

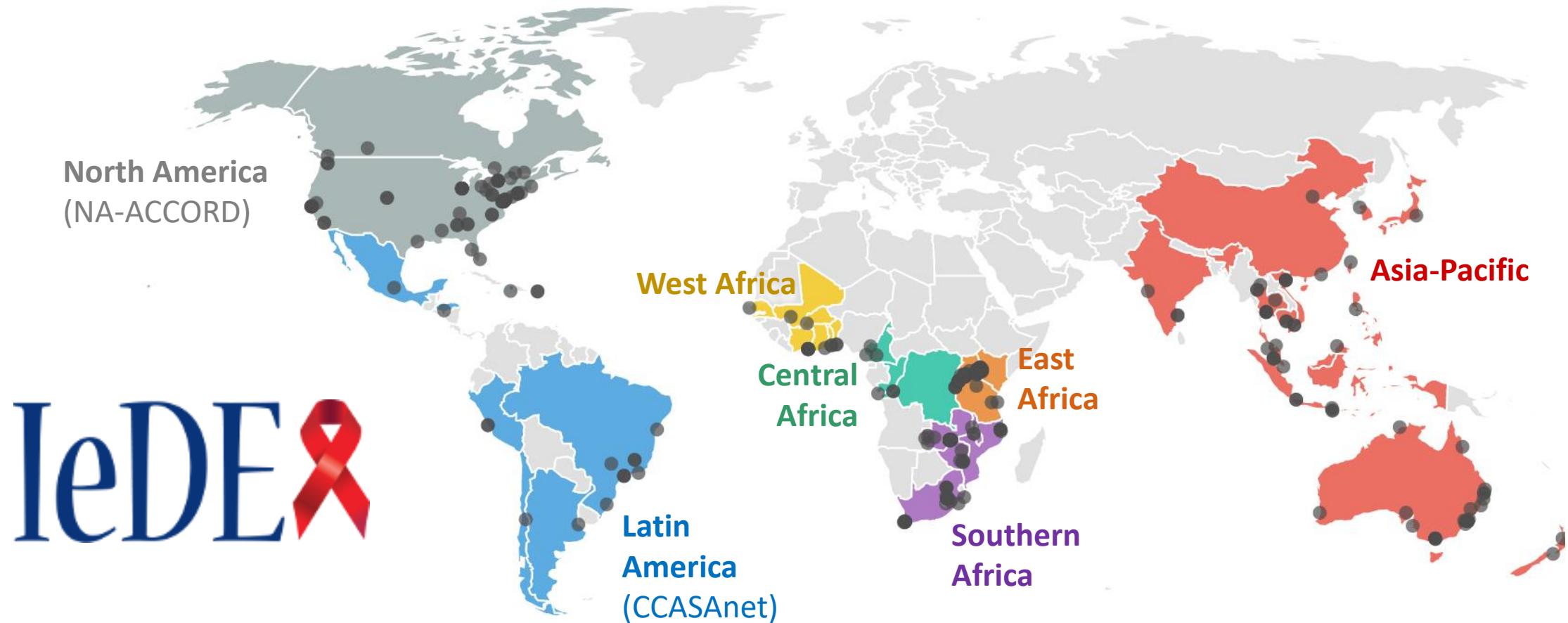
# Streamlining Data Sharing in a Global HIV Research Consortium

