

# The leDEA Data Exchange Standard: a common data model for global HIV cohort collaboration

Stephany N. Duda, Beverly S. Musick, Mary-Ann Davies, Annette H. Sohn, Bruno Ledergerber, Kara Wools-Kaloustian, Catherine C. McGowan, Nicola J. Maxwell, Azar Kariminia, Cam Ha Dao Ostinelli, Brenna C. Hogan, Qiuhi Shi, Karen Malateste, Ruth L. Goodall, Dennis Karsten Kristensen, Erik V. Hansen, Carolyn F. M. Williams, Judith T. Lewis, Constantin T. Yiannoutsos

SND: Dr. Stephany N. Duda, Vanderbilt University School of Medicine, Nashville, TN, USA.

[stephany.duda@vanderbilt.edu](mailto:stephany.duda@vanderbilt.edu)

BSM: Ms. Beverly S. Musick, Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, USA.

[bsmusick@iu.edu](mailto:bsmusick@iu.edu)

MAD: Dr. Mary-Ann Davies, School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa. [mary-ann.davies@uct.ac.za](mailto:mary-ann.davies@uct.ac.za)

AHS: Dr. Annette H. Sohn, TREAT Asia/amfAR – The Foundation for AIDS Research, Bangkok, Thailand.

[annette.sohn@treatasia.org](mailto:annette.sohn@treatasia.org)

BL: Dr. Bruno Ledergerber, Division of Infectious Diseases and Hospital Epidemiology, University of Zurich, Zurich, Switzerland. [bruno.ledergerber@uzh.ch](mailto:bruno.ledergerber@uzh.ch)

KKW: Dr. Kara Wools-Kaloustian, Division of Infectious Diseases, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA. [kwools@iu.edu](mailto:kwools@iu.edu)

CCM: Dr. Catherine C. McGowan, Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA. [c.mcgowan@vumc.org](mailto:c.mcgowan@vumc.org)

NM: Ms. Nicola J. Maxwell, Centre for Infectious Disease Epidemiology and Research, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa. [Nicky.maxwell@uct.ac.za](mailto:Nicky.maxwell@uct.ac.za)

AK: Dr. Azar Kariminia, The Kirby Institute, UNSW Sydney, Australia. [Akariminia@kirby.unsw.edu.au](mailto:Akariminia@kirby.unsw.edu.au)

CDO: Ms. Cam Ha Dao Ostinelli, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland.

[camha.dao@ispm.unibe.ch](mailto:camha.dao@ispm.unibe.ch)

BCH: Ms. Brenna C. Hogan, Johns Hopkins Bloomberg School of Public Health, USA. [bhogan7@jhu.edu](mailto:bhogan7@jhu.edu)

QS: Dr. Qiuhi Shi, Department of Public Health, New York Medical College, Valhalla, NY, USA. [qiuhi\\_shi@nymc.edu](mailto:qiuhi_shi@nymc.edu)

KM: Ms. Karen Malateste, University of Bordeaux, Inserm, French National Research Institute for Sustainable Development (IRD), UMR 1219, Bordeaux, France. [karen.malateste@u-bordeaux.fr](mailto:karen.malateste@u-bordeaux.fr)

RG: Dr. Ruth L. Goodall, Medical Research Council Clinical Trials Unit at University College London, London, UK.

[r.goodall@ucl.ac.uk](mailto:r.goodall@ucl.ac.uk)

DKK: Mr. Dennis Karsten Kristensen, CHIP, Department of Infectious Diseases, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. [dennis.karsten.kristensen@regionh.dk](mailto:dennis.karsten.kristensen@regionh.dk)

EVH: MSc Eng. Erik V. Hansen, CHIP, Department of Infectious Diseases, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. [erik.viuff.hansen@Regionh.dk](mailto:erik.viuff.hansen@Regionh.dk)

CFMW: Dr. Carolyn F. M. Williams, Epidemiology Branch, National Institutes of Allergy and Infectious Diseases, Division of AIDS (NIAID/DAIDS), Rockville, Maryland, USA. [cwilliams@niaid.nih.gov](mailto:cwilliams@niaid.nih.gov)

JTL: Dr. Judith T. Lewis, Vanderbilt Institute of Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA. [judy.lewis@vumc.org](mailto:judy.lewis@vumc.org)

CTY: Dr. Constantin T. Yiannoutsos, Department of Biostatistics, Indiana University Fairbanks School of Public Health, Indianapolis, IN, USA. [cyiannou@iu.edu](mailto:cyiannou@iu.edu)

## Corresponding author:

Stephany N. Duda, PhD

Department of Biomedical Informatics

Vanderbilt University School of Medicine

2525 West End Ave, Suite 1050

Nashville, TN USA

[Stephany.duda@vanderbilt.edu](mailto:Stephany.duda@vanderbilt.edu)

**Keywords:** HIV, common data model, international research, data harmonization, data sharing, standards

## **Abstract**

### *Objective*

To describe content domains and applications of the leDEA Data Exchange Standard, its development history, governance structure, and relationships to other established data models, as well as to share open source, reusable, scalable, and adaptable implementation tools with the informatics community.

