JAPAN ACADEMIC WORKSHOP

The Infectious Diseases Data Observatory (IDDO) and the Benefits of CDISC Standards for Academic Researchers

Kalynn Kennon, Head of Data Engineering & Informatics

Meet the Speaker

Kalynn Kennon

Title:
Head of Data Engineering & Informatics

Organization:

Infectious Diseases Data Observatory (IDDO), University of Oxford

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The author(s) have no real or apparent conflicts of interest to report.*

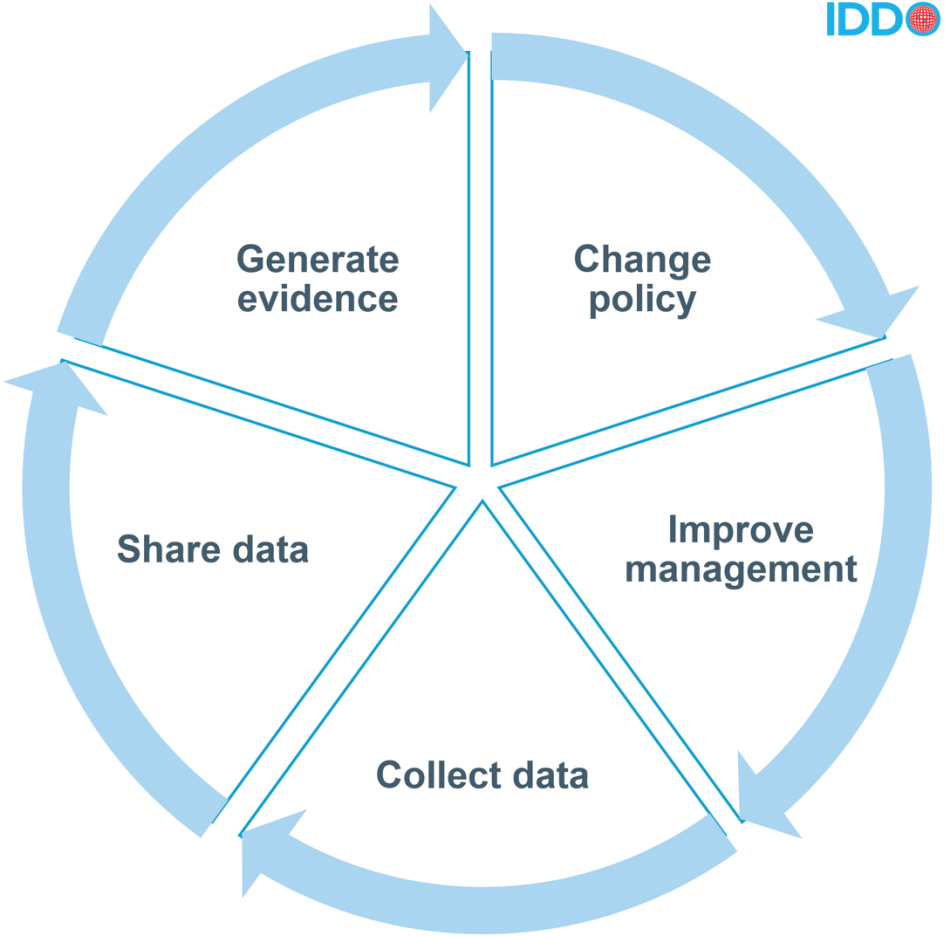
Infectious Diseases Data Observatory

Who are we?

The Infectious Diseases Data Observatory (IDDO) brings together members of the global infectious disease community across the research and humanitarian sectors to collaborate in the generation, analysis and application of data to improve outcomes for patients. IDDO's coordinating office is based in the Centre for Tropical Medicine and Global Health at the University of Oxford.

IDDO's mission

- To accelerate the effective treatment and control of infectious diseases by strengthening research and the **generation of evidence** for policy through **equitable data use**



Pioneers in generating new research evidence from existing data since 2009

- IDDO's model puts equity for global researchers at the forefront of efforts to improve outcomes for infectious disease-affected communities worldwide
- Model began in 2009 with sole malaria focus as the WorldWide Antimalarial Resistance Network (WWARN)
- Using WWARN's success as a blueprint for research into other infectious diseases, IDDO launched in 2016. It now…

