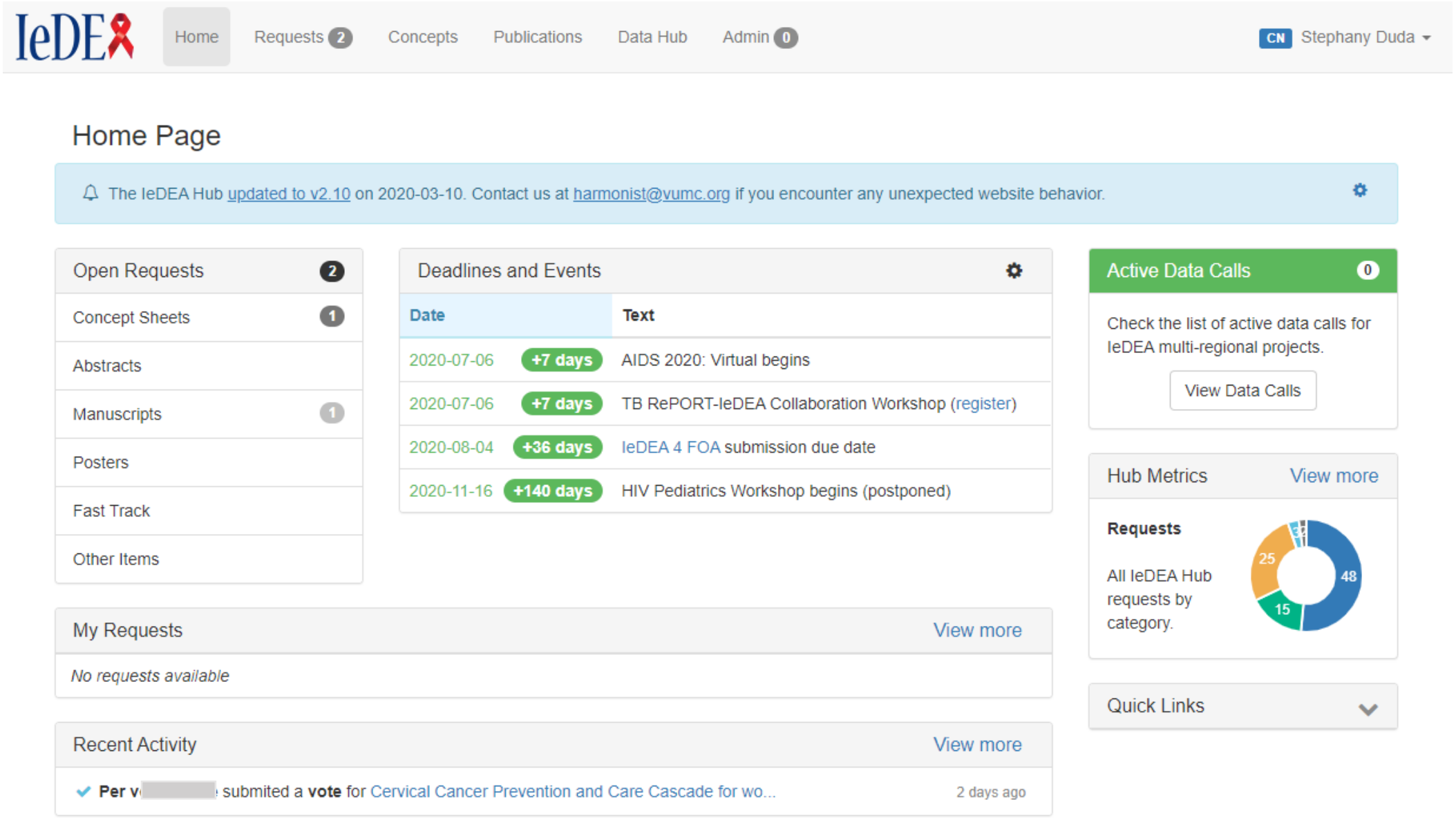
- Not a separate website, just another module installed on REDCap
  Research Electronic Data Capture

IeDEA Hub: Password-Free Login

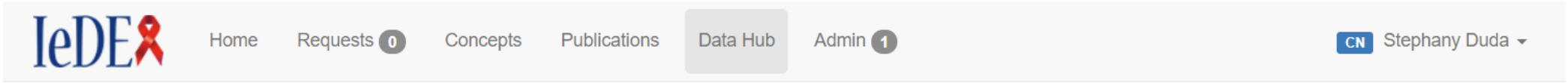


If email address is in a list of active users, send an email with a URL that contains a one-time use token.

IeDEA Hub: Home Page



IeDEA Hub – Data page



Data Hub

The IeDEA Hub provides a set of tools that allow you to **request, submit, and retrieve** IeDEA data. The purpose of these tools it to make it easier to use the IeDEA Data Exchange Standard (DES) and share standardized and quality-checked data in a secure way.

Explore the different types of IeDEA data

Coming Soon

Request IeDEA data for your approved concept

Create Data Request

Check and submit data for an active data call

View Data Calls 0

Retrieve data uploaded for your project

Download Data

Data News

View more

No data news to display

Recent Data Activity

View more

- **Beverly** uploaded EA data for MR180. — 7 hours ago
- **Karu** uploaded CN data for MR180. — 8 hours ago
- **Belinda** downloaded CN data for MR186. — 19 hours ago

Data Log 213 210

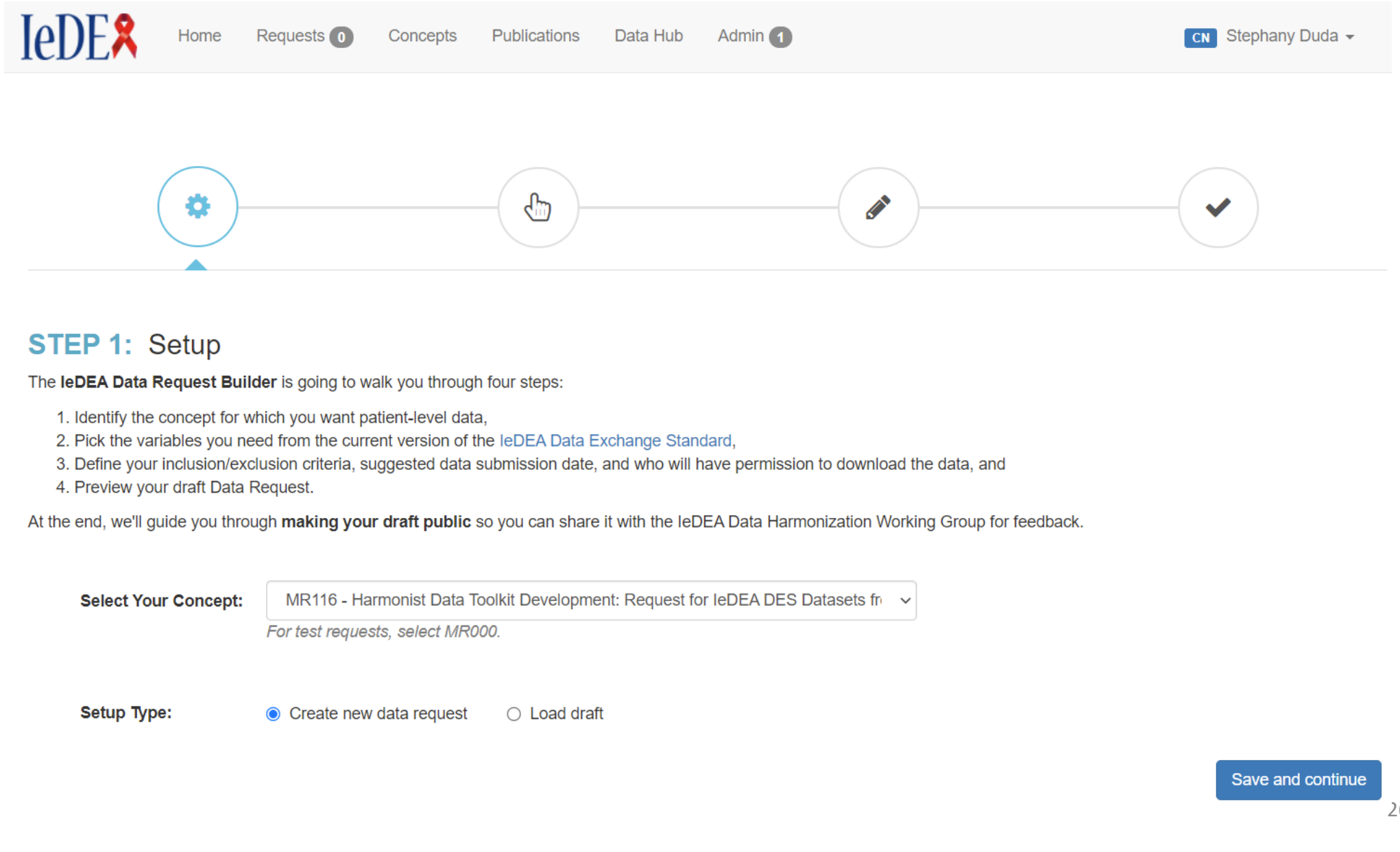
Track uploads and downloads of IeDEA patient-level datasets.