### *Methods*

In 2012, the International Epidemiology Databases to Evaluate AIDS (leDEA) collaboration began development of a data exchange standard, the leDEA DES, to support collaborative global HIV epidemiology research. With the HIV Cohorts Data Exchange Protocol as a template, a global group of data managers, statisticians, clinicians, informaticians, and epidemiologists reviewed existing data schemas and clinic data procedures to develop the HIV data exchange model. The model received a substantial update in 2017, with annual updates thereafter.

### *Findings*

The resulting leDEA DES is a patient-centric common data model designed for HIV research that has been informed by established data models from US-based electronic health records, broad experience in data collection in resource-limited settings, and informatics best practices. The leDEA DES is inherently flexible and continues to grow based on the ongoing stewardship of the leDEA Data Harmonization Working Group with input from external collaborators. Use of the leDEA DES has improved multiregional collaboration within and beyond leDEA, expediting over 95 multiregional research projects using data from more than 400 HIV care and treatment sites across seven global regions. A detailed data model specification and REDCap data entry templates that implement the leDEA DES are publicly available on GitHub.

### *Conclusions*

The leDEA common data model and related resources are powerful tools to foster collaboration and accelerate science across research networks. While currently directed towards observational HIV research and data from resource-limited settings, this model is flexible and extendable to other areas of health research.

## Highlights

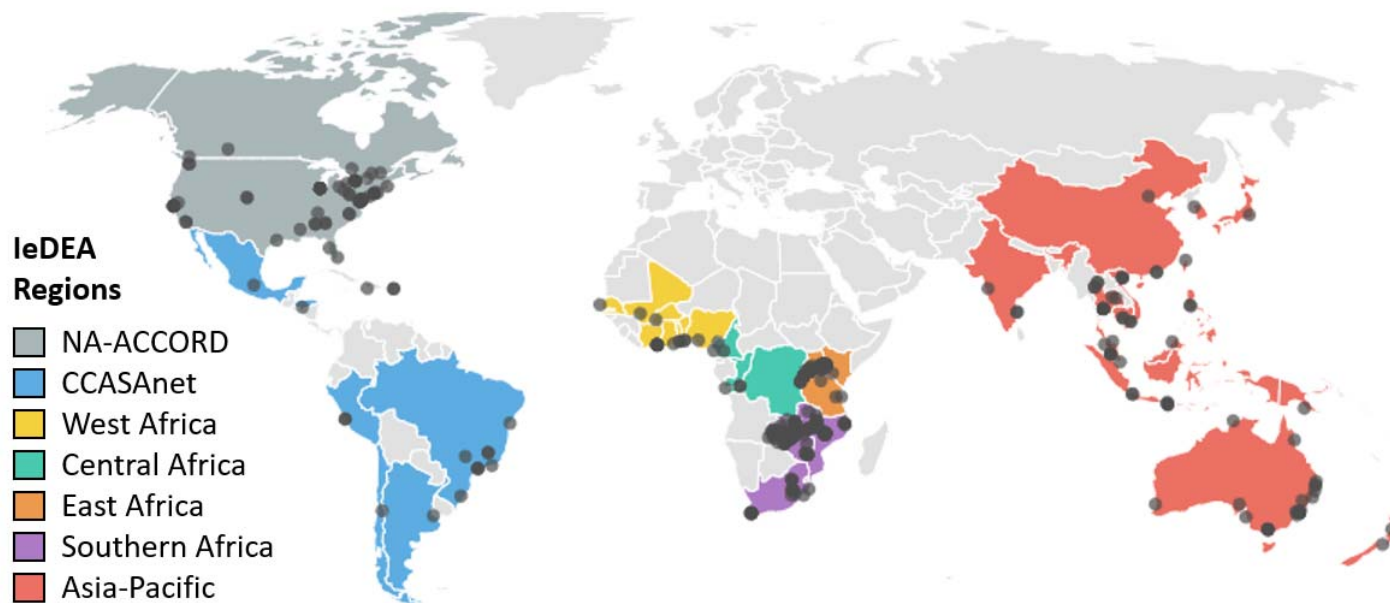
- The leDEA Data Exchange Standard is a data model for HIV epidemiology research.
- The model has expedited 95 projects using data from >400 HIV clinics worldwide.
- A browsable and adaptable version and data collection templates are available online.