Has over 2,000 global research contributors

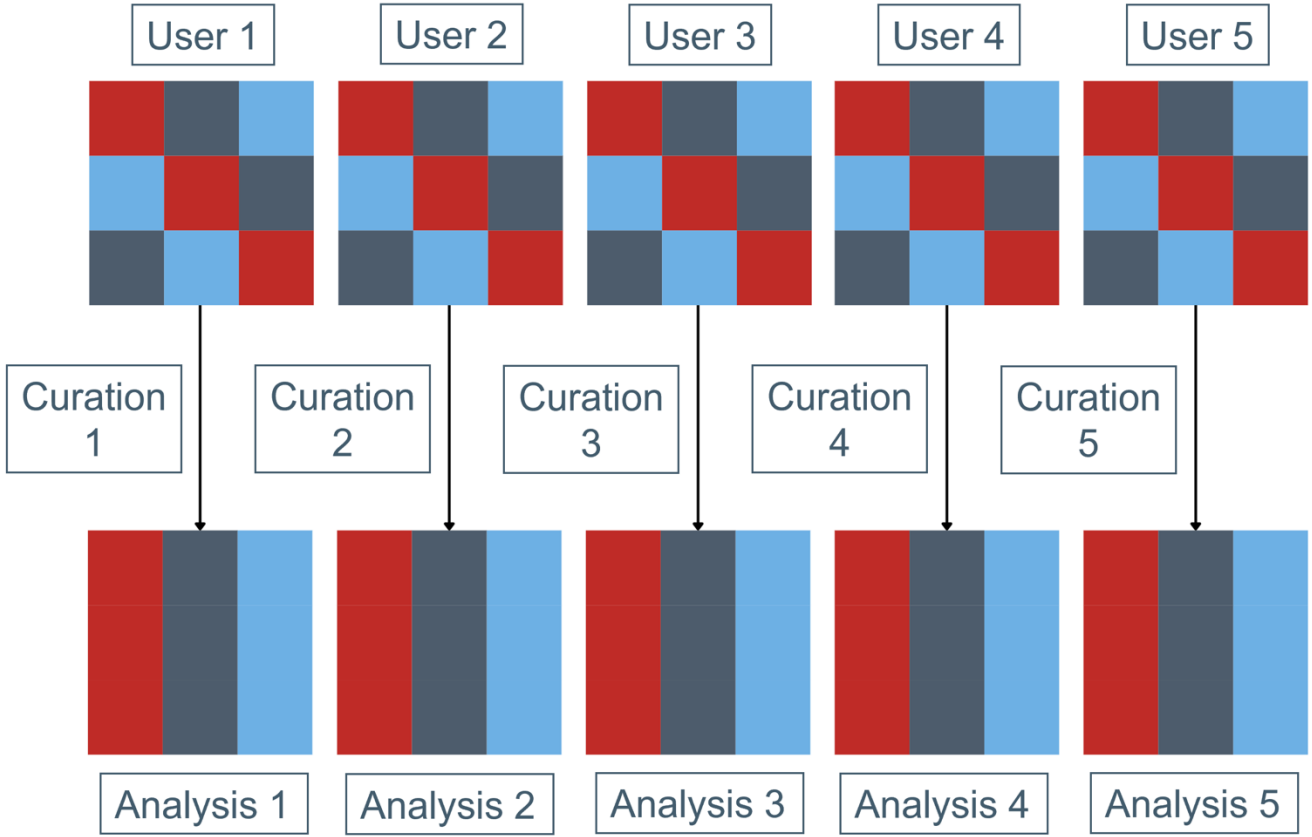
Hosts data from more than 1 million participant infections