Harmonist Data Toolkit

Go to the Toolkit webpage (without a data submission request).

iedeades.org

Creating a Data Request: Step 1



Creating a Data Request: Step 2

STEP 2: Choose Variables

***MR116:** Data Request #63*

Select the IeDEA DES tables and variables you need for your analysis.

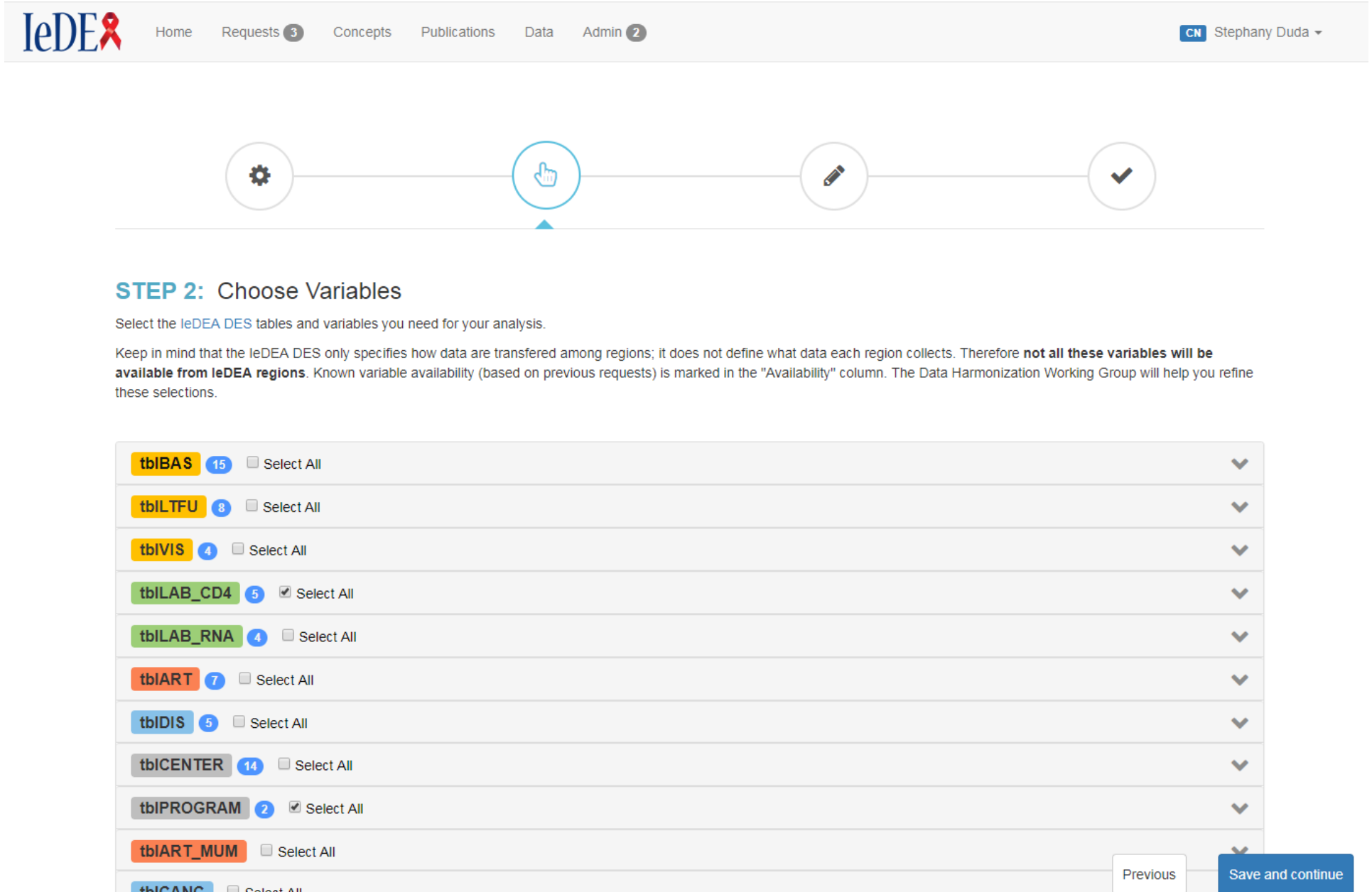
Keep in mind that the IeDEA DES only specifies how data are transfered among regions; it does not define what data each region collects. Therefore **not all these variables will be available from IeDEA regions**. Known variable availability (based on previous requests) is marked in the "Availability" column. The Data Harmonization Working Group will help you refine these selections.

tblBAS 2 ☐ Select All

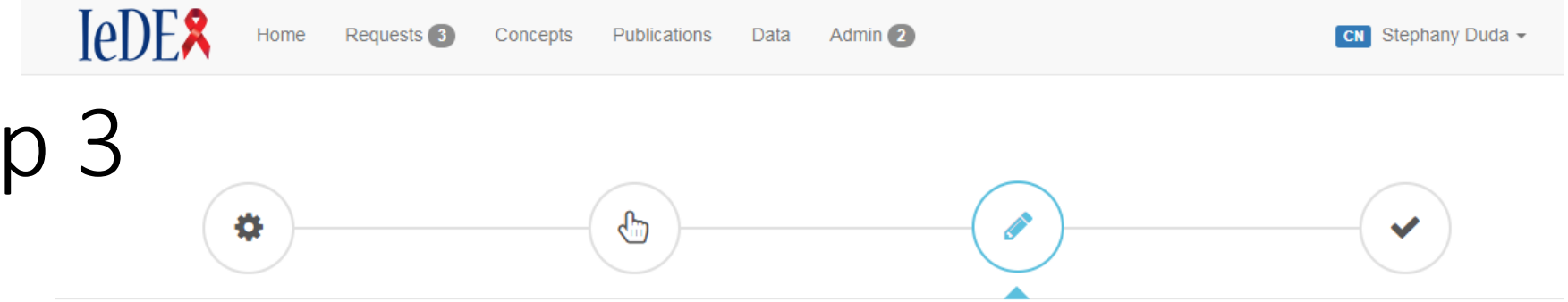
Select	*Field*	*Availability*	*Format*	*Description*
☑	PATIENT		Character or Numeric	Code to identify patient (Cohort Patient ID)
☑	PROGRAM		Character	Direct one-to-one link from the patient to their program
☐	BIRTH_D		YYYY-MM-DD	Birth date
☐	BIRTH_D_A		Character < = Before this date D = Exact to the date M = Exact to the month Y = Exact to the year > = After the date	Date approximation for BIRTH_D

Previous | Save and continue

Creating a Data Request: Step 2.5



Step 3



STEP 3: Add Details

The form below collects some extra information needed to generate the Data Request.

Text Descriptions

Inclusion criteria *(list variable names if possible)*

- Include all HIV-positive older adolescents and adults (>=16 years of age at enrollment) in the IeDEA cohorts from 1996 onward.
- Both ART-naïve and non-naïve patients at enrollment are eligible.
- Include all years of data from CENTERs that diagnose cryptococcal meningitis (CM) and provide data on CM/extrapulmonary cryptococcosis, even if the CENTER only started diagnosing CM recently. (e.g., If a clinic started in 2013, but CM diagnoses only began in 2015, include all data from 2013 onward.)
- Include both patients with CM and patients without CM from these CENTERs.

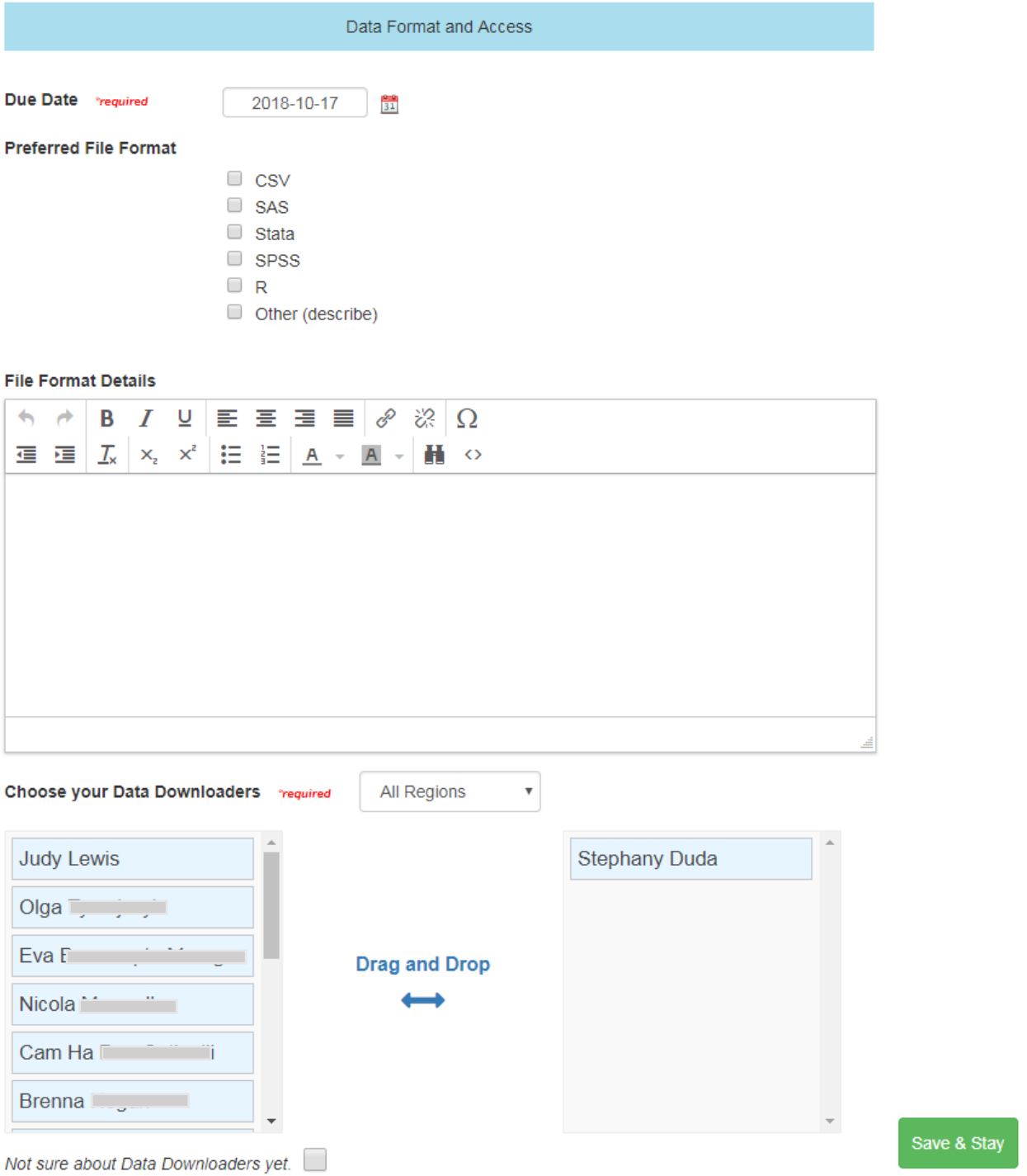
Exclusion criteria *(list variable names if possible)*