## Background and Significance

Shared standards for representing and exchanging data are essential in collaborative research, particularly in the area of HIV/AIDS where multi-site, multi-national, and multiregional collaborations study the global epidemic. To facilitate this work and increase the efficiency of data sharing, collaborating researchers must exchange data using common data models (CDMs), which standardize data tables and variable names, matched code lists, data definitions, and formatting guidelines. Researchers have designed generalized CDMs that capture the broad range of clinical data generated by electronic health record systems and enable rapid data aggregation and distributed analyses across sites [1–4]. Despite their flexibility, generalized data models are not suited to all types of research collaborations: they lack dedicated variables for disease-specific concepts and may not account for varying levels of precision in clinical data obtained in particular in resource-limited settings. Realizing both the advantages of standardized data sharing and the extensive challenges of collecting, harmonizing, processing, validating, and analyzing HIV-specific data obtained from diverse sources, we developed a robust common data model for HIV data exchange that addresses the local context of data collection and the needs of sites from diverse geographic and resource backgrounds.

The International Epidemiology Databases to Evaluate AIDS (IeDEA) consortium is a collaboration of seven regional HIV observational research networks in North America, Latin America and the Caribbean, the Asia-Pacific, and four regions in sub-Saharan Africa (**Figure 1**) [5]. IeDEA was launched in 2006 with support from the U.S. National Institutes of Health (NIH) to conduct global research to address high priority HIV/AIDS research questions [6]. Each of the seven IeDEA regions collates data in a variety of different formats, from participating HIV clinics in its geographic area and pools them in longitudinal databases. In most cases, these data are derived from routine patient care and include demographics, clinical visit measurements such as height and weight, HIV-related diagnoses and laboratory values, plus dosages and dates of antiretroviral and other medications. Data sources include electronic health record systems (e.g., OpenMRS, local or proprietary systems)[7,8], research databases (e.g., MS Access, SPSS, REDCap)[9–11], and paper clinic charts. Regional Data Centers are responsible for receiving data from participating HIV clinics, combining the data into a regional database for local and regional analysis, conducting data quality checks, and sending data to other IeDEA

regions to be combined for multiregional research. NIH developed this federalized structure to build cultures of trust where data sharing was initiated close to the point of data creation and researchers could support constituent countries' research needs, thereby increasing the collaboration's flexibility and inclusiveness. This process has been refined over time, resulting in data from almost two million people living with HIV currently contained in regional databases.



**Figure 1:** Map of the seven leDEA regions and participating HIV clinics.

Overcoming barriers to harmonizing repositories of observational HIV data was the principal motivation behind leDEA's interest in common data models for HIV epidemiology. At the beginning of leDEA, the consortium had no common data exchange standard, leading to excessive time and resources spent on reformatting, cleaning, and merging data for multiregional research. To address this challenge, our team of regional data scientists, clinicians, and analysts decided to create an easy-to-apply common data standard to facilitate HIV data exchange. Here we recount the development of the leDEA Data Exchange Standard (leDEA DES), outline the challenges and decisions behind its creation, and describe the evolving products of this global HIV data collaboration.

## Methods

### *Development*

During the July 2012 XIX International AIDS Conference in Washington, DC, leDEA held a Data Harmonization Stakeholders' Meeting to work towards a data standard to support global HIV epidemiology research. Attendees included over 30 clinicians, epidemiologists, informaticians, data managers, and project managers engaged in leDEA research, along with representatives from leDEA funding agencies. Meeting participants reviewed other available data models, including early versions of the US-based Observational Medical Outcomes Partnership (OMOP) common data model [1], the i2b2 (Informatics for Integrating Biology and the Bedside) data schema [12], and the European HIV Cohorts Data Exchange Protocol (HICDEP). HICDEP was designed in 2003 by European HIV cohorts [13] and further developed within EuroCoord, the HIV research network funded by the European Union specifically to support collaboration among European HIV cohorts [14]. At the time it was the only HIV-specific data model explicitly designed for research data sharing.

After a detailed variable-level comparison with data arising from leDEA regions, the meeting participants advised against unconditional adoption of any existing data model. This arose from the concern that all the generalizable CDMs under consideration required complex coding and data formatting that would require substantial additional resources for data preparation, and the resulting normalized data structure was challenging for clinicians and analysts to interpret. Of all data models reviewed, HICDEP was most in line with anticipated leDEA data management needs, but lacked sufficient flexibility in the variable definitions and coding structure. For example, at the time HICDEP did not include variables for pediatric patients or tuberculosis, and had a limited number of pregnancy-related variables. Therefore, leDEA decided to fork the HICDEP standard by adopting core HICDEP tables and conventions while substantially expanding the variables, tables, and code lists to address broader data needs in leDEA. A working group was assembled to define an leDEA common data model and data exchange procedures. During a 6-month design process, leDEA investigators consistently favored a data model that met practical, on-the-ground needs of local research teams and data managers over more abstracted data models.