Works on 16 research themes with more in the pipeline

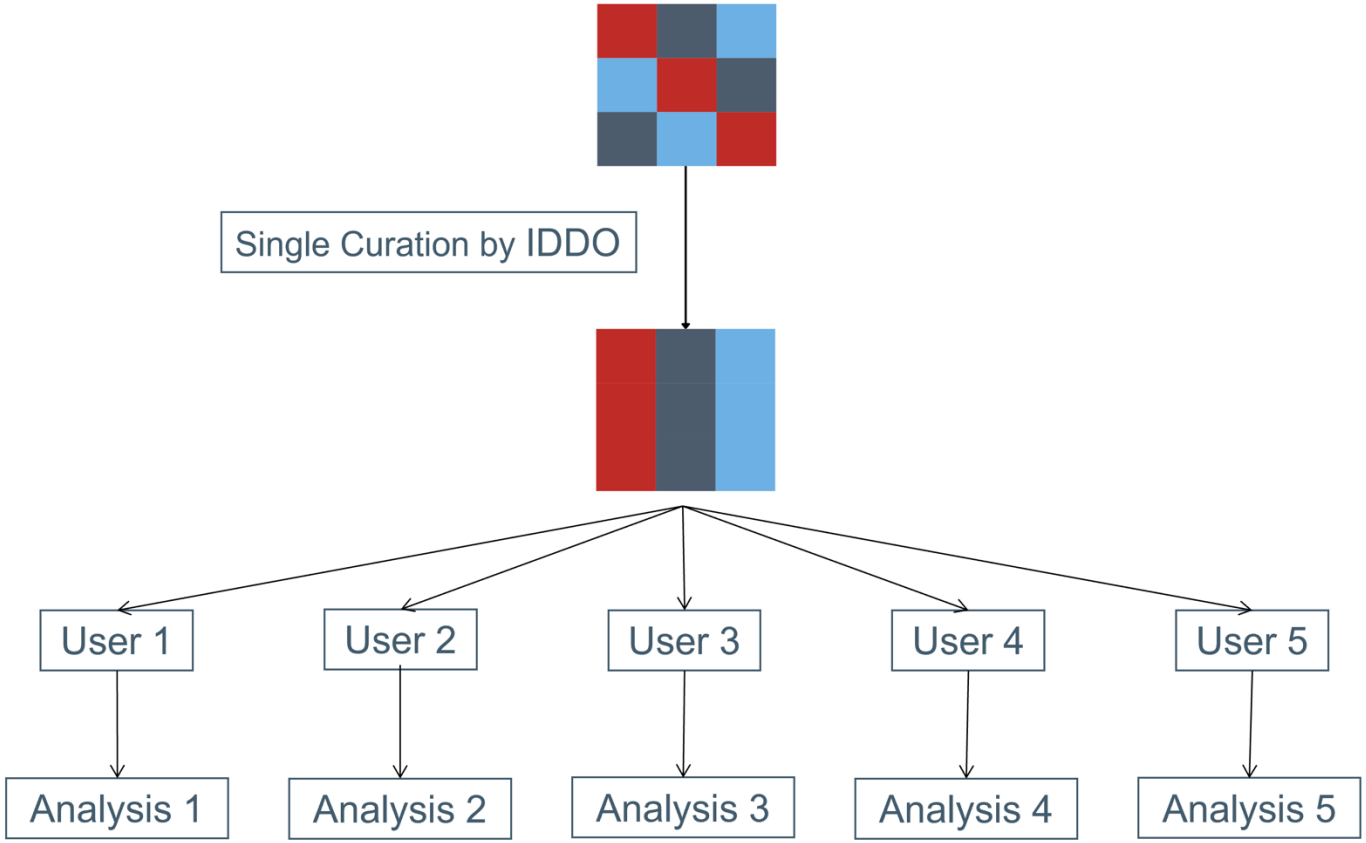
16 areas of focus, at varying stages

- Proposals to launch new research themes originate from those in, or working with, disease endemic research communities