- Exclude CENTERs that do not systematically collect any opportunistic infection (OI) data.
- Exclude CENTERs that have never diagnosed CM.
- Exclude CENTERs that do not collect or provide data on CM for IeDEA (This means exclude CENTERs that do not ever collect or provide DIS_ID = 'CRCO', see definition of CM below.)

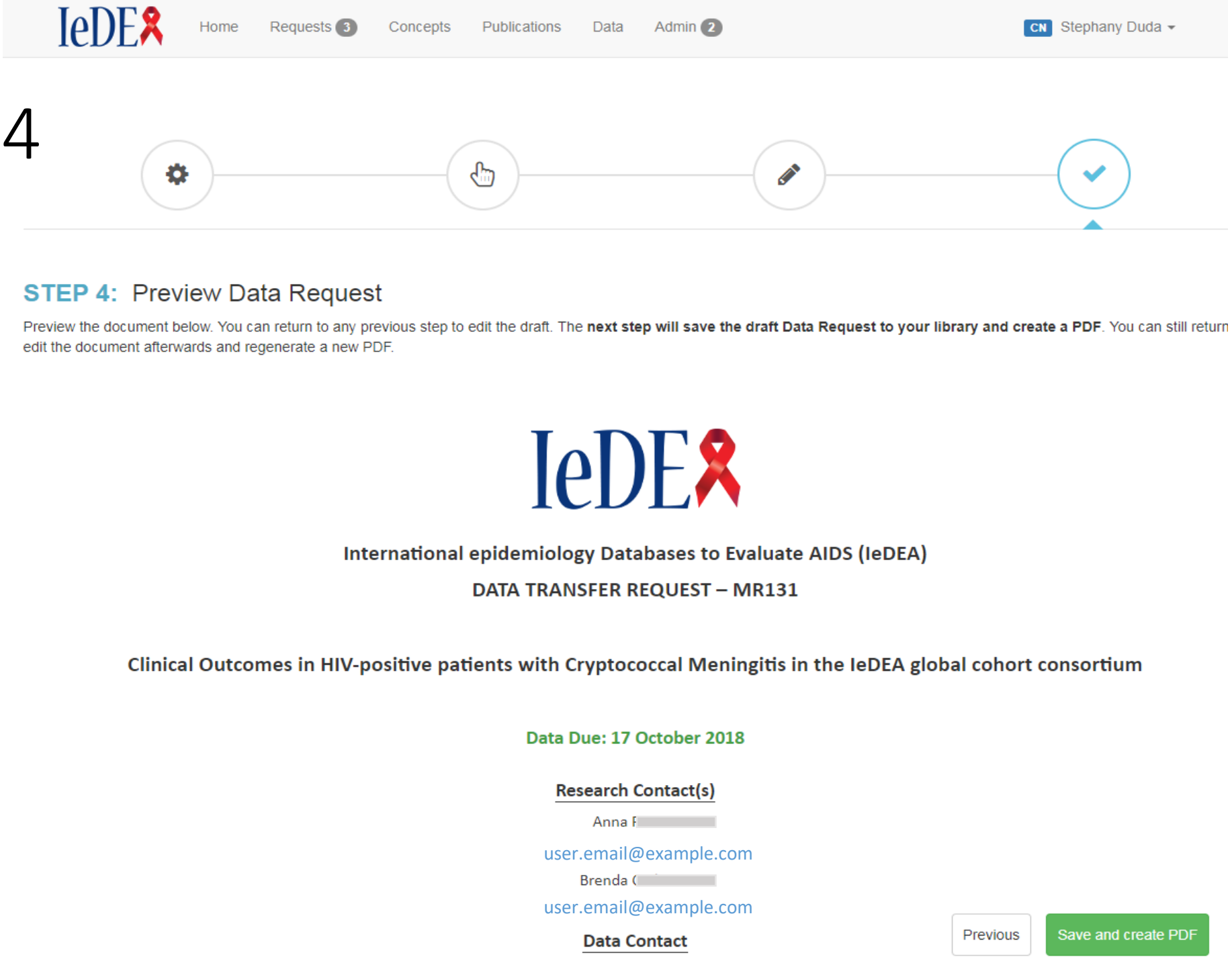
Save & Stay

Step 3.5

New content guided by the IeDEA Data Harmonization Working Group (not part of Word document Data Requests)



Step 4


IeDEA
Home
Requests 3
Concepts
Publications
Data
Admin 2
CN Stephany Duda
STEP 4: Preview Data Request
Preview the document below. You can return to any previous step to edit the draft. The next step will save the draft Data Request to your library and create a PDF. You can still return
edit the document afterwards and regenerate a new PDF.
IeDEA
International epidemiology Databases to Evaluate AIDS (IeDEA)
DATA TRANSFER REQUEST – MR131
Clinical Outcomes in HIV-positive patients with Cryptococcal Meningitis in the IeDEA global cohort consortium
Data Due: 17 October 2018
Research Contact(s)
Anna F
user.email@example.com
Brenda (
user.email@example.com
Data Contact
Previous
Save and create PDF

IeDEA Home Requests 3 Concepts Publications Data Admin 2 CN Stephany Duda

Steps Complete ✔

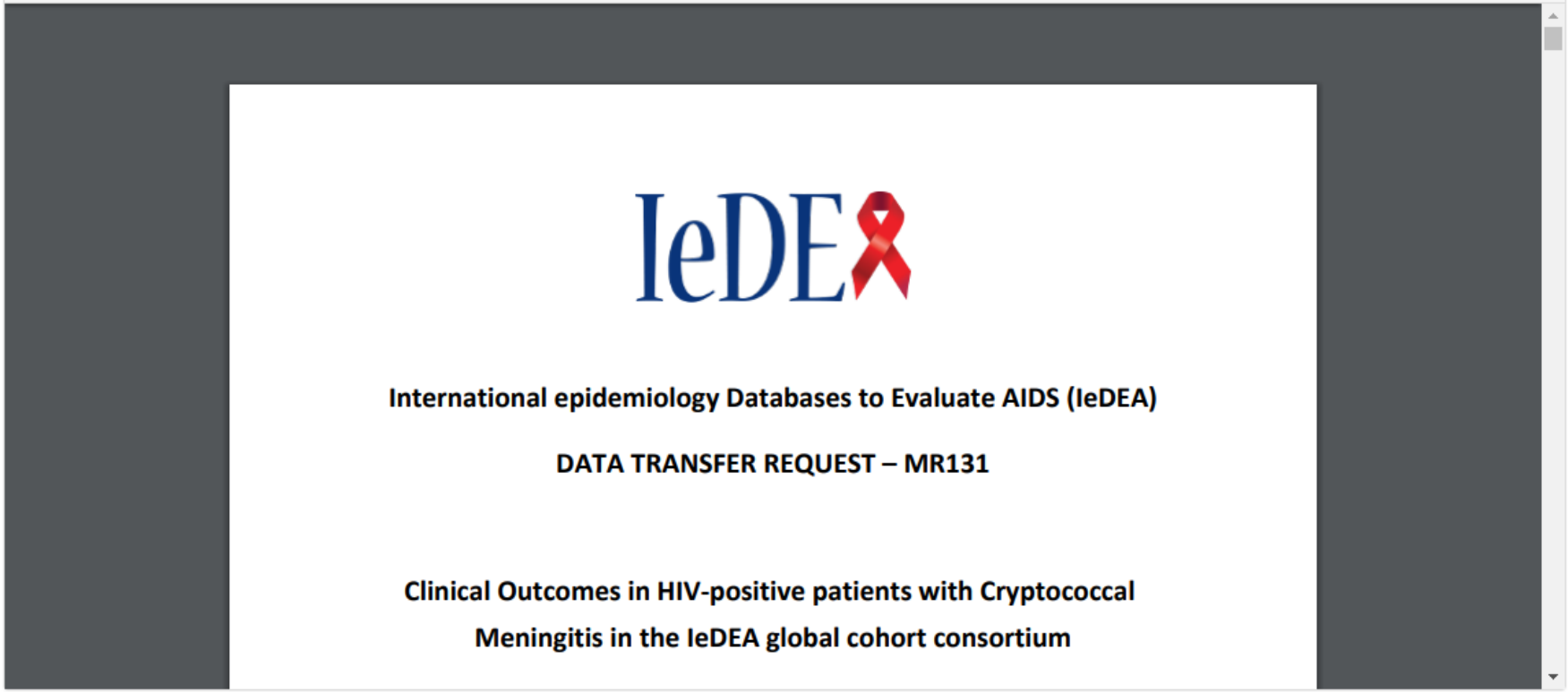
Your Data Request has been generated successfully. You can review and download the PDF below or download a ZIP file with an HTML and a PDF version. If you need to make changes, you can go back to edit your data request.