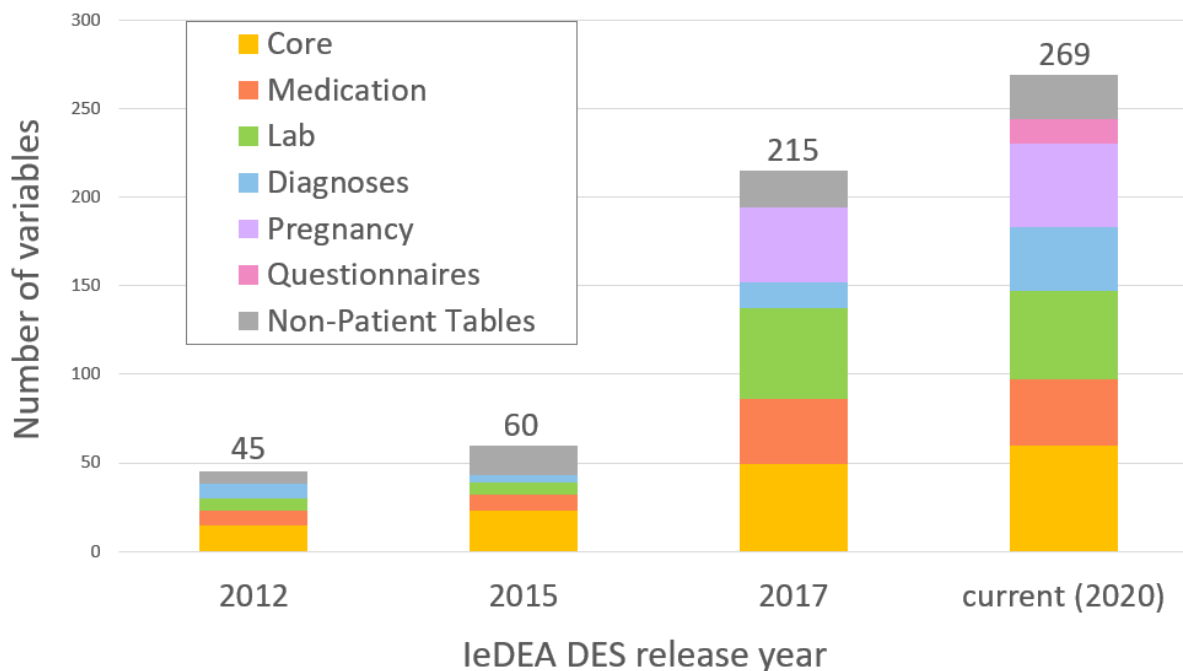
The first version of the leDEA Data Exchange Standard (leDEA DES) was released and implemented for all leDEA projects in 2012. However, the growing volume of multiregional collaborative work involving cross-domain HIV studies and external scientific collaborations necessitated a more complex data model. In 2015, we began an intensive effort to expand the leDEA DES as part of the NIH-funded Big Data to Knowledge (BD2K) program. New data domains were identified through a review of leDEA research proposals that requested data which did not exist in the previous version of the data model and through consultations with experts who helped identified data needs for emerging research priorities. All cadres of leDEA data creators and users participated in extending the data model. Clinicians helped inform the semantics of variables, regional data managers standardized the variable definitions and syntax, and epidemiologists and statisticians provided insights into efficient data representations for analyses. Members of HICDEP and other European HIV cohorts also participated. The resulting major revision of the leDEA DES was released in February 2017, with ongoing annual updates as of 2020.

## Results

### *Content coverage*

From an initial 7 tables and 45 variables in its original 2012 version, the leDEA Data Exchange Standard has evolved to 29 tables and 269 variables in its 2020 version (**Figure 2**). The leDEA DES covers core domains for HIV epidemiology research, including basic demographics, anthropometrics, HIV disease staging, clinic visits, medications, laboratory measurements, clinical diagnoses, hospitalizations, pregnancy data, and the results of screening and diagnostic patient questionnaires.





**Figure 2:** Categories and total counts of variables in the leDEA Data Exchange Standard by major version release year

Initial patient data, including the patient identifier, birth date, sex at birth, probable route of HIV infection, and HIV clinic enrollment date, are stored in a required “basic” table (`tblBAS`), which is paired with an optional patient follow-up table that includes date of loss to follow-up from clinic, last date known alive, and date of death (`tblLTFU`). Most other patient tables, such as laboratory testing, medication, and visit tables, can contain multiple rows per patient. Through BD2K support, we added tables on pregnancy and newborn care, allowing linkage of mother-child pairs when both are enrolled in the cohort, or facilitate capture of mother-centric or child-centric data when only one member of the mother-child dyad is enrolled. Other expansions included variables for HIV status disclosure, orphan status and caregiver information for children, reasons for starting antiretroviral therapy (ART), drug resistance testing, and cancer diagnoses.