Stephany Duda PhD and Judy Lewis PhD

Vanderbilt University Medical Center

October 1, 2021

# IeDEA – A global HIV research consortium

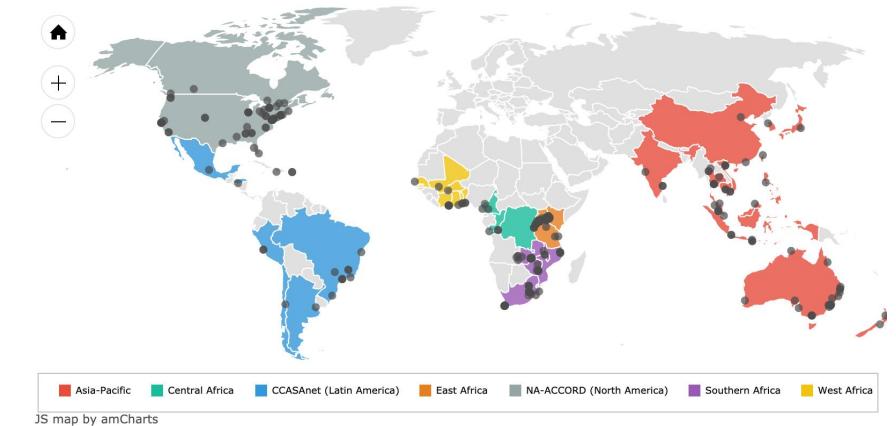


## **International epidemiology Databases to Evaluate AIDS**

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

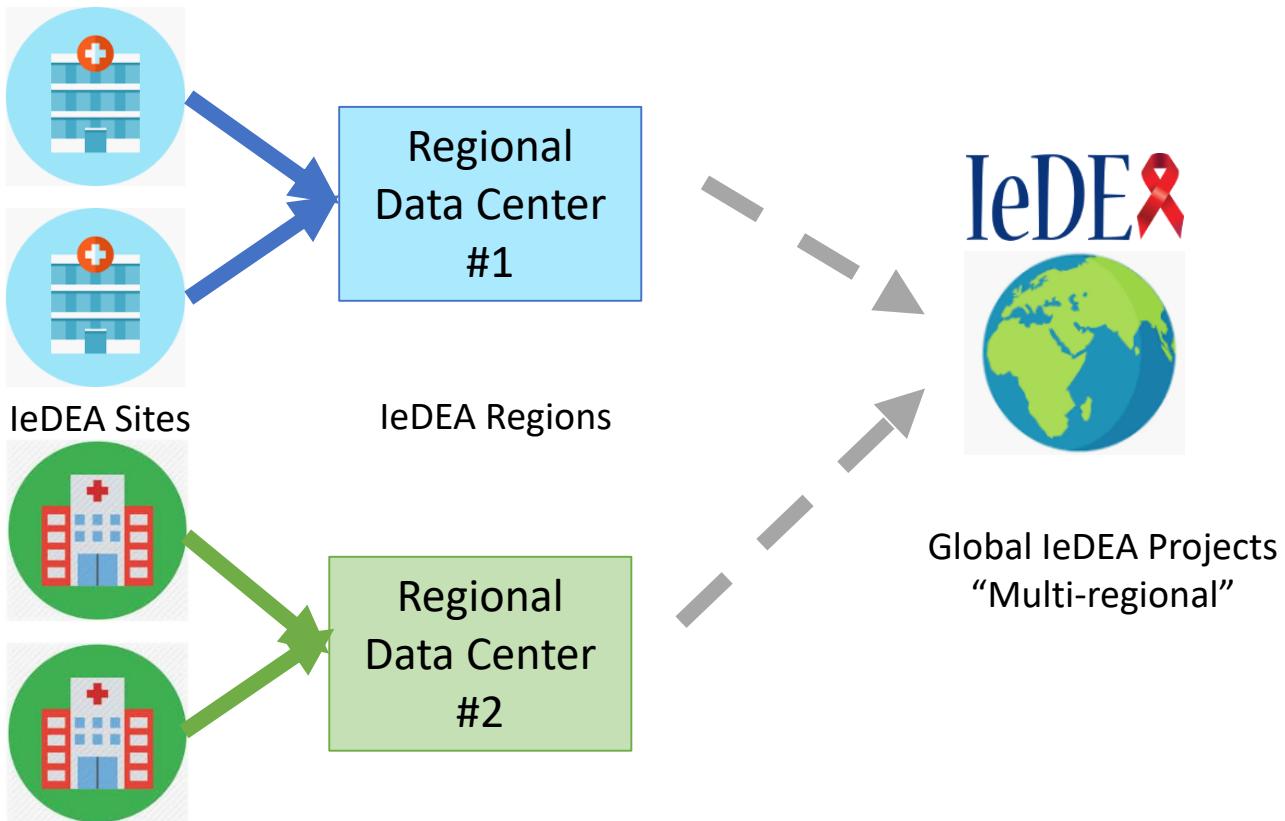
# IeDEA is a collaboration, not a single network

- 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

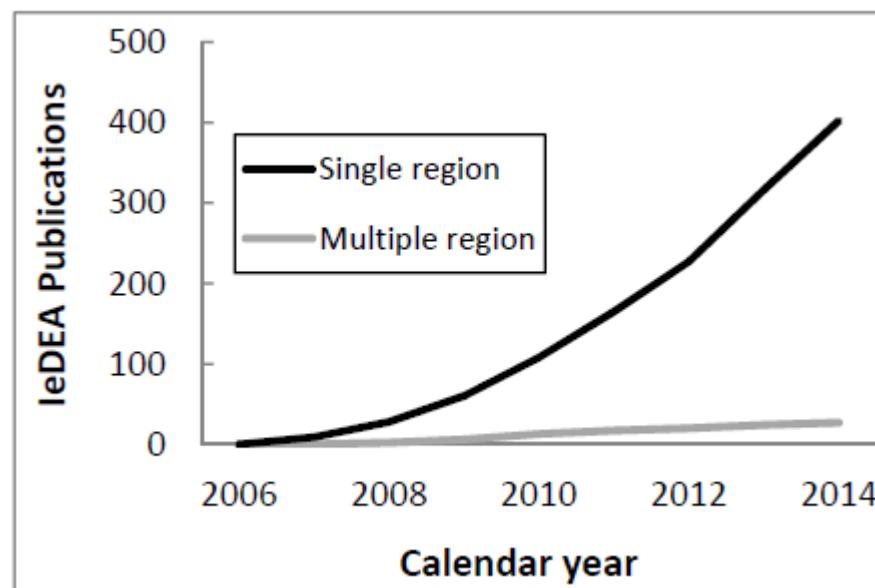
- Data from every clinic can be different.  
- Data at every Regional Data Center can be different.  
- Global IeDEA data are not stored centrally – subsets of the data are merged for specific projects.
- Sites and Regions have the ultimate say in whether their data is included for a specific project.



# In the Early Days of leDEA...

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

Cumulative number  
of leDEA 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  
*leDEA 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) leDEA Data Exchange Standard (DES)

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



**Data Transfer Protocol for leDEA Multi-regional Collaboration**

Approved by the leDEA Executive Committee on Thursday August 13, 2015

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

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

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

Table Name	Description	Not Yet Designated
tblART	antiretroviral drugs	
tblBAS	basic Information	
tblCANC	cancer diagnosis	
tblCENTER	site-specific information	
tblCEP	clinical events including serious non-AIDS conditions	X
tblDELIVERY_CHILD	delivery information related to child	X
tblDELIVERY_MUN	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	patients 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	

**tbCENTER**  
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 leDEA 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
- 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



leDEA International epidemiologic Database for Evaluate AIDS		
Data Transfer Protocol for leDEA Multi-regional Collaborations		
Approved by the leDEA Executive Committee on Thursday August 16, 2012		
Appendix A (leDEA-DT Policy) revised by leDEA Executive Committee Working Group on April 14, 2015		
<b>DEFINITION</b> Relation to HICDP/NON-HICDP		
Field	Format	Description
CENTER	character	Code for Clinic/Hospital where specimen is seen. Needs to be unique across all sites.
PROGRAM	character	Program with which the center is associated.
NAME	character	Proper name to identify center.
CITY	character	[Optional] Proper name to identify city.
PROVINCIAL	character	[Optional] Proper name to identify province.
DISTRICT	character	[Optional] Proper name to identify district.
CTY	character	[Optional] Proper name to identify county.
COORD_OF_LAT	Numeric	Latitude of the site/ location.
COORDINATION	Numeric	Longitude of the site/ location.
LEVEL	integer	Code for level of care.
ADULTPED	character	Population the center serves. "PEDI", "ADULT", or "BOTH".
OPENED	YYYY-mm-dd	[Optional] date of opening of clinic; earliest date for which data can be collected.
CLOSED	YYYY-mm-dd	Date of closing of dataset.
EX_WRTTS	YYYY-mm-dd	Date when the last event added to the cohort.
SHUT_CENTER	YYYY-mm-dd	Date when the site was dropped from the cohort.

# What is REDCap ?



~5440 sites  
142 countries  
1.3M users

[projectredcap.org](http://projectredcap.org)

# Representing the leDEA 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

Actions: [Modify instrument](#)[Download PDF of instrument\(s\)](#)

# Example: Tables

## 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 definition

Every Patient ID must have one and only one entry in `tblBAS`.  
Use HTML tags to define sections or other document structure.