The **next step is to make your draft public and request a review** by the IeDEA Data Harmonization Working Group. The working group chairs may ask to meet with you to revise this document together, or may ask if you can discuss this data request on the next WG call.

View in My Drafts | Send for Review

< Back to Edit Data Request

Data Request — Download ZIP



IeDEA

International epidemiology Databases to Evaluate AIDS (IeDEA)

DATA TRANSFER REQUEST – MR131

Clinical Outcomes in HIV-positive patients with Cryptococcal Meningitis in the IeDEA global cohort consortium

Automatically Generated Data Request

International epidemiology Databases to Evaluate AIDS (IeDEA)

DATA TRANSFER REQUEST – MR131

Clinical Outcomes in HIV-positive patients with Cryptococcal Meningitis in the IeDEA global cohort consortium

Data Due: 17 October 2018

Research Contact(s)

Anna
user.email@example.com

Brenda
user.email@example.com

Data Contact

Stephany Duda
user.email@example.com

Data Request Version: 07 March 2019

tblLTFU - Death and dropout

Field	Format	Description
PATIENT	Character or Numeric	Code to identify patient (Cohort Patient ID)
DROP_Y	Numeric 0 = No 1 = Yes	Has the patient dropped out?
DROP_D	YYYY-MM-DD	If patient has dropped out, Date of last visit
DROP_D_A	Character D = Exact to the date M = Exact to the month Y = Exact to the year > = After the date U = Unknown	Date approximation for DROP_D
DROP_RS	Numeric	Reason for Drop
DEATH_Y	Numeric 0 = No 1 = Yes	Has the patient died?
DEATH_D	YYYY-MM-DD	Date of Death
DEATH_D_A	Character D = Exact to the date M = Exact to the month Y = Exact to the year > = After the date U = Unknown	Date approximation for DEATH_D

Why this matters

- No more manually assembled Word documents! Point-and-click data request builder saves time.
- The list of requestable variables is pulled from the machine-readable IeDEA DES. The tool automatically adds required variables if not selected
  - Therefore all data requests will be compliant with the standard.
- One “source of truth” for data requests (on the Hub)
- All resources in GitHub and the REDCapRepo

And…

- Data requests are stored in a machine-readable format now too.

< Back to Data

Check and Submit Data

IeDEA data is submitted securely through the **Harmonist Data Toolkit**. The Toolkit will

1. Scan your uploaded files to run data format and quality checks,
2. Auto-generate dataset reports for you to download, and
3. Allow data upload to the **secure Harmonist cloud** (for data transfer only).

Data files should be formatted according to the IeDEA Data Exchange Standard (IeDEA DES).

The following IeDEA Concepts have active Data Requests. Please review the request details or select the request for which you wish to upload data.

Not started | Partial Data | Complete Data | Data Not Available | Region Not Requested | Other Status

View Data Activity Log | View Past Data Calls

Open Data Calls

No Open Data Calls available

Completed Data Calls

Due Date	Data Request Details	Data Contact	AP	CA	CN	EA	NA	SA	WA	Actions
2021-09-08 -23 days	**MR180** Infectious and non-infectious causes of liver disease among adults living with HIV from low and middle income countries, the Sentinel Research Network of IeDEA *Draft ID: 54* Data Request | MR180 Concept |	Kare (WA)	✓	✓	✓	✓	✗	✓	✓	Upload Data Change Status
2021-09-15 -16 days	**MR181** Diagnostic value of serological markers for liver steatosis and fibrosis in people living with HIV: a multicenter study from the IeDEA Network *Draft ID: 55* Data Request | MR181 Concept |	H (CN)	✓	✓	✓	✓	✗	✓	✓	Upload Data Change Status
2021-09-20 -11 days	**MR185** Associations of mental health and substance use disorders with virological failure by sex among aging adults living with HIV from low/middle-income countries, the Sentinel Research Network of IeDEA *Draft ID: 60* Data Request | MR185 Concept |	Kath CA)	✓	✓	✓	✓	✗	✓	✓	Upload Data Change Status

The Harmonist Project

(1) Support for Data Model
IeDEA Data Exchange Standard → Clear data standards

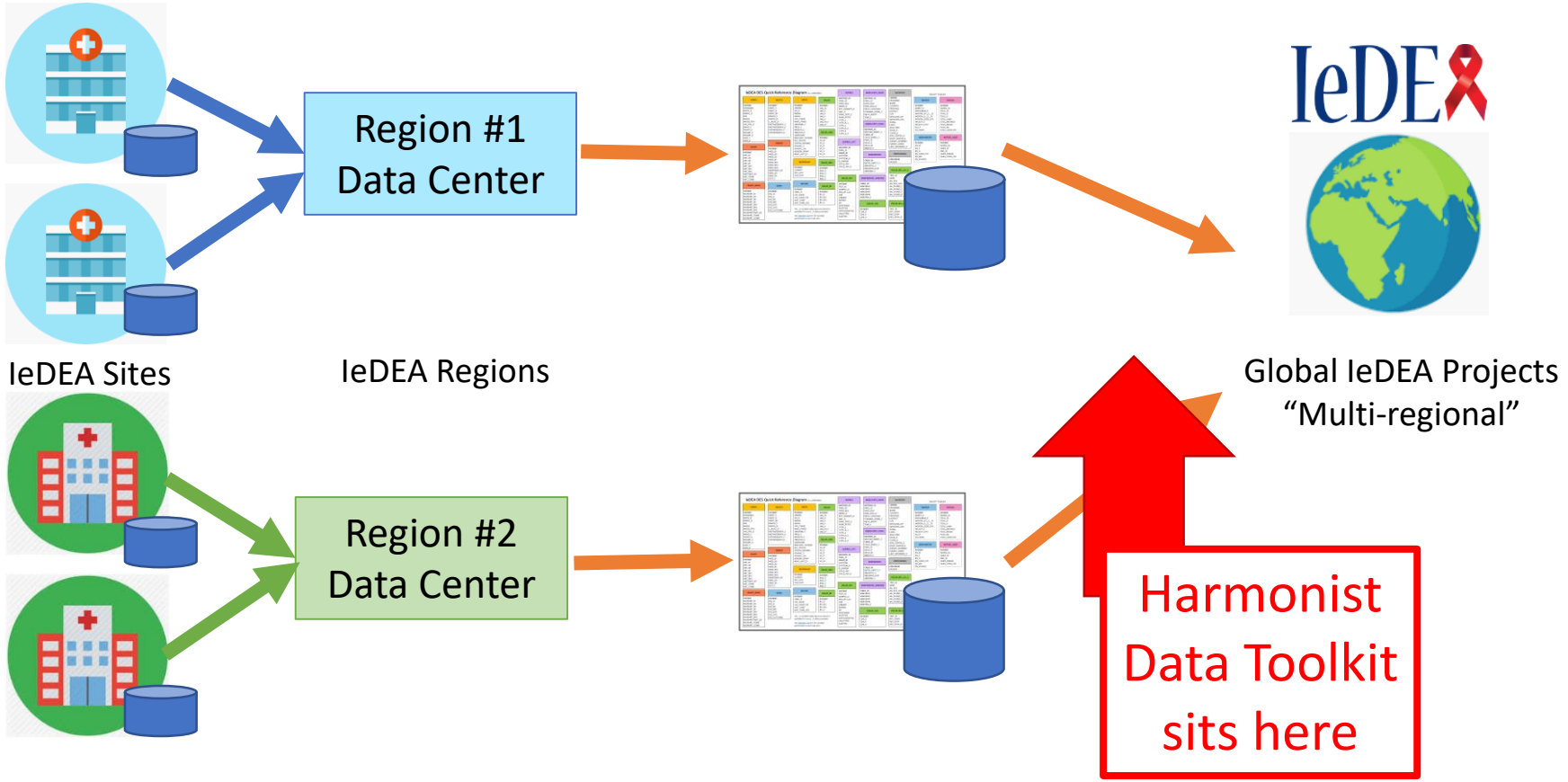
(2) Consortium Portal
Harmonist Hub → Managed research workflow