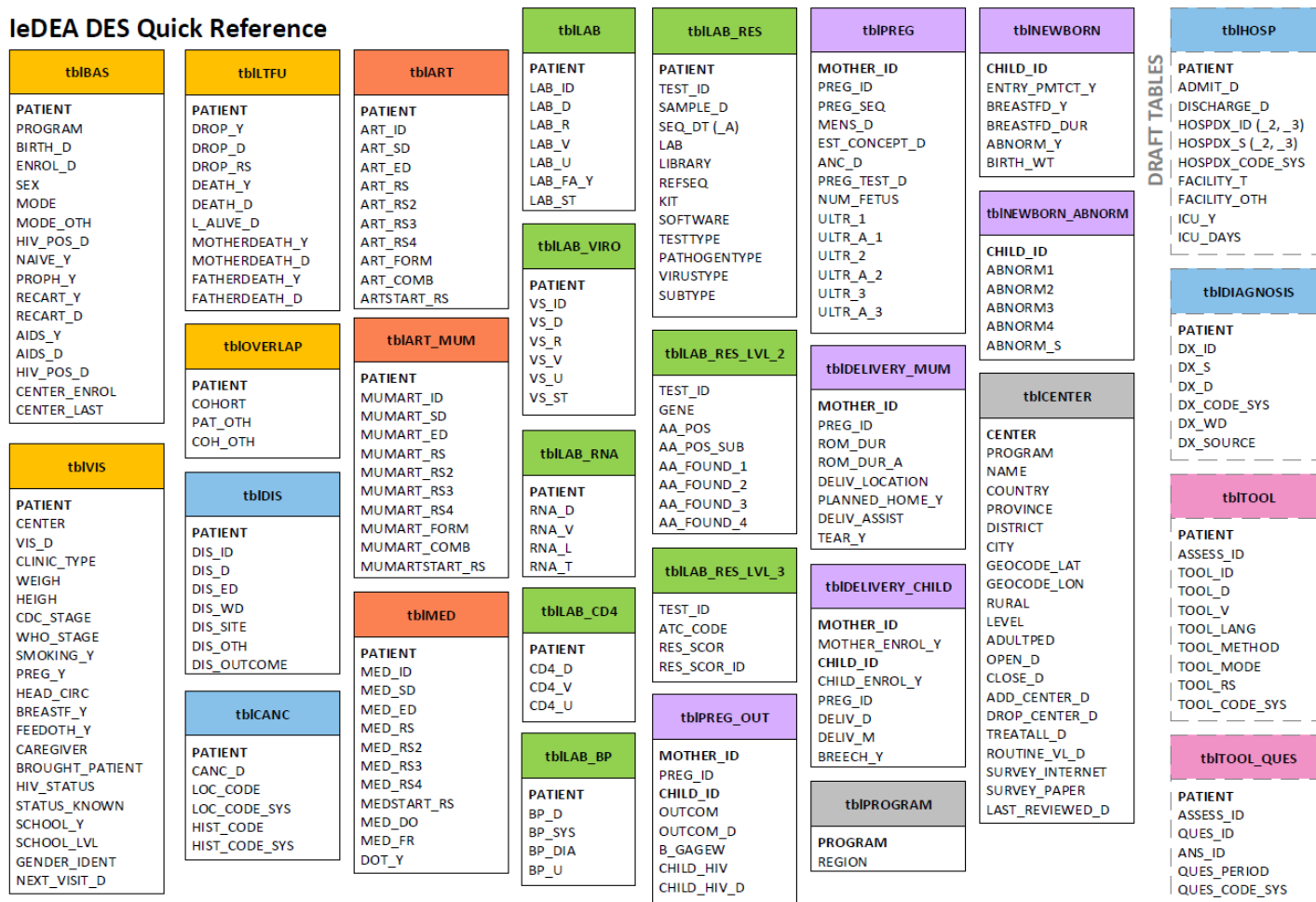
**Figure 3** depicts a diagram of the tables and variables of the leDEA DES data model.

Two non-patient tables, `tblCENTER` and `tblPROGRAM`, focus on data about the location of care. `tblCENTER` captures the name and location of the HIV care and treatment center, clinic population, level of care (e.g., primary to

tertiary), community setting (e.g., urban or rural), and the opening and closure dates of the research dataset.

tblPROGRAM links multiple HIV clinics that are part of the same health care program or organization.

### leDEA DES Quick Reference



**Figure 3:** Diagram of tables and variables for the leDEA DES data model. "Table Basic" (tblBAS), shown top left, is the index table containing records for all patients in the cohort. All date variables (*variablename\_D*) have optional date approximation variables (*variablename\_D\_A*) (date approximation variables not shown). Updated versions and variable definitions are posted at [jedeades.org](http://jedeades.org).

### Design for HIV data from resource-limited settings

A key innovation of the leDEA DES is that it supports the exchange of core HIV care data from a variety of settings ranging from a rural health clinic in South Africa, to a public hospital in Peru, to an academic medical center in Canada.

The database structure is straightforward, reflects real-world clinical concepts (e.g., labs, visits, medications), and is not fully normalized, allowing it to be manipulated with limited computing resources by non-experts. Code lists include inexpensive point-of-care/rapid laboratory tests often used in resource-limited settings. Every patient-related date variable in the leDEA DES has an optional “date approximation” field that allows data managers to specify uncertainty in the accuracy of a date. Such flexibility in defining variables, is critical for representing data extracted from clinics with less detailed medical records. Patient visits are mapped to clinics in the visits table (`tblVIS`), allowing HIV care and treatment programs to track patients who visit multiple leDEA sites as part of routine care.