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

Expand

# Initial REDCap Mapping

**leDEA**  
International open  
Databases to Evaluate  
Antiretroviral  
Strategies

Data Transfer Protocol for leDEA

Approved by the leDEA Executive Committee

Appendix A (leDEA DES) by leDEA Data Harmonization Group on April 14, 2015

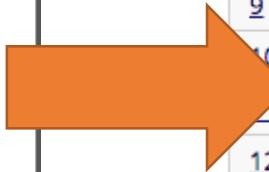
**Appendix A: leDEA-DES Tables**  
(note: the Data Harmonization Working Group will be responsible for recent revisions of this section of the data transfer protocol both tables are approved and designated as either HICDEP+ or Non-HICDEP provide additional documentation of the modified or additional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

Date last revised: Tuesday April 14, 2015 (Text in green represents 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	nucleotide 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 newborns	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:	Code for the site situation (facility location) 1 = Urban 2 = Mostly urban 3 = Mostly rural 4 = Rural 9 = Unknown
LEVEL	numeric:	Code for level of care 1 = Health centre 2 = District hospital 3 = Regional, provincial or university hospital 9 = Unknown
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



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

# 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": "", "table_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_deprecated_1": "", "variable_deprecated_d": "", "has_codes": "", "code_format": "", "code_list": "", "code_file": "", "codes_print": "", "variable_metadata_complete": ""}, {"record_id": "1", "redcap_repeat_instrument": "variable metadata", "redcap_repeat_instance": 1, "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": "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_deprecated_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, "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_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: \nhttp://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_deprecated_d": "", "has_codes": "1", "code_format": "3", "code_list": "", "code_file": "", "codes_print": "0", "variable_metadata_complete": "0"}, {"record_id": "1", "redcap_repeat_instrument": "variable metadata", "redcap_repeat_instance": 3, "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_SD", "data_format": "4", "description": "Date of initiation of treatment", "variable_order": "3", "code_text": "", "variable_required_1": "1", "variable_deprecated_1": "0", "variable_deprecated_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": 4, "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_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_instrument": "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_format": "4", "description": "Date of stopping of treatment", "variable_order": "5", "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"}]
```

# Data Model Browser

**leDEA DES Browser**

**leDEA Data Exchange Standard**

This site provides an auto-generated, web-browsable version of the leDEA Data Exchange Standard (leDEA DES), a **common data model for sharing observational HIV data** developed by the [International epidemiology Databases to Evaluate AIDS \(leDEA\)](#). More information on the data model is available on our [GitHub page](#).

leDEA DES Quick Reference Diagram ( [Download PDF](#), last updated 2020-01-03)  
leDEA Multiregional Data Transfer Protocol (Word Document) ( [Download](#), last updated 2017-02-17)

[>Show Draft](#) [>Show Deprecated](#)

Data Tables	
Table	Content
tblART	Antiretroviral medication
tblART_MUM	Antiretroviral Medication of mother in cases where mother is not enrolled in cohort
tblIBAS	*Required Basic information
tblCANC	Diagnosis of cancer
tblCENTER	Site-specific information
tblDELIVERY_CHILD	Delivery information related to child
tblDELIVERY_MUM	Delivery information related to mother

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



# 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  
*leDEA 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  
*leDEA Data Exchange Standard*

→ Clear data standards

(2) Consortium Portal  
*Harmonist Hub*

→ Managed research workflow

(3) Data Quality  
*Harmonist Data Toolkit*

How can leDEA investigators  
request data for projects?