(3) Data Quality and Sharing Tool
Harmonist Data Toolkit → High quality data

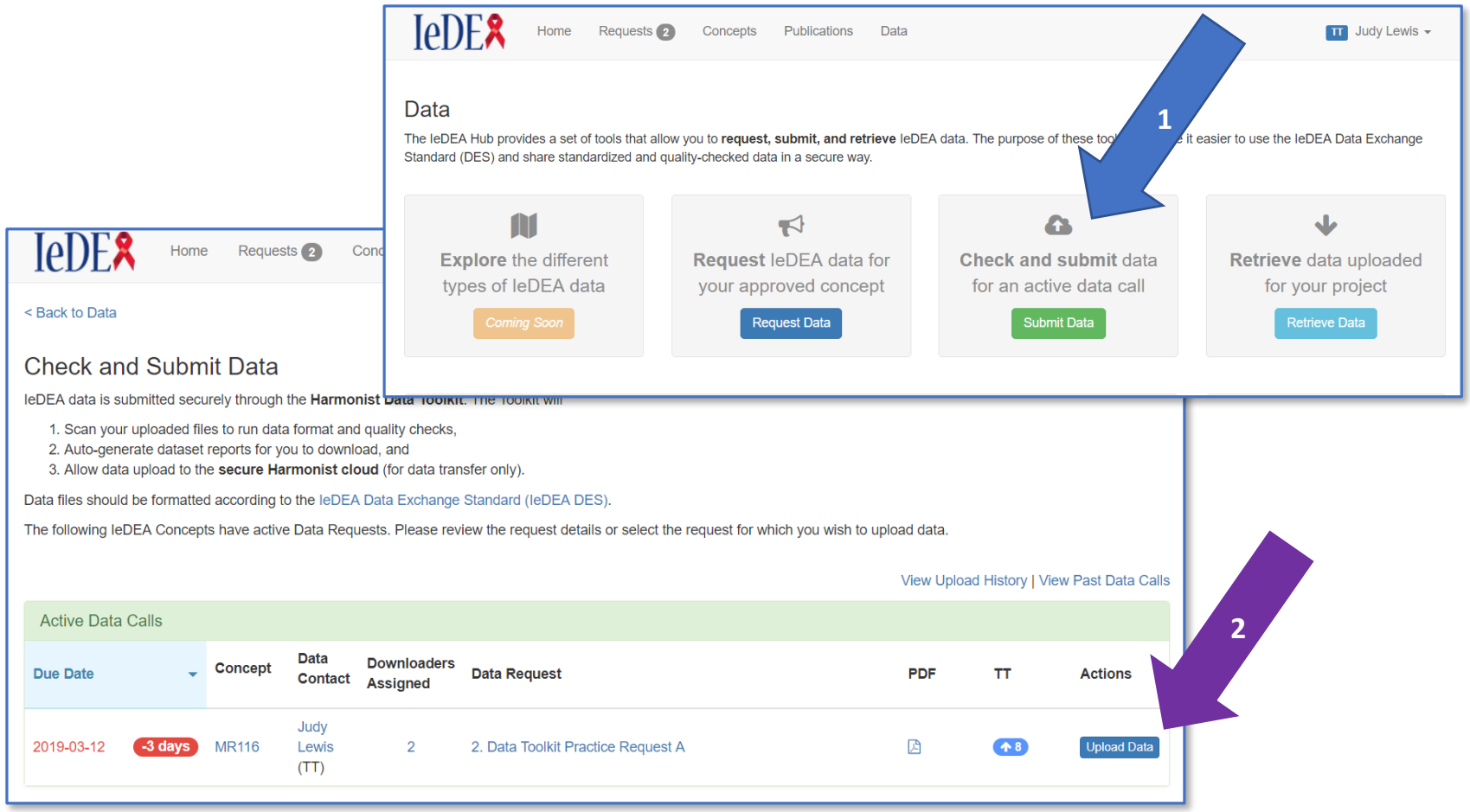
IeDEA Harmonist Data Toolkit

- Collaborative project with all seven IeDEA regions
- Web application
- Developed with freely available/open source tools (R, Shiny, REDCap)
- Designed to evolve with data exchange standard
- Features:
  - Ensures datasets conform to common data model
  - Performs data quality checks
  - Generates reproducible reports
  - Submits approved datasets to secure cloud storage

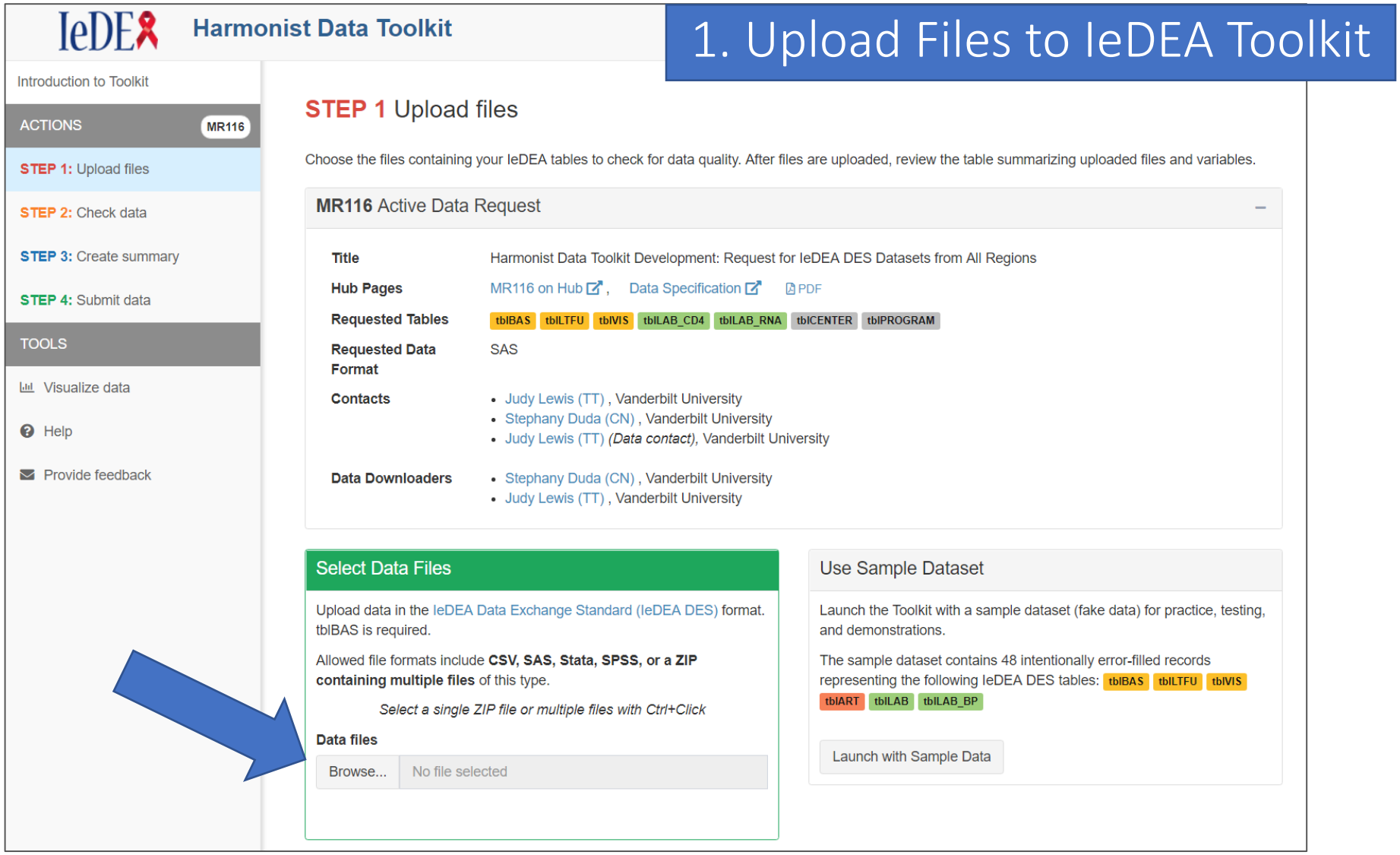
Flow of IeDEA Data for Global Projects



Workflow Begins in IeDEA Project Portal (IeDEA Hub)



1. Upload Files to IeDEA Toolkit



2. File Review/Data Quality Checks

IeDEA **Harmonist Data Toolkit**

Introduction to Toolkit

ACTIONS MR116

STEP 1: Upload files

STEP 2: Check data

STEP 3: Create summary

STEP 4: Submit data

TOOLS

Visualize data

Help

Provide feedback

Exit Data Toolkit

STEP 1 Upload files

Choose the files containing your IeDEA tables to check for data quality. After files are uploaded, review the table summarizing uploaded files and variables.