To accommodate multiple types of local data management software across leDEA contributing sites, the leDEA DES does not define a file format for central data transfer. DES data files can be exchanged using common formats such as CSV, generic database formats such as SQL, or proprietary data formats like SAS and Stata, depending on the needs of the data recipients and the platforms available to the sender.

### *Naming conventions*

leDEA adopted many core HIV variables from HICDEP along with their naming conventions. Table names are prefixed with “tbl” to distinguish tables from variables. Variable suffixes indicate the type and format of the variable. **Table 1** lists the most frequently used suffixes as formalized in the leDEA DES. These naming conventions are used to automate basic data quality checks on leDEA datasets. For example, if the binary variable for AIDS diagnosis is coded “No” (`AIDS_Y=0`), then there should be no date in the associated `AIDS_D` field because the patient has not been diagnosed with AIDS.

**Table 1:** Variable name suffixes used in the leDEA DES to designate data type and content

Variable Suffix	Variable Meaning
<b>_Y</b>	Yes/No/Unknown
<b>_D</b>	Date in YYYY-MM-DD format
<b>_D_A</b>	Date approximation code
<b>_ID</b>	Identifier (except PATIENT)
<b>_OTH</b>	Descriptive field for “other (specify)”
<b>_S</b>	Descriptive field (string) for original data
<b>_R</b>	Result (of non-numeric test)
<b>_V</b>	Value (of numeric test result)
<b>_T</b>	Type (of facility, laboratory equipment)
<b>_U</b>	Unit of measurement code
<b>_RS</b>	Reason (for medication change, loss to follow-up, death)

### *Use of terminologies*

The leDEA DES supports the use of standardized terminologies distributed by the World Health Organization (WHO), as these coding systems have no licensing fees and are most likely to be used globally and in resource-limited settings. Medications in the antiretrovirals table (`tblART`), general medications table (`tblMED`), and drug resistance table (`tblLAB_RES_LVL_3`) are coded according to the Anatomical Therapeutic Chemical (ATC) Classification System maintained by WHO [15]. Diagnoses in the hospitalizations and diagnoses tables (`tblHOSP`, `tblDIAGNOSIS`) are coded using WHO’s International Statistical Classification of Diseases and Related Health Problems (ICD) [16]. The leDEA Data Harmonization Working Group provides suggested subsets of ICD-10 and ATC codes to simplify coding, but other coding systems can be used if defined in a table's coding system (`_CODE_SYS`) variables. The official lists of AIDS-defining disease diagnoses (`tblDIS`) specified by the WHO and the U.S. Centers for Disease Control and Prevention are represented using an internally managed list of 2-to-4-character strings [17,18]. This concession maintains historic compatibility with HICDEP (for CDC codes) and facilitates mapping of this critical HIV care data, as the WHO list of AIDS conditions does not map one-to-one to ICD codes. International Organization for Standardization (ISO) standards are used for date formats, country codes, and language codes [19–21].

## *Governance*

The leDEA DES is maintained by the Data Harmonization Working Group, which meets monthly to assess new research proposals and their data needs. Upkeep of the DES requires approximately four hours per month, including collating new variable proposals and revising them to conform to naming conventions, researching other data models, and communicating with domain experts. During an annual cross-consortium data meeting, working group members discuss proposed leDEA DES additions with the European HICDEP team and consider new HICDEP variables for inclusion. Backwards compatibility is preserved for all coding lists, though old variables may be deprecated with advance notice if needed to maintain consistency of the data model. A formal draft revision is prepared annually by the DHWG and circulated to all leDEA regions for review. The revisions are presented at an leDEA-wide meeting for approval by the leDEA Executive Committee before release.

## *Availability*

The leDEA DES is openly available for use and adaptation by other research organizations. Documentation of the leDEA DES is provided at <http://ledeades.org> and recorded on fairsharing.org and the NIAID resources page [22,23]. To support use of the leDEA DES for practical data collection, the version of the leDEA DES described here has been implemented as a set of data collection forms for the REDCap platform. These REDCap data dictionaries, which were developed for REDCap version 8.2.0 and above, are posted on GitHub at <https://github.com/leDEA> in CDISC ODM-XML format, along with implementation guidelines and a REDCap data dictionary for the communally maintained tblCENTER.