Duplicated Effort in Data Standardization

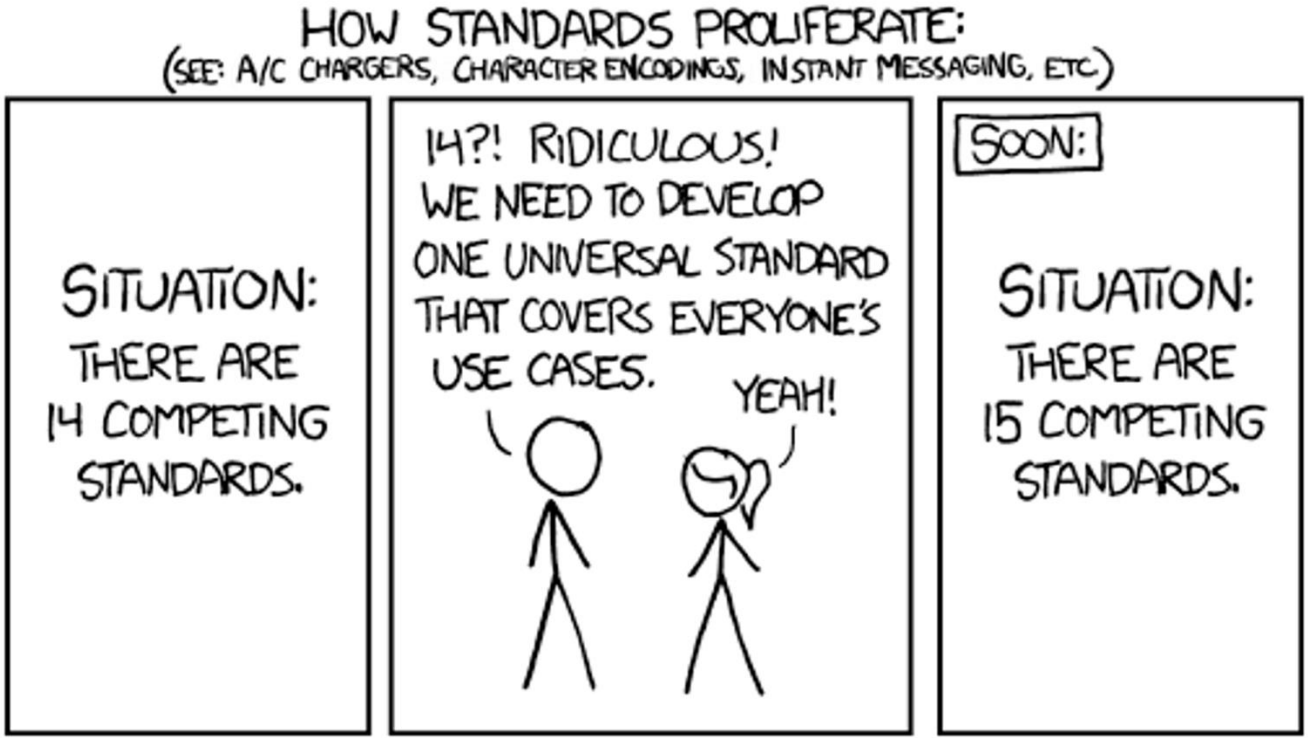


Data Standardization of data shared with IDDO