MR116 Active Data Request +

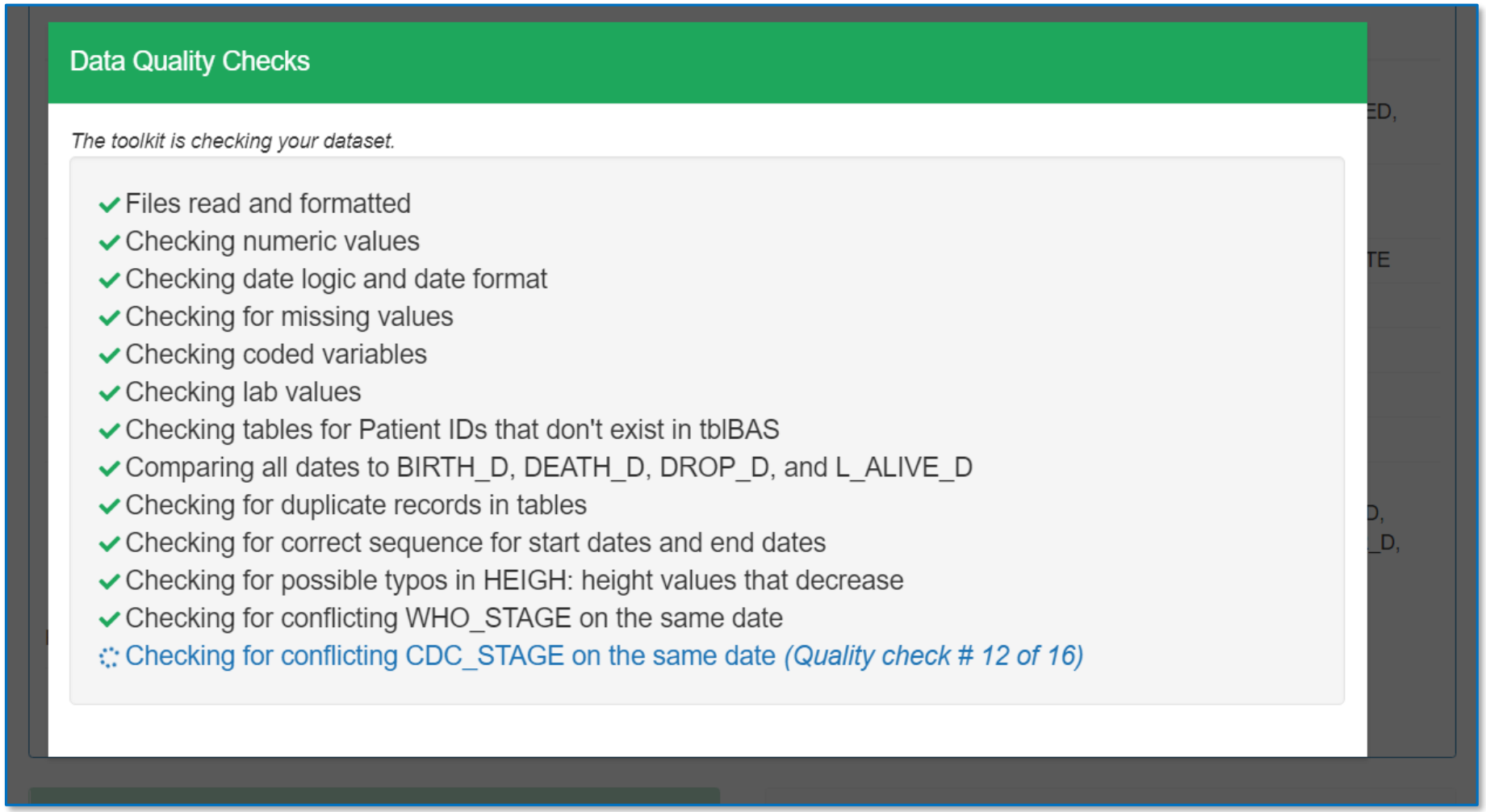
Missing Variables

The following variables requested by MR116 were not found:

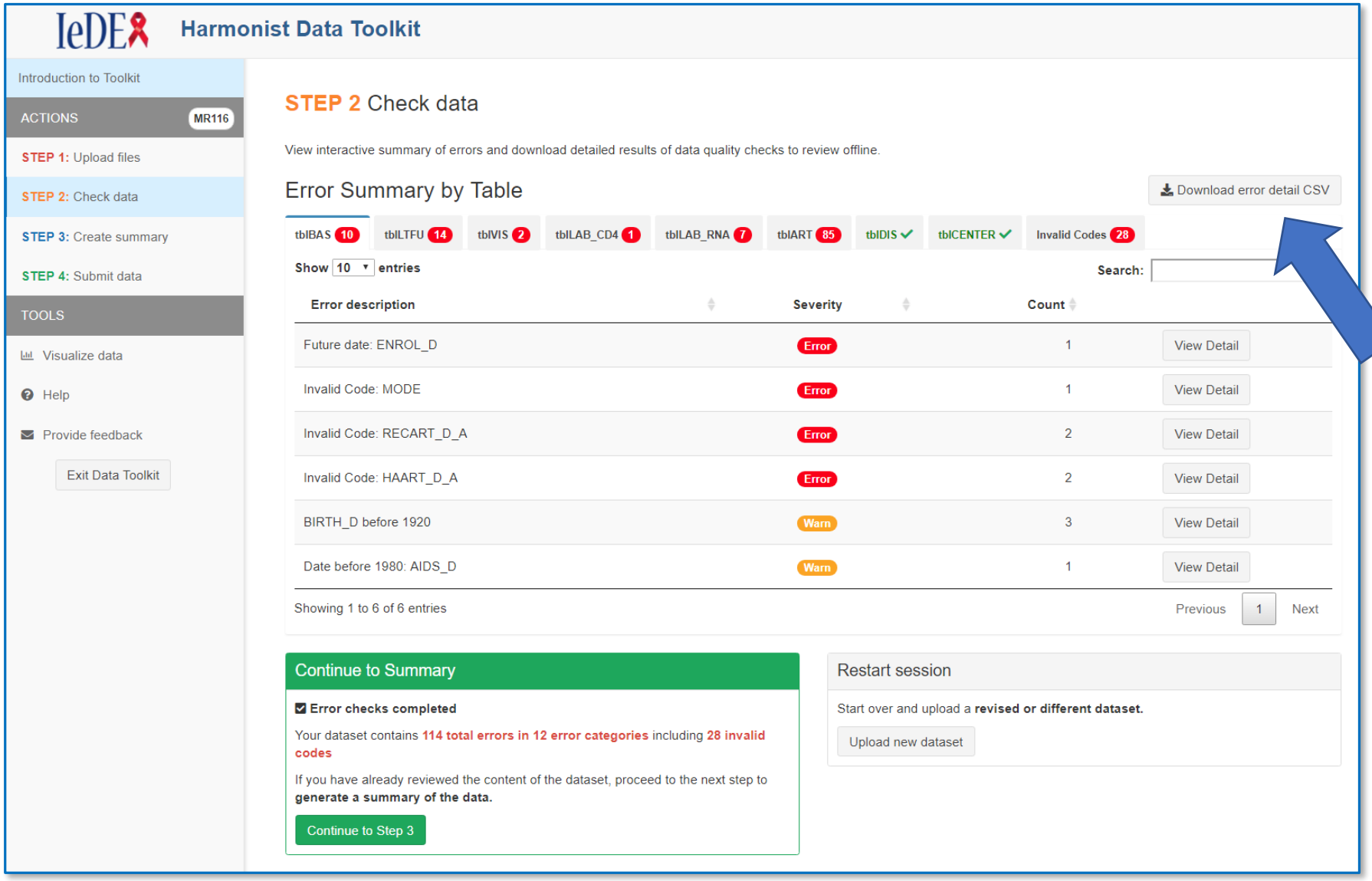
- tblVIS NEXT_VISIT_D, NEXT_VISIT_D_A
- tblPROGRAM Table missing (2 variables)

Summary of Uploaded IeDEA Tables −

Table	Records	IeDEA DES Variables	Extra Variables
tblBAS	28089	PATIENT, PROGRAM, BIRTH_D, BIRTH_D_A, ENROL_D, ENROL_D_A, GENDER, MODE, NAIVE_Y, PROPH_Y, RECART_Y, RECART_D, RECART_D_A, HAART_D, HAART_D_A, AIDS_Y, AIDS_D, AIDS_D_A	CENTER, YEAR_ENROLLED, SITE
tblLTFU	28089	PATIENT, DROP_Y, DROP_D, DROP_D_A, DROP_RS, DEATH_Y, DEATH_D, DEATH_D_A, L_ALIVE_D, L_ALIVE_D_A	CENTER, SITE
tblVIS	1084237	PATIENT, CENTER, VIS_D, VIS_D_A, CDC_STAGE, WHO_STAGE, PREG_Y	GENDER_ID, SITE
tblLAB_CD4	298041	PATIENT, CD4_D, CD4_D_A, CD4_V, CD4_U	CENTER, SITE
tblLAB_RNA	159234	PATIENT, RNA_D, RNA_D_A, RNA_V	CENTER, SITE
tblART	140435	PATIENT, ART_ID, ART_SD, ART_SD_A, ART_ED, ART_ED_A	CENTER, SITE
tblDIS	3750	PATIENT, DIS_ID, DIS_D, DIS_D_A, DIS_ED, DIS_ED_A, DIS_OUTCOME	CENTER, SITE
tblCENTER	9	CENTER, PROGRAM, NAME, COUNTRY, PROVINCE, DISTRICT, CITY, GEOCODE_LAT, GEOCODE_LON, RURAL, LEVEL, ADULTPED, OPEN_D, CLOSE_D	REGION, ADD_CENTER_D, DROP_CENTER_D, SITE


Data Quality Checks
The toolkit is checking your dataset.
Files read and formatted
Checking numeric values
Checking date logic and date format
Checking for missing values
Checking coded variables
Checking lab values
Checking tables for Patient IDs that don't exist in tblBAS
Comparing all dates to BIRTH_D, DEATH_D, DROP_D, and L_ALIVE_D
Checking for duplicate records in tables
Checking for correct sequence for start dates and end dates
Checking for possible typos in HEIGH: height values that decrease
Checking for conflicting WHO_STAGE on the same date
Checking for conflicting CDC_STAGE on the same date (Quality check # 12 of 16)

3. Data Quality Results



4. Reproducible Reports

STEP 3 Create summary

Generate and download customized reports summarizing uploaded dataset.