## *Impact*

Data from more than 400 HIV care and treatment clinics participating in leDEA have been successfully mapped to the DES for multiregional data exchange. Since its launch in 2012, the DES has facilitated data exchange for more than 95 projects and 49 publications across leDEA regions and with other external collaborators such as the World Health Organization and the Joint United Nations Programme on HIV/AIDS (UNAIDS)[24–26]. Multiregional leDEA data also have been used by other research groups, including the Collaborative Initiative for Paediatric HIV Education and Research (CIPHER)[27–29], the Collaboration of Observational HIV Epidemiological Research Europe Study (COHERE)

[30,31], and the ALPHA network [32]. leDEA also contributes data and scientific expertise to mathematical modeling consortia such as the Cost-Effectiveness of Preventing AIDS Complications (CEPAC) modeling group [33,34] and the UNAIDS Reference Group on Estimates, Modelling and Projections [35–40]. Clear benefits of the leDEA DES have included substantial increases in data quality observed by data managers and analysts, tangible time savings from standardization when performing repeated data processing, and increased collaboration and perceived added value among participating organizations. Three leDEA regions have adopted the leDEA DES as their principal data storage format and the European HICDEP collaboration has incorporated most of leDEA’s additions into its own data model.

## **Discussion**

The leDEA DES is an accessible data standard with an active community of contributors, a proven tool to improve efficiency and facilitate global HIV research collaborations, and a resource for meaningful multiregional collaboration within and beyond leDEA. Its use in leDEA has facilitated numerous projects, publications, and collaborations. leDEA chose to develop a patient-centric, disease-centric data model rather than using a generalized CDM to reduce resource utilization. Mapping clinical data to a generalized CDM and extracting them for analysis often involves extensive data abstraction, transformation, and recoding steps and consequently more personnel time, data management expertise, and data infrastructure. Our approach towards common data model adoption reflects our “on the ground” experience with data challenges in global collaborations.

Plans for future growth of the leDEA DES include developing new tables and variables to better represent patient antiretroviral treatment adherence, entry points into care (e.g., voluntary counseling and testing, antenatal clinics), and participation in research trials, as well as HIV care program characteristics. As HIV care is transformed from acute infectious disease management to chronic disease care, the leDEA DES will expand to include more tables and variables to cover non-communicable disease outcomes. The model is also being expanded to include tuberculosis disease and care. Expansion of the leDEA DES continues to be informed by common data models originally developed for electronic

health record data (e.g., Observational Medical Outcomes Partnership (OMOP) CDM, PCORnet Distributed Research Network CDM [1–3]), as well as electronic health record database schemas designed for HIV care, such as the Academic Model Providing Access to Healthcare (AMPATH) Medical Record System schema and the HIV minimum dataset proposed by Tierney et al. [41,42].

The structure of the leDEA DES allows for the development of research support software, and we are currently developing software platforms where leDEA DES data can be loaded, quality-checked, and summarized. Through the continuous expansion and improvement of its data domains, and because of its proven utility in worldwide collaborations and accessibility, the leDEA DES is positioned as a complete solution for HIV-specific data collection, definition, and storage in the “big epidemiology data” context.

## **Conclusions**

The leDEA DES is an accessible data exchange standard that continues to be expanded based on feedback from global collaborators and shaped by evolving HIV epidemics and new research interests. The leDEA data model can be leveraged and adapted by other research collaborations seeking to implement a defined structure for disease-specific, patient-centric research data exchange. The ongoing data standardization efforts promoted by leDEA and the increased awareness of the benefits of global data harmonization are intended to meaningfully strengthen collaborations studying the impact of the global response to the HIV epidemic.

## Acknowledgments

The authors would like to acknowledge the support of the members of the leDEA Data Harmonization Working Group, the leDEA Executive Committee, the HIV Cohorts Data Exchange Protocol collaboration, and other groups that have contributed to the development of the leDEA DES over time. The authors also thank Dr. Pernille Iversen, Mr. Jesper Kjaer, Mr. Larry Riggen, Dr. Rosemary McKaig, and Dr. Firas Wehbe for their support in the original creation of the leDEA Data Exchange Standard.

*Funding:* The International Epidemiology Databases to Evaluate AIDS (leDEA) is supported by the U.S. National Institutes of Health's National Institute of Allergy and Infectious Diseases, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, the National Cancer Institute, the National Institute of Mental Health, the National Institute on Drug Abuse, the National Heart, Lung, and Blood Institute, the National Institute on Alcohol Abuse and Alcoholism, the National Institute of Diabetes and Digestive and Kidney Diseases, the Fogarty International Center, and the National Library of Medicine: Asia-Pacific, U01AI069907; CCASAnet, U01AI069923; Central Africa, U01AI096299; East Africa, U01AI069911; NA-ACCORD, U01AI069918; Southern Africa, U01AI069924; West Africa, U01AI069919. Informatics resources are supported by the Harmonist project, R24AI124872. EuroCoord was funded through the European Union Seventh Framework program (FP7/2001-2013) under the grant agreement no. 260694. This work is solely the responsibility of the authors and does not necessarily represent the official views of any of the institutions mentioned above.