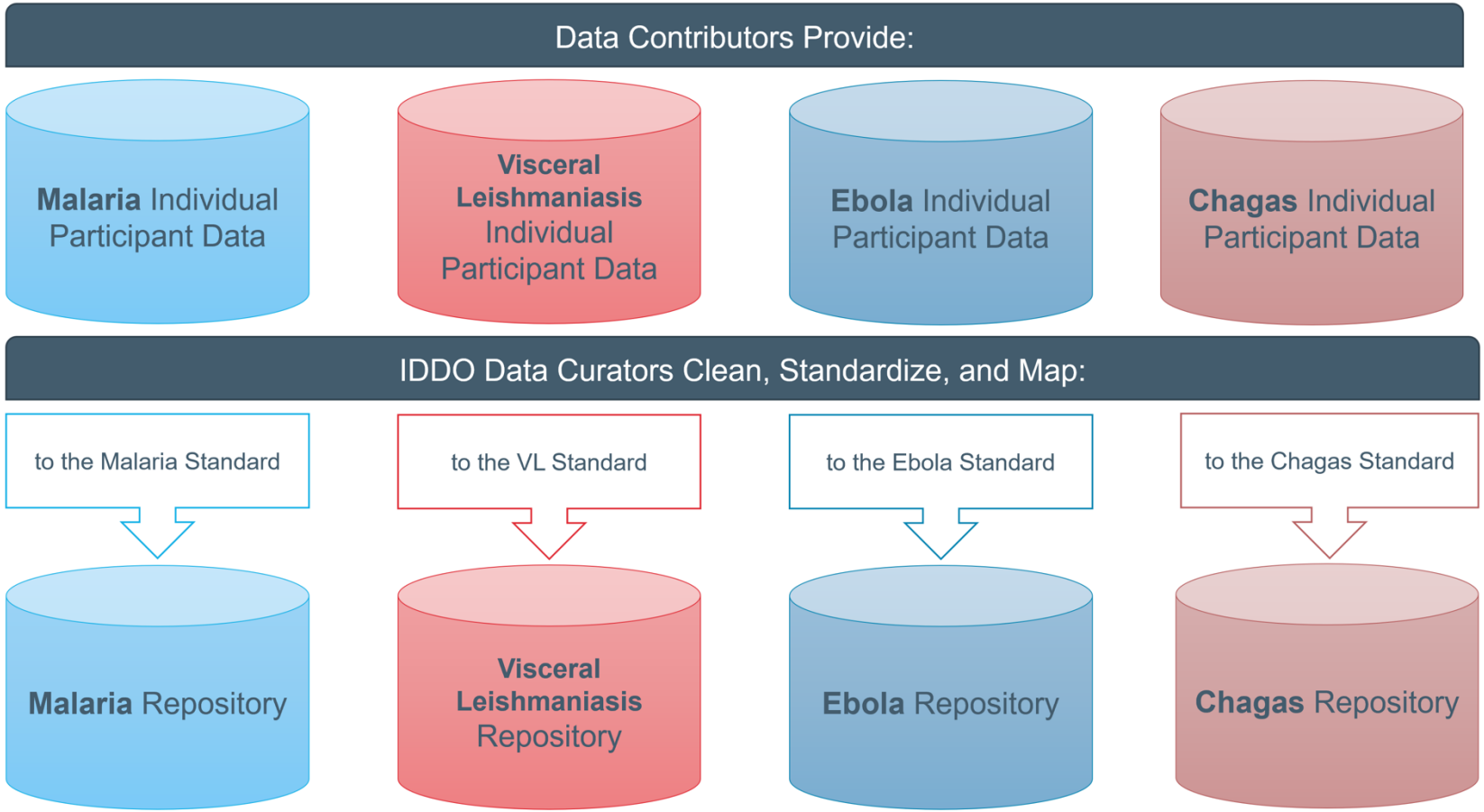


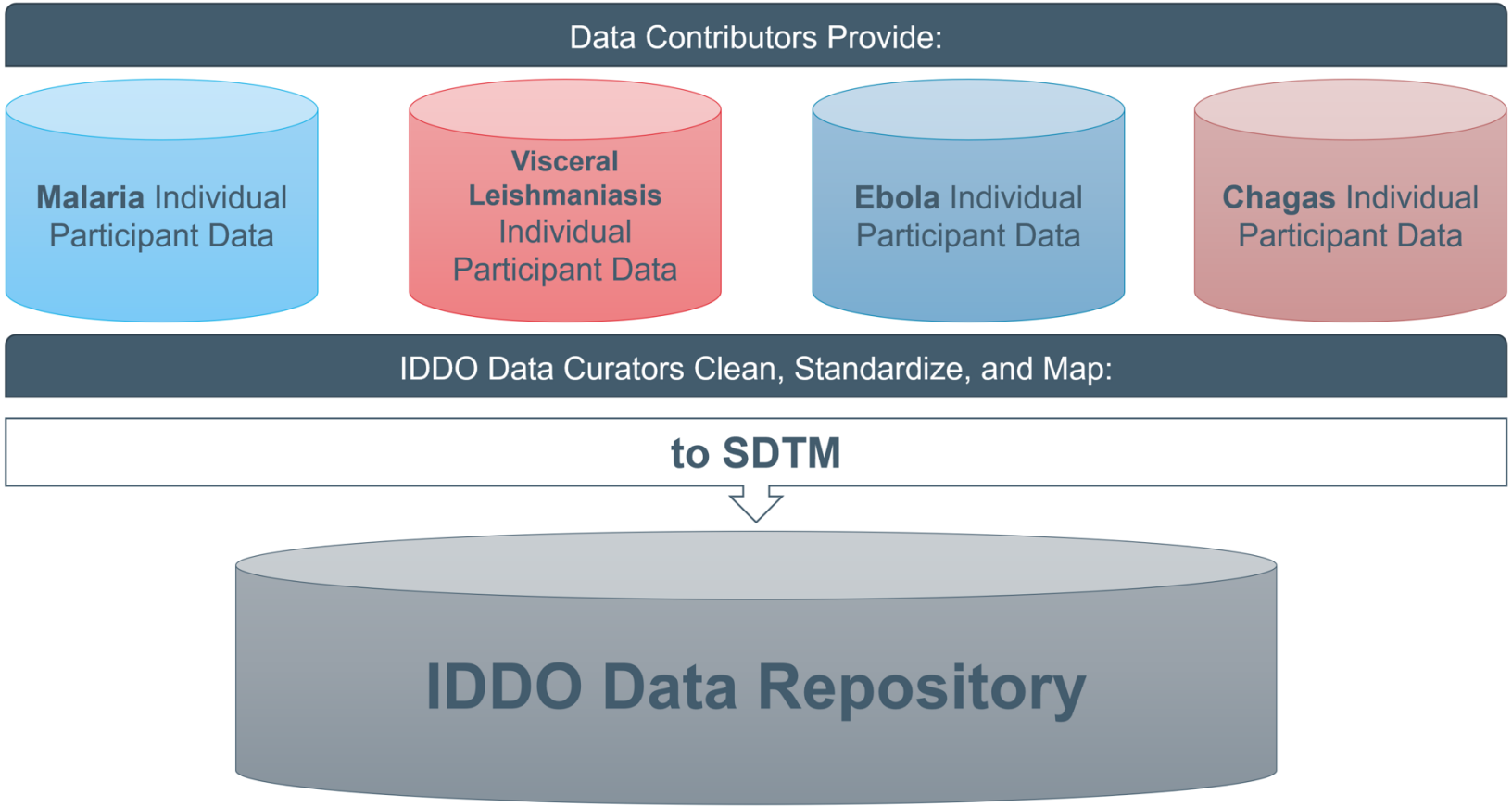
CDSIC SDTM Data Standard

Why do we use it?