# Before...

- leDEA 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 leDEA DES**.

Original data request  
(created by hand!)

**STANDARD OPERATING PROCEDURES FOR**  
**DATA TRANSFER**

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

Concept # 2.011  
Version 1.1 February 2014

Contacts

Kara Wools-Kaloustian M.D. M.S.  
Co-PI East African leDEA  
Indiana University School of Medicine  
[Kwools@iupui.edu](mailto: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](mailto:rvreeman@iu.edu)

Beverly Musick, M.S.  
Regional Data Manager East Africa leDEA  
Indiana University School of Medicine  
[bmusick@iupui.edu](mailto:bmusick@iupui.edu)

**Data files**

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



## International Epidemiologic Databases to Evaluate AIDS

Home &gt;&gt; Concepts Overview &gt;&gt; Concept Status

## Concept Status

[ALL CONCEPTS](#) [DATA AND APPROVED](#) [ABS AND MS](#)

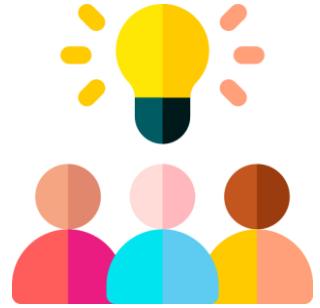
Lead Investigator	Region	Working Group	APPLY	RESET
	- Any -			

Concept Number	Title	Lead Investigator
2.042	Outcomes of children and adolescents treated with raltegravir in the leDEA consortium	Mary Davis
2.041	Descriptions and outcomes of HIV infected patients treated for tuberculosis without microbiologic confirmation in HIV care programs within the leDEA consortium	Kara Woo-Kalo
2.040	Diagnostic, treatment and outcome of extra-pulmonary Tuberculosis in HIV-co-infected adults and children	Mari Ballieux
2.039	leDEA-WHO collaboration 2016: Adolescent treatment outcomes in the leDEA global consortium	Anne Sohr
2.038	leDEA-WHO Collaboration: Global analysis of delays from ART eligibility to antiretroviral treatment initiation among adults	Denis Nash-Cerbo
2.037	leDEA-WHO collaboration: global analysis of the pre-ART cascade and delay from diagnosis to start of antiretroviral therapy in HIV-infected children aged 0-19 years.	Valerie Lero
2.036	leDEA-WHO Collaboration: Global analysis of retention in care in initial HIV care and treatment program	Elizabeth Zani
2.035	Update of mortality estimates of HIV-infected patients for Spectrum	Consuelo Yian
2.034	Association between clinic-level factors and individual retention, engagement, and loss to follow-up following ART initiation in the leDEA collaboration from 2009 to 2014	Bill V... - I...

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 leDEA 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 leDEA's scientific portfolio
- 
- Not a separate website, just another module installed on  REDCap  
Research Electronic Data Capture

# IeDEA Hub: Password-Free Login

The screenshot shows the IeDEA Hub login page. At the top, it says "Access IeDEA Hub". Below that, a message states: "IeDEA Hub ([iedeahub.org](https://iedeahub.org)) is the internal coordination website for the International Epidemiology Databases to Evaluate AIDS. Please visit [iedea.org](https://iedea.org) for our public site." A "Request Access Link" button is visible. An "Email Address" input field contains the placeholder "Type your email address.". A "Get Link" button is below the input field. A yellow arrow points from the "Get Link" button to an incoming email message. The email message is from "IeDEA Hub <harmonist@vanderbilt.edu>" and is titled "Review Hub Access". It includes a timestamp "Mon 3/5/2018 1:38 PM" and a recipient name "Stephany Norah". The body of the email contains a link: "Here is your link to access the IeDEA Hub: <https://redcap.vanderbilt.edu/plugins/iedea/index.php?token=MQ1>". A note at the bottom of the email says: "This link is unique to you and should not be forwarded to others. This link will expire in 7 days. You can request a new link at any time, which will invalidate the old link. If you are logging into the Hub from a public computer, please remember to log out of the Hub to invalidate the link." To the right of the email, a callout box contains the text: "If email address is in a list of active users, send an email with a URL that contains a one-time use token."

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

# leDEA Hub: Home Page

leDEA  Home Requests 2 Concepts Publications Data Hub Admin 0 CN Stephany Duda ▾

## Home Page

 The leDEA Hub updated to v2.10 on 2020-03-10. Contact us at [harmonist@yumc.org](mailto:harmonist@yumc.org) if you encounter any unexpected website behavior.