Customize Summary Report

File format for report

PDF

Data subgroup(s) for report

All

(Optional) Short title for report heading

Select report content

- ☑ Summary statistics of tables
- ☑ Summary of data quality checks
- ☑ Histograms of dates

Date Histogram Options

Choose years to include in histograms

Years: 2000 - present

Generate summary PDF report

IeDEA Harmonist Data Toolkit Report

Dataset submitted from IeDEA Region: Harmonist Test

Report date: 2019-03-07

Dataset Summary

Total number of patients in dataset: 28089

Table 1: Table Summary

Table	Records	Patients	Age at Enrollment					
			0-4	5-9	10-14	15-19	20-24	Adults 25+
tblBAS	28089	28089	0	0	0	971	3342	23776
tblLTFU	28089	28089	0	0	0	971	3342	23776
tblVIS	1084237	26820	0	0	0	963	3270	22587
tblLAB_CD4	298041	27304	0	0	0	945	3273	23086
tblLAB_RNA	159234	15524	0	0	0	471	2010	13043
tblART	140435	25840	0	0	0	889	2971	21980
tblDIS	3750	3215	0	0	0	100	345	2770

Table 2: SITE in Dataset

SITE	Patients	tblLTFU	tblVIS	tblLAB_CD4	tblLAB_RNA	tblART	tblDIS
Hogwarts	594	594	537	557	461	536	36
Hufflepuff	11763	11763	11744	11447	239	11530	1933
Muggleton	5248	5248	5240	5145	4898	4397	650
Potterburg	3208	3208	3208	3168	3116	2723	69
Ravenclaw	1079						
Slytherin	458						
Snapetown	4099						
Wizardville	1640						

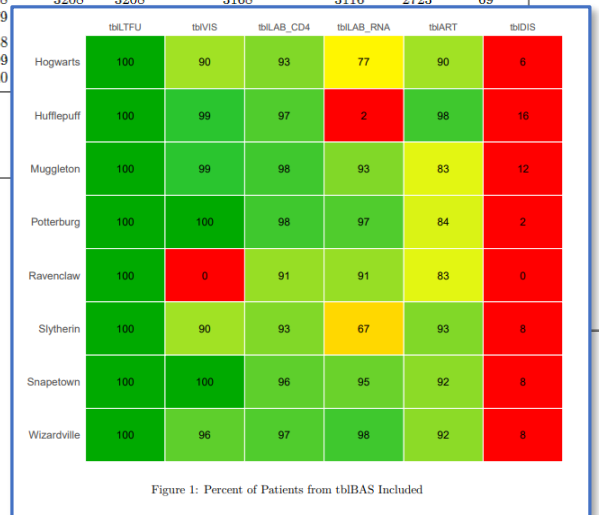


Figure 1: Percent of Patients from tblBAS Included

Table 3: Number of observations per year

Variable	< 2011	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Enrolled	14728	2517	2527	2859	2391	1439	1332	295	0	0	28088
Visits	411772	115301	117341	124501	118630	120098	71094	5500	0	0	1084237
Deaths	1064	249	232	234	170	216	141	10	0	0	2316
Transfers Out	193	19	19	48	94	145	139	14	0	0	671
Viral Load	59549	15201	15740	16776	16726	17731	15799	1708	0	0	159230
CD4	122521	30414	30781	30465	33865	26465	20152	3378	0	0	298041

Histograms of important dates by SITE

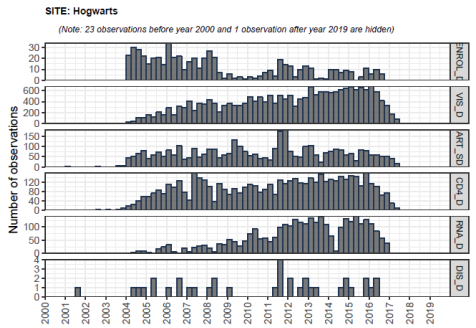
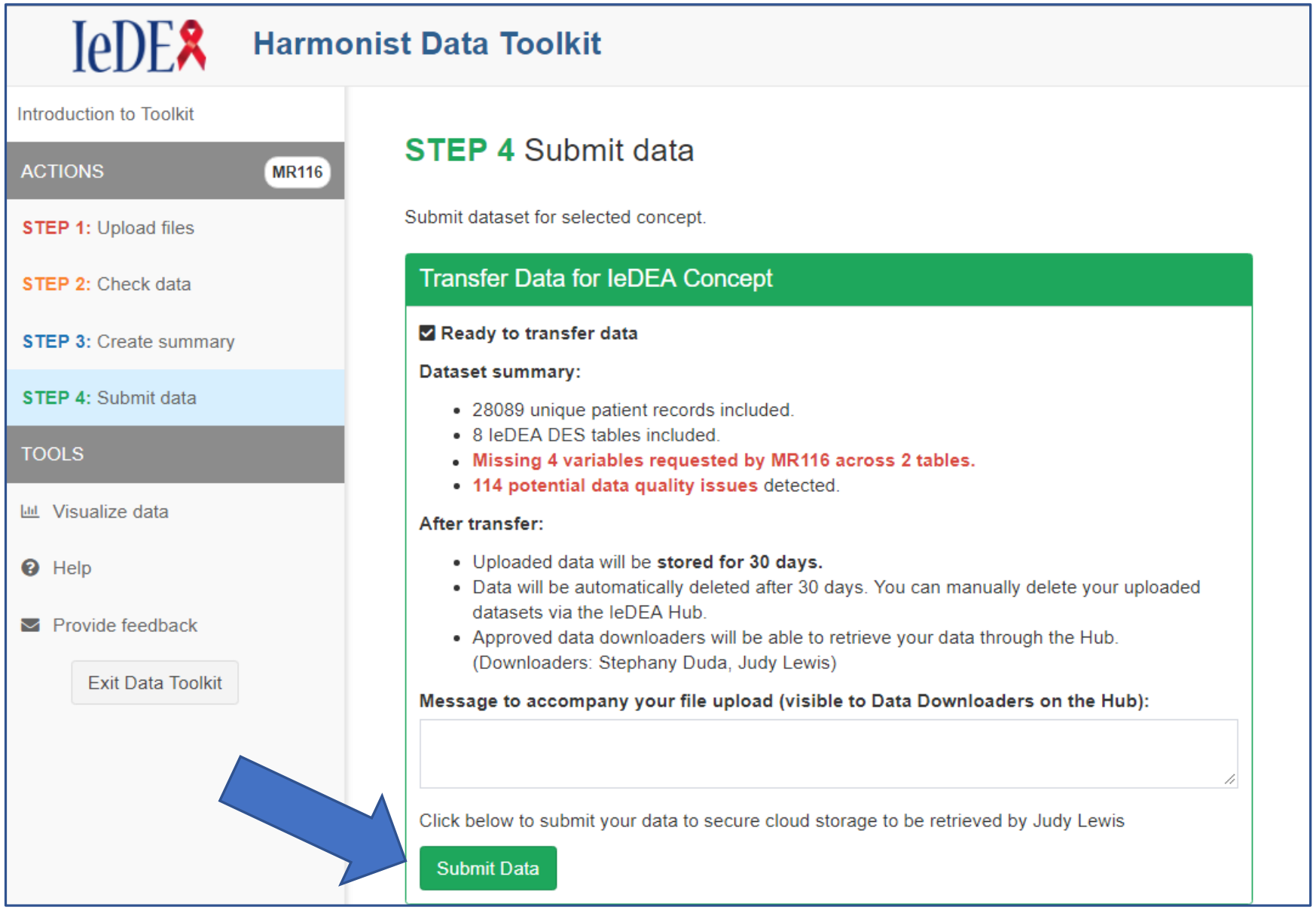