PERMANENT LINK TO THIS COMIC: HTTPS://XKCD.COM/927/


Data Contributors Provide:
Malaria Individual Participant Data
Visceral Leishmaniasis Individual Participant Data
Ebola Individual Participant Data
Chagas Individual Participant Data
IDDO Data Curators Clean, Standardize, and Map:
to the Malaria Standard
to the VL Standard
to the Ebola Standard
to the Chagas Standard
Malaria Repository
Visceral Leishmaniasis Repository
Ebola Repository
Chagas Repository


Data Contributors Provide:
Malaria Individual Participant Data
Visceral Leishmaniasis Individual Participant Data
Ebola Individual Participant Data
Chagas Individual Participant Data
IDDO Data Curators Clean, Standardize, and Map:
to SDTM
IDDO Data Repository

Other Reasons for Choosing SDTM

- The ability to maintain granularity for those who want it, but also to standardize data to make easier for those who don't
- The flexibility to accommodate the differences in both study design and in the amounts and types of data collected in the studies contributed to IDDO
  - Metadata collected in the standard itself helps capture the differences in study information (e.g., the variety of timing variables) especially at theindividual test/event/intervention level

Data Standardization and Enrichment

INTRT	INDECOD	INCLAS	INCLASCD
Avigan	FAVIPIRAVIR	ANTIVIRALS FOR SYSTEMIC USE	J05
Favipiravir	FAVIPIRAVIR	ANTIVIRALS FOR SYSTEMIC USE	J05
Areplivir	FAVIPIRAVIR	ANTIVIRALS FOR SYSTEMIC USE	J05
Amoxicillin	AMOXICILLIN	ANTIBACTERIALS FOR SYSTEMIC USE	J01
Cefuroxime	CEFUROXIME	ANTIBACTERIALS FOR SYSTEMIC USE	J01
Flucloxacillin	FLUCLOXACILLIN	ANTIBACTERIALS FOR SYSTEMIC USE	J01

--ORRES, --TERM, --TRT allow for maintaining the details from the individual datasets

--STRESC/N, --MODIFY, --DECOD allow for enrichment and binning things together to facilitate easier analysis

SDTM Flexibility for Differing Levels of Metadata

SATERM	SAPRESP	SAOCCUR	SADTC	SASTDTC	SAENDTC	SADY	SASTDY	SAENDY	SADUR	SAEVLINT	SAEVINTX	Questions from CRF
Fever	Y	Y	14/07/2021			1						Date of assessment [14/07/2021] Do you have a fever? [YES]
SATERM	**SAPRESP**	**SAOCCUR**	**SADTC**	**SASTDTC**	**SAENDTC**	**SADY**	**SASTDY**	**SAENDY**	**SADUR**	**SAEVLINT**	**SAEVINTX**	Questions from CRF
Fever	Y	Y	14/07/2021	10/07/2021		1	-4					Date of assessment [14/07/2021] Do you have a fever? [YES] When did the fever start? [10/07/2021]
SATERM	**SAPRESP**	**SAOCCUR**	**SADTC**	**SASTDTC**	**SAENDTC**	**SADY**	**SASTDY**	**SAENDY**	**SADUR**	**SAEVLINT**	**SAEVINTX**	Questions from CRF
Fever	Y	Y	14/07/2021			1			P5D			Date of assessment [14/07/2021] Do you have a fever? [YES] For how long have you had a fever? [5 days]
SATERM	**SAPRESP**	**SAOCCUR**	**SADTC**	**SASTDTC**	**SAENDTC**	**SADY**	**SASTDY**	**SAENDY**	**SADUR**	**SAEVLINT**	**SAEVINTX**	Questions from CRF
Fever	Y	Y	14/07/2021			1				-P10D		Date of assessment [14/07/2021] Have you had a fever in the last 10 days? [YES]
SATERM	**SAPRESP**	**SAOCCUR**	**SADTC**	**SASTDTC**	**SAENDTC**	**SADY**	**SASTDY**	**SAENDY**	**SADUR**	**SAEVLINT**	**SAEVINTX**	Questions from CRF
Fever	Y	Y	14/07/2021	10/07/2021	13/07/2021	1	-4	-1		-PT48H		Date of assessment [14/07/2021] Have you had a fever in the last 48 hours? [YES] When did the fever start? [10/07/2021] When did the fever end? [13/7/2021]
SATERM	**SAPRESP**	**SAOCCUR**	**SADTC**	**SASTDTC**	**SAENDTC**	**SADY**	**SASTDY**	**SAENDY**	**SADUR**	**SAEVLINT**	**SAEVINTX**	Questions from CRF
Fever	Y	Y	14/07/2021			121					RECENTLY	Date of assessment [14/07/2021] Have you felt feverish recently? [YES]

CDISC Standards

How else do we use them?

CDISC Case Report Form development

As a key component of its data platform development process, IDDO works closely with research communities to develop a standard Case Report Form (CRF) tailored to each disease. The CRF does not prescribe what data to collect, but rather provides researchers with a standardised means of recording any data they do choose to collect from a study. This supports efficient, scientifically-valid generation and reporting of clinical data to streamline development of new treatments, regulatory submission and post-marketing research, as well as enabling data sharing, comparison and aggregation for high-quality, novel research outputs to address knowledge gaps.

Through the process described in the workflow below, IDDO works closely with researchers, regulators, pharma and policymakers with support from CDISC to develop consensus-based, freely-available data standards in compliance with existing CDISC standards .