Open Requests	2
Concept Sheets	1
Abstracts	
Manuscripts	1
Posters	
Fast Track	
Other Items	

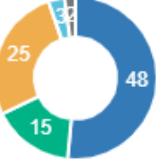
My Requests	<a href="#">View more</a>
<i>No requests available</i>	

Recent Activity	<a href="#">View more</a>
✓ Per v submitted a vote for Cervical Cancer Prevention and Care Cascade for wo...	2 days ago

Deadlines and Events	
Date	Text
2020-07-06	+7 days AIDS 2020: Virtual begins
2020-07-06	+7 days TB RePORT-leDEA Collaboration Workshop ( <a href="#">register</a> )
2020-08-04	+36 days leDEA 4 FOA submission due date
2020-11-16	+140 days HIV Pediatrics Workshop begins (postponed)

Active Data Calls	0
Check the list of active data calls for leDEA multi-regional projects.	
<a href="#">View Data Calls</a>	

Hub Metrics	<a href="#">View more</a>
<b>Requests</b> All leDEA Hub requests by category.	



Category	Value
Category A	25
Category B	48
Category C	15
Category D	3

Quick Links
<a href="#">Link 1</a> ▾

# leDEA Hub – Data page

The leDEA Hub navigation bar includes links for Home, Requests (0), Concepts, Publications, Data Hub (the active tab), Admin (1), and a user dropdown for Stephany Duda.

## Data Hub

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

The Data Hub section features four main tools:

- Request leDEA data for your approved concept** (highlighted with a red box)
- Check and submit data for an active data call**
- Retrieve data uploaded for your project**
- Explore the different types of leDEA data** (Coming Soon)

### Data News

No data news to display

### Recent Data Activity

↑ Beverly [REDACTED] uploaded EA data for MR180.	7 hours ago
↑ Karu [REDACTED] uploaded CN data for MR180.	8 hours ago
↓ Belinda C [REDACTED] downloaded CN data for MR186.	19 hours ago

### Data Log

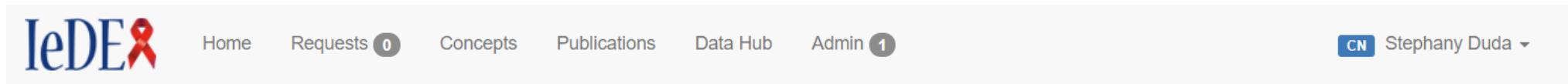
Track uploads and downloads of leDEA patient-level datasets.

### Harmonist Data Toolkit

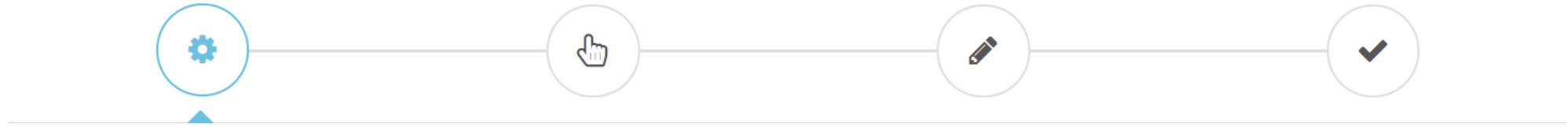
Go to the Toolkit webpage (without a data submission request).

### iedeadeas.org

# Creating a Data Request: Step 1



The image shows the LeDEA website's navigation bar. It includes the LeDEA logo with a red ribbon icon, a search bar, and links for Home, Requests (0), Concepts, Publications, Data Hub, Admin (1), and a user profile for CN Stephany Duda.



## STEP 1: Setup

The **LeDEA Data Request Builder** is going to walk you through four steps:

1. Identify the concept for which you want patient-level data,
2. Pick the variables you need from the current version of the [LeDEA Data Exchange Standard](#),
3. Define your inclusion/exclusion criteria, suggested data submission date, and who will have permission to download the data, and
4. Preview your draft Data Request.

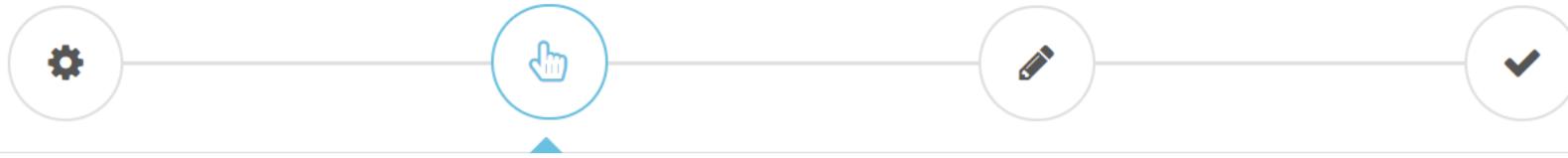
At the end, we'll guide you through [making your draft public](#) so you can share it with the LeDEA Data Harmonization Working Group for feedback.

**Select Your Concept:**  For test requests, select MR000.

**Setup Type:**  Create new data request  Load draft

[Save and continue](#)

# Creating a Data Request: Step 2



## STEP 2: Choose Variables

**MR116:** Data Request #63

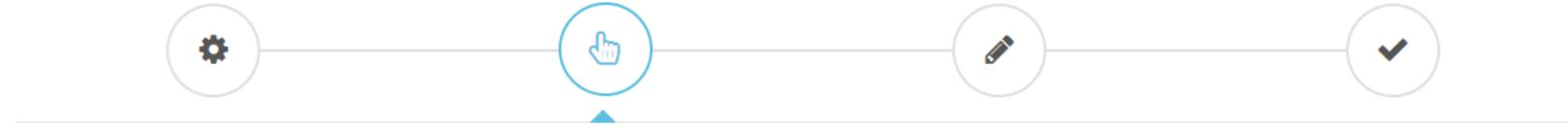
Select the leDEA DES tables and variables you need for your analysis.

Keep in mind that the leDEA DES only specifies how data are transferred among regions; it does not define what data each region collects. Therefore **not all these variables will be available from leDEA 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.

tbIBAS <span style="background-color: yellow;">2</span> <input type="checkbox"/> Select All				
Select	Field	Availability	Format	Description
<input checked="" type="checkbox"/>	PATIENT		Character or Numeric	Code to identify patient (Cohort Patient ID)
<input checked="" type="checkbox"/>	PROGRAM		Character	Direct one-to-one link from the patient to their program
<input type="checkbox"/>	BIRTH_D		YYYY-MM-DD	Birth date
<input type="checkbox"/>	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

# Creating a Data Request: Step 2.5

The LeDEA navigation bar includes links for Home, Requests (3), Concepts, Publications, Data, Admin (2), and a user profile for Stephany Duda.



## STEP 2: Choose Variables

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

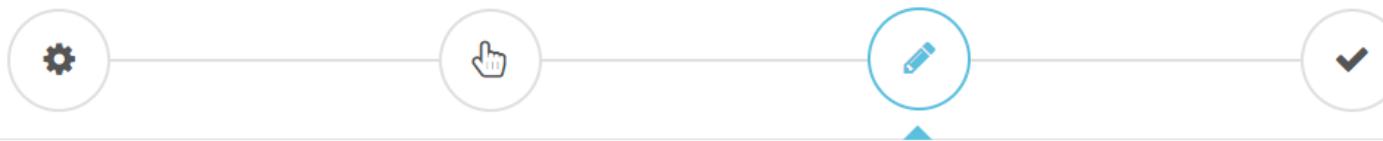