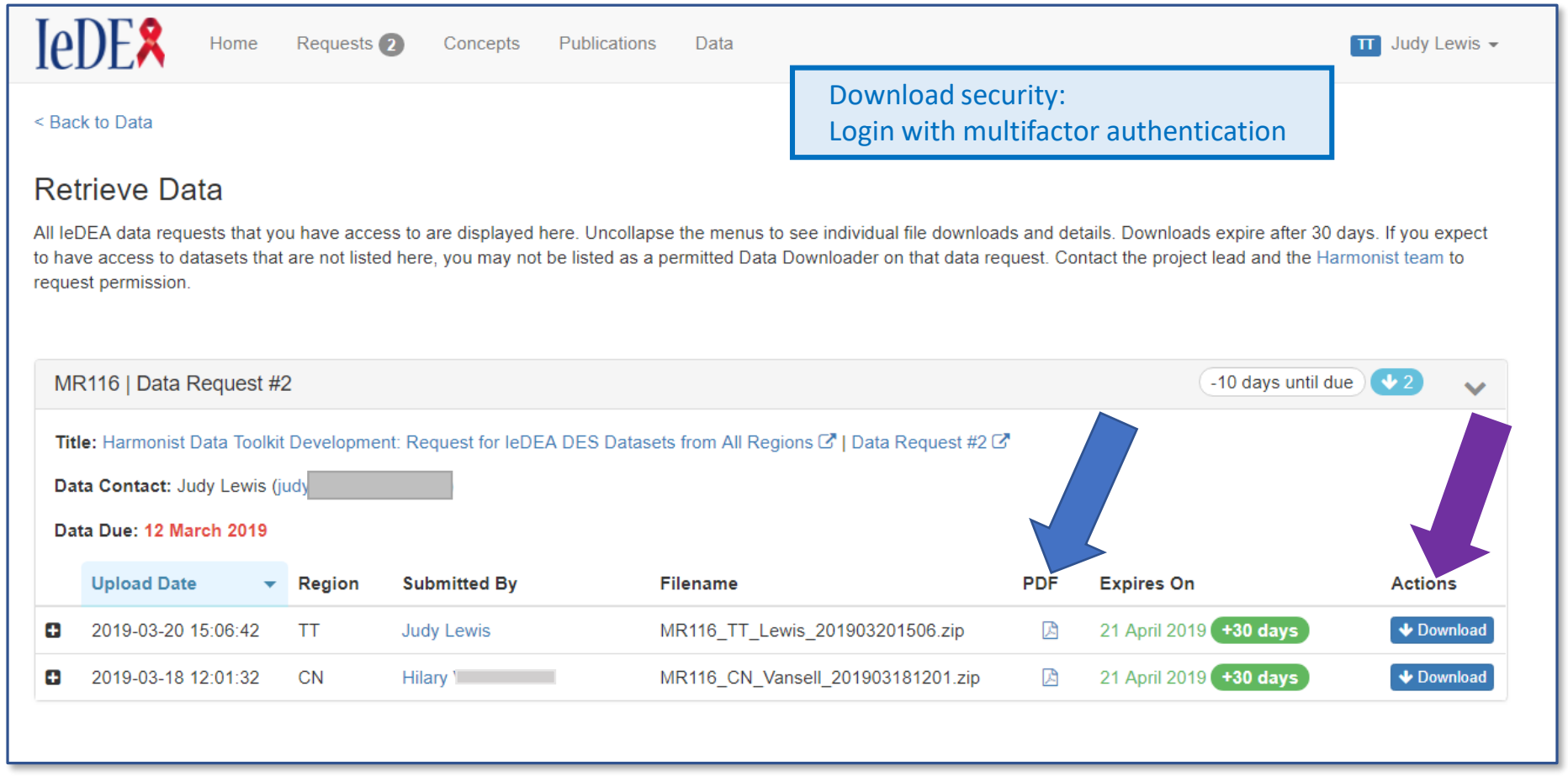
Table 4: Summary statistics from uploaded tables

Value	Count	Percent
Sex		
Male	17453	62.1
Female	10636	37.9
Missing	0	0.0
Deceased		
No	25773	91.8
Yes	2316	8.2
Missing	0	0.0
Treatment Naive at Enrollment		
No	2940	10.5
Yes	25149	89.5
Missing	0	0.0
Receive Antiretroviral Therapy (ART)		
Yes	25840	92.0
Missing	2249	8.0

5. Transfer Data



6. Recipient Downloads Data

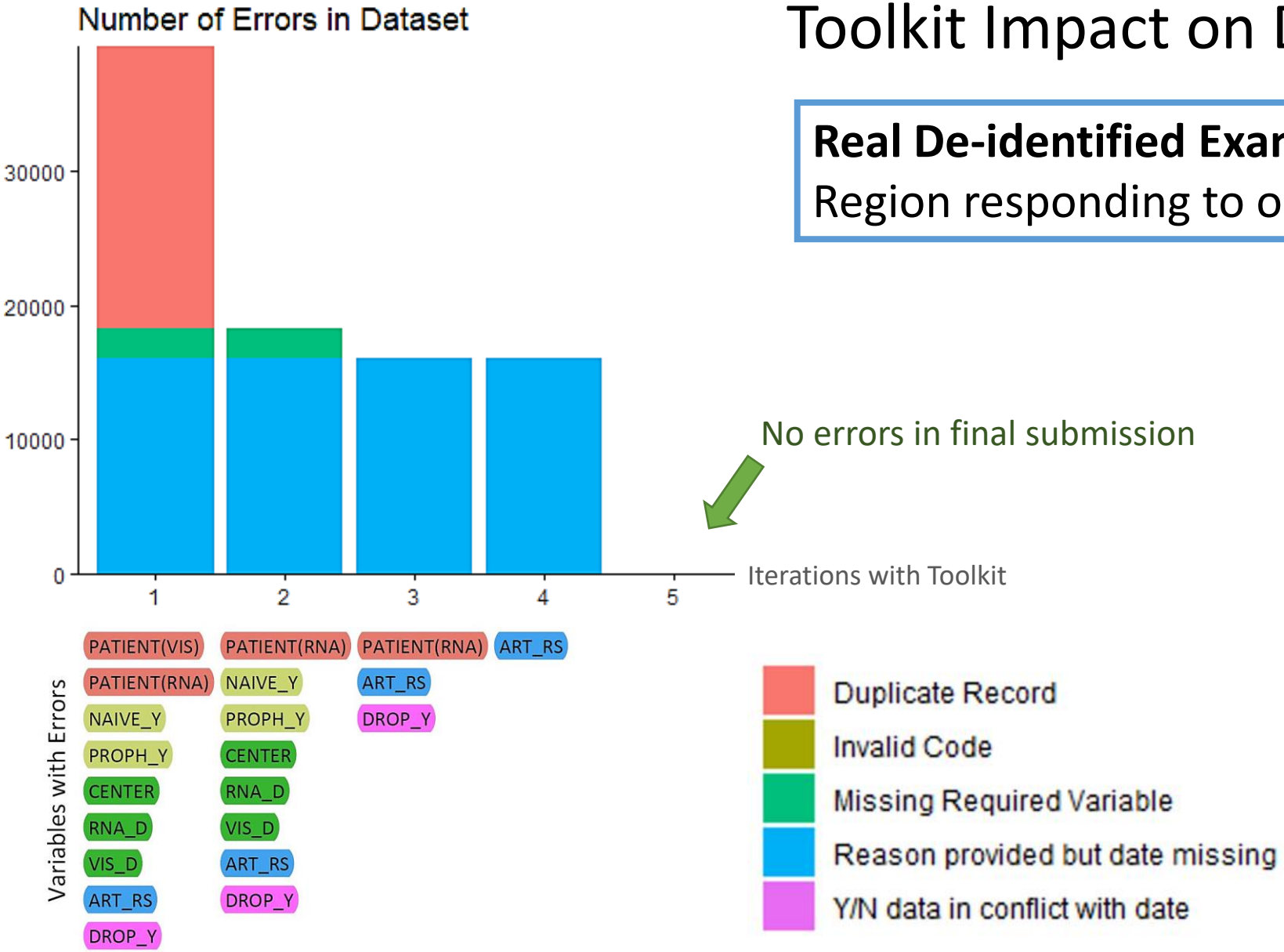


Toolkit Impact on Data Quality

- The Toolkit was released in March 2019
  - > 800 datasets have been checked and summarized
  - 1,800 to 986,089 patients per dataset
  - 81 secure file transfers for 17 multiregional IeDEA projects
- Regional data managers uploaded datasets and reviewed data quality results multiple times before final transfer
- Data managers used Toolkit data quality reports to improve datasets before transfer
- The number and types of errors decreased with each iteration of Toolkit use, average of 61% decrease in errors

Toolkit Impact on Data Quality

Real De-identified Example:
Region responding to one data request



Why This Matters

- High quality data is essential to meaningful research.
- Tools like this can help:
  - Improve adherence to data model and standards
  - Reduce time for data preparation and checking
  - Highlight data completeness and coding problems
  - Increase security and uniform workflow for data exchange
- Generalized design allows software to be adapted to other domains.

Lessons Learned

- Close collaboration with stakeholders and users is key
  - Frequent contact with stakeholders (e.g., admins, data managers, working groups)
  - Structured testing and training exercises with users
  - In-person (now virtual) meetings to collaborate on design
- Using REDCap as a platform facilitates development and sharing
- Defining details in REDCap and using the REDCap API make it possible to design tools that adapt with the changing data model

Tools have been well received by IeDEA; multiregional concept proposals and data requests have doubled.

Development & Opportunities

- Future tasks
  - Expand features
  - Gather insights from other consortia
  - Improve generalizability of tools

- We welcome all questions and input!
  - Contact us at harmonist@vumc.org
- Code available

  - github.com/IeDEA/Harmonist
  - github.com/vanderbilt-redcap
  - redcap.vanderbilt.edu/consortium/modules/index.php

Thank You

- Harmonist and REDCap technical teams
- IeDEA leadership, Data Harmonization Working Group, and collaborators
- Other consortia helping to make our processes generalizable!

Research reported in this publication was supported by the National Institute Of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R24AI124872. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.