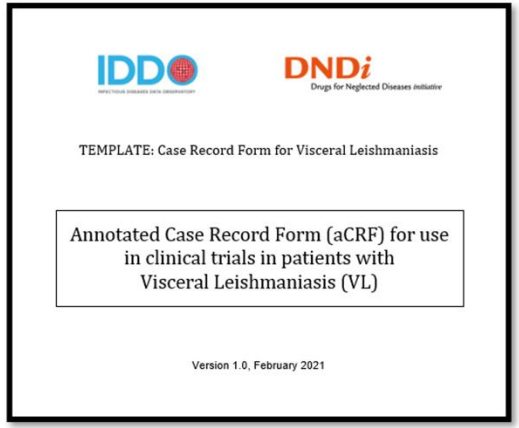
WWARN

Visit: http://www.wwarn.org/working-together/malaria-standards-CDISC

TEMPLATE:
Annotated Case Record Form (CRF) for use in clinical trials in patients with uncomplicated malaria (phase II/III)

V2.0 09th, APRIL, 2018

The WorldWide Antimalarial Resistance Network (WWARN) would like to acknowledge the following groups for their participation and contribution to the development of the CDISC malaria data standard: Bill & Melinda Gates Foundation, Clinical Data Interchange Standards Consortium (CDISC), Critical Path Institute (C-PATH), United States



IDDO DNDi

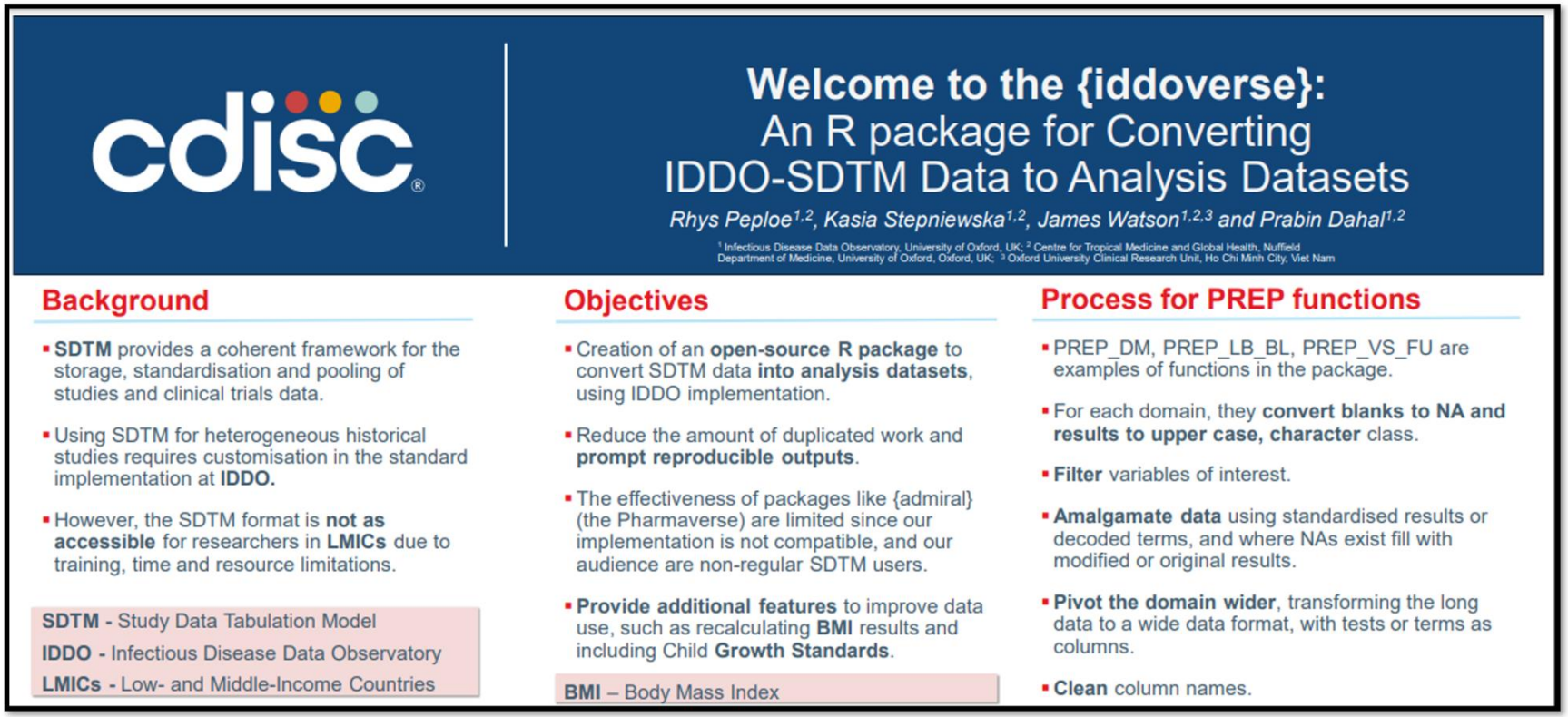
TEMPLATE: Case Record Form for Visceral Leishmaniasis