Keep in mind that the leDEA DES only specifies how data are transferred among regions; it does not define what data each region collects. Therefore **not all these variables will be available from leDEA regions**. Known variable availability (based on previous requests) is marked in the "Availability"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

tbIBAS	15	<input type="checkbox"/> Select All	▼
tbILTFU	8	<input type="checkbox"/> Select All	▼
tblVIS	4	<input type="checkbox"/> Select All	▼
tbILAB_CD4	5	<input checked="" type="checkbox"/> Select All	▼
tbILAB_RNA	4	<input type="checkbox"/> Select All	▼
tbIART	7	<input type="checkbox"/> Select All	▼
tbIDIS	5	<input type="checkbox"/> Select All	▼
tbICENTER	14	<input type="checkbox"/> Select All	▼
tbIPROGRAM	2	<input checked="" type="checkbox"/> Select All	▼
tbIART_MUM		<input type="checkbox"/> Select All	▼
tbICANC		<input type="checkbox"/> Select All	▼

Previous

Save and continue

# Step 3

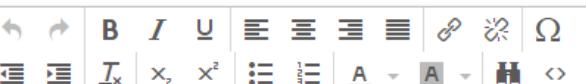


## STEP 3: Add Details

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 ( $\geq 16$  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 leDEA Data  
Harmonization  
Working Group  
(not part of Word  
document Data  
Requests)

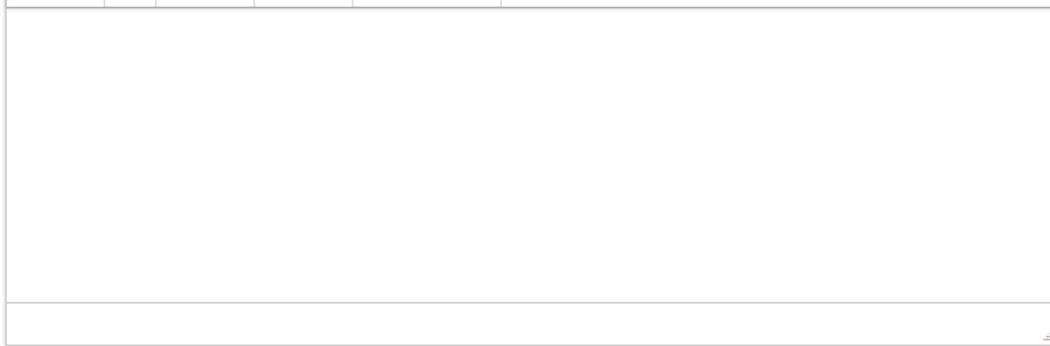


Due Date \*required 2018-10-17 

Preferred File Format

- CSV
- SAS
- Stata
- SPSS
- R
- Other (describe)

File Format Details

Choose your Data Downloaders \*required All Regions ▾

Judy Lewis

Olga [REDACTED]

Eva E [REDACTED]

Nicola [REDACTED]

Cam Ha [REDACTED]

Brenna [REDACTED]

Drag and Drop

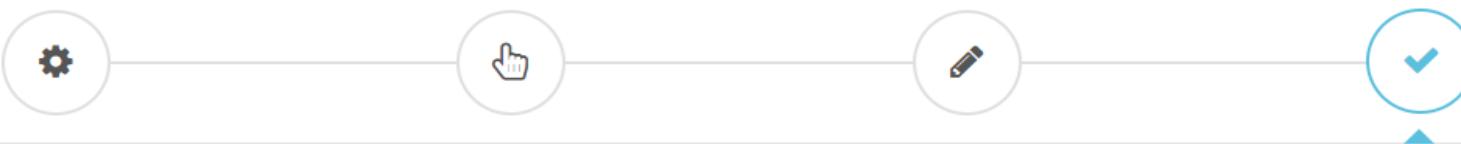


Stephany Duda

*Not sure about Data Downloaders yet.*

Save & Stay

# Step 4



## 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.



International epidemiology Databases to Evaluate AIDS (leDEA)

DATA TRANSFER REQUEST – MR131

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

Data Due: 17 October 2018

### Research Contact(s)

Anna F [REDACTED]

[user.email@example.com](mailto:user.email@example.com)

Brenda C [REDACTED]

[user.email@example.com](mailto:user.email@example.com)

### Data Contact

[Previous](#)

[Save and create PDF](#)

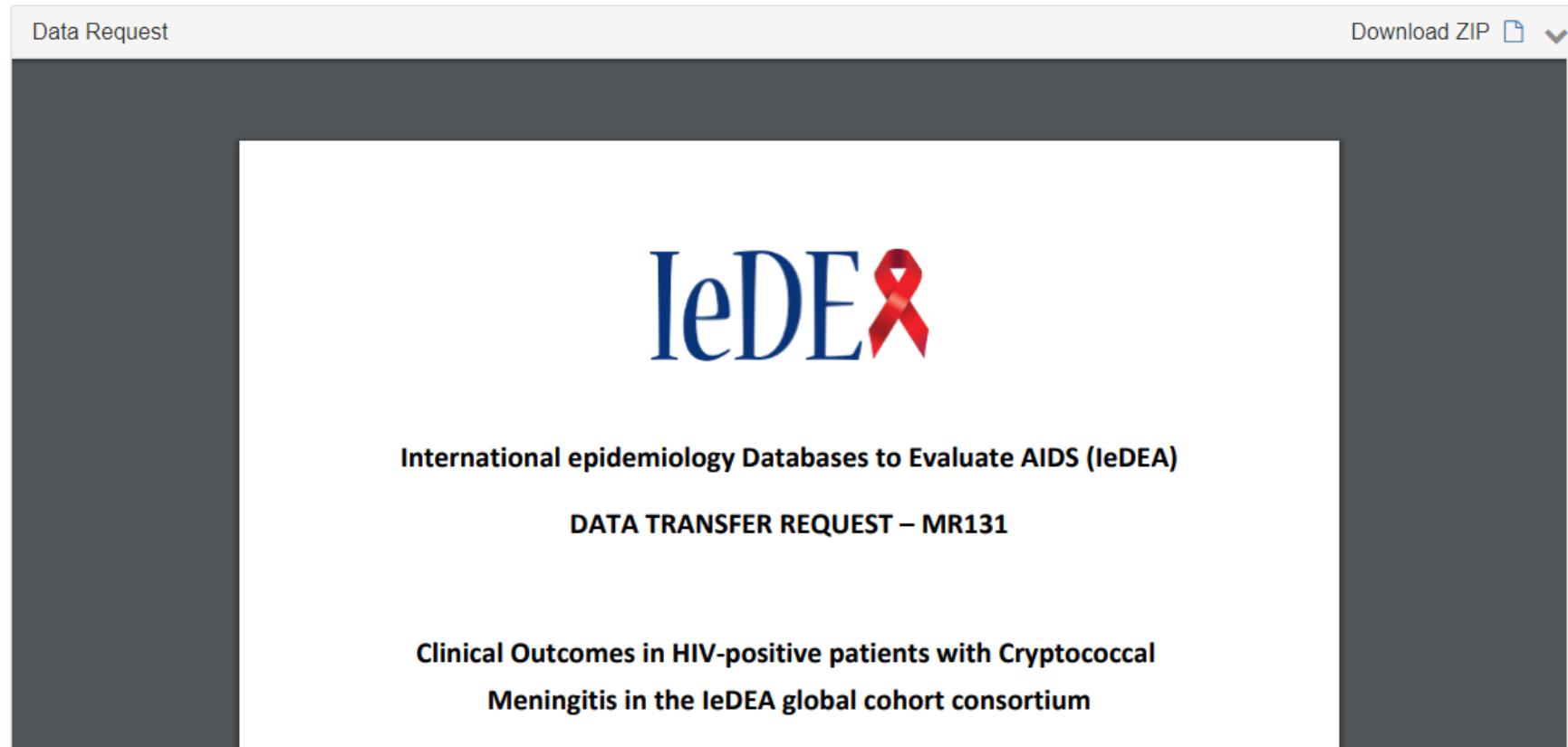
## Steps Complete ✓

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  ▾



The screenshot shows a PDF document titled "Clinical Outcomes in HIV-positive patients with Cryptococcal Meningitis in the IeDEA global cohort consortium". The document is a DATA TRANSFER REQUEST – MR131. It features the IeDEA logo at the top, followed by the title and subtitle. The background of the document is white, and the text is black. The overall layout is clean and professional.



International epidemiology Databases to Evaluate AIDS (leDEA)

### DATA TRANSFER REQUEST – MR131

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

Data Due: 17 October 2018

##### Research Contact(s)

Anna [REDACTED]  
user.email@example.com

Brenda [REDACTED]  
user.email@example.com

##### Data Contact

Stephany Duda  
user.email@example.com

## Automatically Generated Data Request

#### 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

leDEA 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 leDEA Data Exchange Standard (leDEA DES).

The following leDEA 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	<a href="#">View Data Activity Log</a>	<a href="#">View Past Data Calls</a>
<b>Open Data Calls</b>							
<i>No Open Data Calls available</i>							

Completed Data Calls												
Due Date	Data Request Details			Data Contact	AP	CA	CN	EA	NA	SA	WA	Actions
2021-09-08 -23 days	<b>MR180</b> Infectious and non-infectious causes of liver disease among adults living with HIV from low and middle income countries, the Sentinel Research Network of leDEA		Draft ID: 54	Karen [REDACTED] (WA)	<input checked="" type="checkbox"/>	<a href="#">Upload Data</a>						
2021-09-15 -16 days	<b>MR181</b> Diagnostic value of serological markers for liver steatosis and fibrosis in people living with HIV: a multicenter study from the leDEA Network		Draft ID: 55	Hilary [REDACTED] (CN)	<input checked="" type="checkbox"/>	<a href="#">Upload Data</a>						
2021-09-20 -11 days	<b>MR185</b> 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 leDEA		Draft ID: 60	Kathy [REDACTED] (CA)	<input checked="" type="checkbox"/>	<a href="#">Upload Data</a>						