Annotated Case Record Form (aCRF) for use in clinical trials in patients with Visceral Leishmaniasis (VL)

Version 1.0, February 2021

https://www.iddo.org/cdisc-case-report-form-development

Packages for Analysis Datasets



Welcome to the {iddoverse}: An R package for Converting IDDO-SDTM Data to Analysis Datasets

Rhys Peploe[1,2], Kasia Stepniewska[1,2], James Watson[1,2,3] and Prabin Dahal[1,2]

[1] Infectious Disease Data Observatory, University of Oxford, UK; [2] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [3] Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam

Background

- **SDTM** provides a coherent framework for the storage, standardisation and pooling of studies and clinical trials data.
- Using SDTM for heterogeneous historical studies requires customisation in the standard implementation at **IDDO.**
- However, the SDTM format is **not as accessible** for researchers in **LMICs** due to training, time and resource limitations.

SDTM - Study Data Tabulation Model
IDDO - Infectious Disease Data Observatory
LMICs - Low- and Middle-Income Countries

Objectives

- Creation of an **open-source R package** to convert SDTM data **into analysis datasets**, using IDDO implementation.
- Reduce the amount of duplicated work and **prompt reproducible outputs**.
- The effectiveness of packages like {admiral} (the Pharmaverse) are limited since our implementation is not compatible, and our audience are non-regular SDTM users.
- **Provide additional features** to improve data use, such as recalculating **BMI** results and including Child **Growth Standards**.

BMI – Body Mass Index

Process for PREP functions

- PREP_DM, PREP_LB_BL, PREP_VS_FU are examples of functions in the package.
- For each domain, they **convert blanks to NA and results to upper case, character** class.
- **Filter** variables of interest.
- **Amalgamate data** using standardised results or decoded terms, and where NAs exist fill with modified or original results.
- **Pivot the domain wider**, transforming the long data to a wide data format, with tests or terms as columns.
- **Clean** column names.

https://www.cdisc.org/sites/default/files/2024-04/CDISC_EI_2024-Poster.pdf

Benefits to Academic Groups

Why use CDISC standards?

Enhanced Data Quality, Reliability, and Reproducibility

- **STANDARDIZATION**: The use of CDISC standards ensures data is collected, structured, and stored in a consistent manner
  - For example: A multi-center trial studying the effects of a new drug can maintain uniform data formats and definitions across all sites, ensuring high-quality and comparable data and reducing errors and inconsistencies in data collection

- **QUALITY CONTROL**: The use of CDISC standards enables robust validation checks and tools for quality control
  - For example: CDISC and the CDISC user community supports a wide variety of open-source tools (e.g., CDISC Library, R, Python, GitHub repositories) to check both quality and conformance of CDISC implementation, enabling groups to customize and use these solutions without a need for significant financial or human resource investment

Increased Data Sharing and Reuse Capabilities

- **REUSABILITY:** CDISC standards support the reuse of data internally for secondary purposes such as further research, audits, and quality control
  - For example, SDTM data from several completed studies on cardiovascular disease can be easily pooled for a secondary analysis on long-term effects and safety, without the need for extensive reformatting of individual datasets
- **COLLABORATION**: Using CDISC standards facilitates data sharing and integration across different systems and organizations
  - For example, research groups using CDISC can easily share their dataset with another institution for a meta-analysis, knowing the data is consistently formatted, structured, and described, facilitating understanding and use of the data

Improved Efficiency in Data Collection and Analysis Pipelines

- **COST SAVINGS**: Standardized data collection and reporting reduce the time and resources needed for both data management and analysis
  - For example: data following consistent and expected formats can be managed and analyzed with greater automation, reducing the costs associated with custom data handling, formatting, and manual cleaning
- **ACCELERATED TIMELINES**: Streamlined data management and analysis processes allow for shortened conception-implementation timelines
  - For example: An internal “library” of annotated CRFs, data capture systems, validation checks, and analysis code templates can be reused and quickly customized and implemented for each trial

Supporting Academics in CDSIC

How can I help?

CDISC Interchanges

- Attend the Interchanges!
- Submit abstracts for the Interchanges!
- Use cases and presentations from academic groups are essential to share knowledge and advance the use of CDISC standards by the academic community

https://www.cdisc.org/partnerships/coordinating-committees/europe

Thank You!