	<b>MR186</b> Mental Health and Substance Use Disorders in Aging Adults Living with HIV		Draft ID: 61	Leanne [REDACTED] (CA)	<input checked="" type="checkbox"/>	<a href="#">Upload Data</a>						

# The Harmonist Project

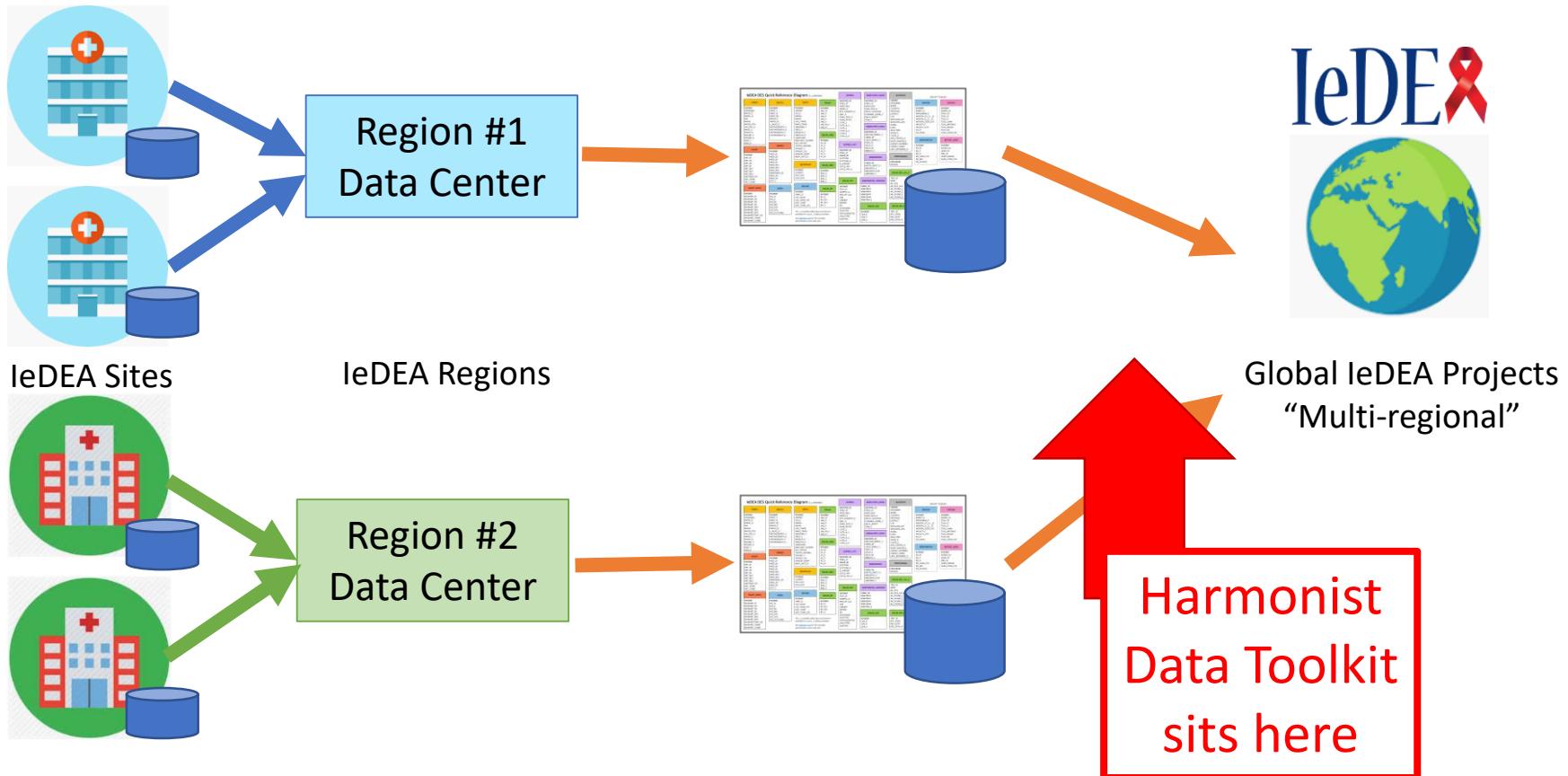


- (1) Support for Data Model  
*leDEA Data Exchange Standard* → Clear data standards
- (2) Consortium Portal  
*Harmonist Hub* → Managed research workflow
- (3) Data Quality and Sharing Tool  
*Harmonist Data Toolkit* → High quality data

# 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 leDEA Project Portal (leDEA Hub)

The screenshot shows the leDEA Hub interface. At the top, there's a navigation bar with links for Home, Requests (2), Concepts, Publications, and Data. A user profile for "Judy Lewis" is on the right.

The main content area has a title "Data" and a sub-section "Check and Submit Data". It includes a "Coming Soon" button for exploring leDEA data types. Below this, there are four cards: "Request leDEA data for your approved concept" (Request Data button), "Check and submit data for an active data call" (Submit Data button, highlighted by a blue arrow labeled "1"), and "Retrieve data uploaded for your project" (Retrieve Data button).

Under "Check and Submit Data", there's a list of steps:

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).

Text below the steps says: "Data files should be formatted according to the [leDEA Data Exchange Standard \(leDEA DES\)](#)".

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

View Upload History | View Past Data Calls

Active Data Calls						PDF	TT	Actions
Due Date	Concept	Data Contact	Downloaders Assigned	Data Request				
2019-03-12	-3 days	MR116	Judy Lewis (TT)	2	2. Data Toolkit Practice Request A			<a href="#">Upload Data</a>

A large purple arrow labeled "2" points to the "Upload Data" button in the table row.



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](#)

## STEP 1 Upload files

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

### MR116 Active Data Request

**Title** Harmonist Data Toolkit Development: Request for leDEA DES Datasets from All Regions

**Hub Pages** MR116 on Hub [↗](#), Data Specification [↗](#) [PDF](#)

**Requested Tables** [tblBAS](#) [tblLT FU](#) [tblVIS](#) [tblLAB\\_CD4](#) [tblLAB\\_RNA](#) [tblCENTER](#) [tblPROGRAM](#)

**Requested Data Format**

**Contacts**

- Judy Lewis (TT) , Vanderbilt University
- Stephany Duda (CN) , Vanderbilt University
- Judy Lewis (TT) (*Data contact*) , Vanderbilt University

**Data Downloaders**

- Stephany Duda (CN) , Vanderbilt University
- Judy Lewis (TT) , Vanderbilt University

### Select Data Files

Upload data in the leDEA Data Exchange Standard (leDEA DES) format. **tblBAS** is required.

Allowed file formats include **CSV**, **SAS**, **Stata**, **SPSS**, or a **ZIP** containing multiple files of this type.

Select a single ZIP file or multiple files with Ctrl+Click

**Data files**

[Browse...](#) No file selected

### Use Sample Dataset

Launch the Toolkit with a sample dataset (fake data) for practice, testing, and demonstrations.

The sample dataset contains 48 intentionally error-filled records representing the following leDEA DES tables: [tblBAS](#) [tblLT FU](#) [tblVIS](#) [tblART](#) [tblLAB](#) [tblLAB\\_BP](#)

[Launch with Sample Data](#)



# 1. Upload Files to leDEA Toolkit

## 2. File Review/Data Quality Checks

### leDEA 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 leDEA 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 leDEA Tables

Table	Records	leDEA 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
tblLTU	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

**leDE** 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 2 Check data

View interactive summary of errors and download detailed results of data quality checks to review offline.

### Error Summary by Table

[Download error detail CSV](#)

Error description	Severity	Count	Action
Future date: ENROL_D	Error	1	<a href="#">View Detail</a>
Invalid Code: MODE	Error	1	<a href="#">View Detail</a>
Invalid Code: RECART_D_A	Error	2	<a href="#">View Detail</a>
Invalid Code: HAART_D_A	Error	2	<a href="#">View Detail</a>
BIRTH_D before 1920	Warn	3	<a href="#">View Detail</a>
Date before 1980: AIDS_D	Warn	1	<a href="#">View Detail</a>

Show 10 entries

Search:

Showing 1 to 6 of 6 entries

Previous 1 Next

**Continue to Summary**

Error checks completed

Your dataset contains **114 total errors** in **12 error categories** including **28 invalid codes**

If you have already reviewed the content of the dataset, proceed to the next step to generate a **summary of the data**.

[Continue to Step 3](#)

**Restart session**

Start over and upload a **revised or different dataset**.

[Upload new dataset](#)



# 4. Reproducible Reports

**STEP 3 Create summary**

Generate and download customized reports summarizing uploaded dataset.

**Customize Summary Report**

**File format for report**: PDF

**Select report content**:

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

**Data subgroup(s) for report**: All

**(Optional) Short title for report heading**:

**Date Histogram Options**  
Choose years to include in histograms  
Years: 2000 - present

**Generate summary PDF report**

**Dataset Summary**

Total number of patients in dataset: 28089

Report date: 2019-03-07

**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	3168	3116	2723	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**

**SITE: Hogwarts**  
(Note: 23 observations before year 2000 and 1 observation after year 2019 are hidden)

**Table 4: Summary statistics from uploaded tables**

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

Figure 1: Percent of Patients from tblBAS Included

## 5. Transfer Data

**leDEA** 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 4 Submit data

Submit dataset for selected concept.

#### Transfer Data for leDEA Concept

Ready to transfer data

**Dataset summary:**

- 28089 unique patient records included.
- 8 leDEA DES tables included.
- **Missing 4 variables requested by MR116 across 2 tables.**
- **114 potential data quality issues** detected.

**After transfer:**

- Uploaded data will be **stored for 30 days**.
- Data will be automatically deleted after 30 days. You can manually delete your uploaded datasets via the leDEA Hub.
- Approved data downloaders will be able to retrieve your data through the Hub.  
(Downloaders: Stephany Duda, Judy Lewis)

**Message to accompany your file upload (visible to Data Downloaders on the Hub):**

Click below to submit your data to secure cloud storage to be retrieved by Judy Lewis

**Submit Data**

## 6. Recipient Downloads Data

leDEA Home Requests 2 Concepts Publications Data Judy Lewis >

< Back to Data

Download security:  
Login with multifactor authentication

### Retrieve Data

All leDEA data requests that you have access to are displayed here. Uncollapse the menus to see individual file downloads and details. Downloads expire after 30 days. If you expect to have access to datasets that are not listed here, you may not be listed as a permitted Data Downloader on that data request. Contact the project lead and the Harmonist team to request permission.

MR116 | Data Request #2

-10 days until due 2

Title: Harmonist Data Toolkit Development: Request for leDEA DES Datasets from All Regions | Data Request #2

Data Contact: Judy Lewis (judy)

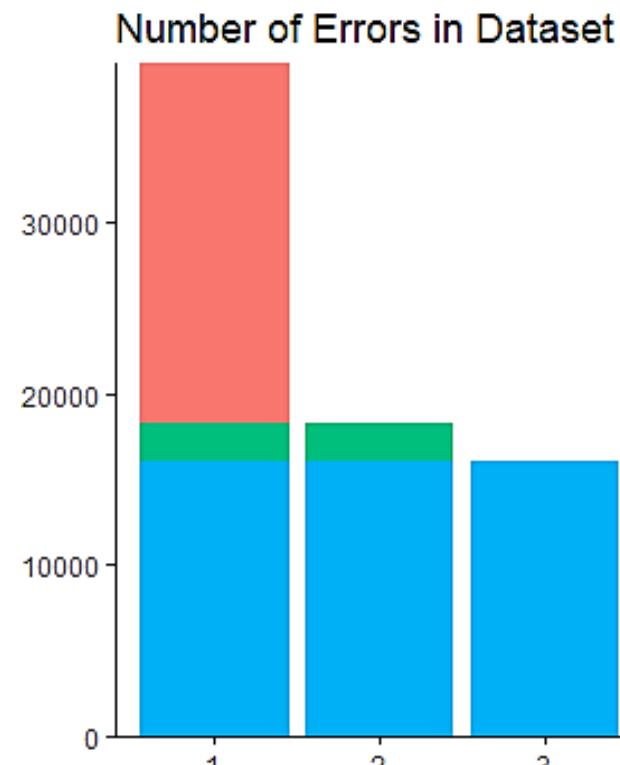
Data Due: 12 March 2019

Upload Date	Region	Submitted By	Filename	PDF	Expires On	Actions
+ 2019-03-20 15:06:42	TT	Judy Lewis	MR116_TT_Lewis_201903201506.zip		21 April 2019 +30 days	
+ 2019-03-18 12:01:32	CN	Hilary Vansell	MR116_CN_Vansell_201903181201.zip		21 April 2019 +30 days	



# 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



PATIENT(VIS)	PATIENT(RNA)	PATIENT(RNA)	ART_RS
PATIENT(RNA)	NAIVE_Y	ART_RS	
NAIVE_Y	PROPH_Y	DROP_Y	
PROPH_Y	CENTER		
CENTER	RNA_D		
RNA_D	VIS_D		
VIS_D	ART_RS		
ART_RS	DROP_Y		
DROP_Y			

# Toolkit Impact on Data Quality

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

No errors in final submission



- Duplicate Record
- Invalid Code
- Missing Required Variable
- Reason provided but date missing
- Y/N data in conflict with date

# 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 leDEA; 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](mailto:harmonist@vumc.org)



- Code available
  - [github.com/leDEA/Harmonist](https://github.com/leDEA/Harmonist)
  - [github.com/vanderbilt-redcap](https://github.com/vanderbilt-redcap)
  - [redcap.vanderbilt.edu/consortium/modules/index.php](https://redcap.vanderbilt.edu/consortium/modules/index.php)

# Thank You

- Harmonist and REDCap technical teams
- leDEA 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.