## References

- [1] E.A. Voss, R. Makadia, A. Matcho, Q. Ma, C. Knoll, M. Schuemie, F.J. DeFalco, A. Londhe, V. Zhu, P.B. Ryan, Feasibility and utility of applications of the common data model to multiple, disparate observational health databases, *J. Am. Med. Inform. Assoc. JAMIA*. 22 (2015) 553–564. <https://doi.org/10.1093/jamia/ocu023>.
- [2] R.L. Fleurence, L.H. Curtis, R.M. Califf, R. Platt, J.V. Selby, J.S. Brown, Launching PCORnet, a national patient-centered clinical research network, *J. Am. Med. Inform. Assoc. JAMIA*. 21 (2014) 578–582. <https://doi.org/10.1136/amiainl-2014-002747>.
- [3] PCORnet Common Data Model (CDM), (n.d.). <https://pcornet.org/data-driven-common-model/> (accessed June 30, 2020).
- [4] Sentinel Common Data Model | Sentinel Initiative, (n.d.). <https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model> (accessed November 13, 2019).
- [5] leDEA International epidemiology Databases to Evaluate AIDS, (n.d.). <https://www.iedea.org/> (accessed January 31, 2019).
- [6] RFA-AI-05-014: International Epidemiologic Databases to Evaluate AIDS (IEDEA), (n.d.). <http://grants.nih.gov/grants/guide/rfa-files/RFA-AI-05-014.html> (accessed September 7, 2015).
- [7] B.W. Mamlin, P.G. Biondich, B.A. Wolfe, H. Fraser, D. Jazayeri, C. Allen, J. Miranda, W.M. Tierney, Cooking Up An Open Source EMR For Developing Countries: OpenMRS – A Recipe For Successful Collaboration, *AMIA. Annu. Symp. Proc. 2006* (2006) 529–533.
- [8] A. Marcelin, C. Perodin, C. Baja, A. Bright, J. Duperval, M. Duplan, F. Dérilus, S. Duda, J. Pape, Developing an Electronic Medical Record for Interlinked Care Services in Haiti, *Stud. Health Technol. Inform.* 216 (2015) 883.
- [9] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, J.G. Conde, Research Electronic Data Capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support, *J. Biomed. Inform.* 42 (2009) 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>.
- [10] Microsoft Access, (n.d.). <https://www.microsoft.com/Microsoft/Access>.
- [11] SPSS Software, (2019). <https://www.ibm.com/analytics/spss-statistics-software> (accessed November 13, 2019).
- [12] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *J. Am. Med. Inform. Assoc. JAMIA*. 17 (2010) 124–130. <https://doi.org/10.1136/jamia.2009.000893>.
- [13] J. Kjaer, B. Ledergerber, HIV cohort collaborations: proposal for harmonization of data exchange, *Antivir. Ther.* 9 (2004) 631–633.
- [14] HIV Cohorts Data Exchange Protocol Wiki, (n.d.). <https://hicdep.org/> (accessed November 13, 2019).
- [15] WHOCC - ATC/DDD Index, (n.d.). [https://www.whocc.no/atc\\_ddd\\_index/](https://www.whocc.no/atc_ddd_index/) (accessed November 13, 2019).
- [16] WHO | International Classification of Diseases, WHO. (n.d.). <http://www.who.int/classifications/icd/en/> (accessed November 13, 2019).
- [17] World Health Organization, WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children., World Health Organization, Geneva, 2007. <https://www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf> (accessed November 12, 2019).
- [18] U.S. Centers for Disease Control and Prevention, 1993 Revised Classification System for HIV Infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults, (n.d.). <http://www.cdc.gov/mmwr/preview/mmwrhtml/00018871.htm> (accessed November 13, 2019).
- [19] International Organization for Standardization, ISO 8601 Date and time format, ISO. (n.d.). <http://www.iso.org/cms/render/live/en/sites/isoorg/home/standards/popular-standards/iso-8601-date-and-time-format.html> (accessed November 13, 2019).
- [20] International Organization for Standardization, ISO 3166 Country Codes, ISO. (n.d.). <http://www.iso.org/cms/render/live/en/sites/isoorg/home/standards/popular-standards/iso-3166-country-codes.html> (accessed November 13, 2019).
- [21] International Organization for Standardization, ISO 639 Language codes, ISO. (n.d.). <http://www.iso.org/cms/render/live/en/sites/isoorg/home/standards/popular-standards/iso-639-language-codes.html> (accessed November 13, 2019).
- [22] FAIRsharing: bsg-s001423: leDEA DES, (n.d.). <https://fairsharing.org/bsg-s001423/> (accessed December 20, 2019).

- [23] International Epidemiologic Database to Evaluate AIDS Data Exchange Standard (IeDEA DES) | NIH: National Institute of Allergy and Infectious Diseases, (n.d.). <https://www.niaid.nih.gov/research/iedea-data-exchange-standard> (accessed January 3, 2020).
- [24] E. Zaniewski, O. Tymejczyk, A. Kariminia, S. Desmonde, V. Leroy, N. Ford, A.H. Sohn, D. Nash, M. Yotebieng, M. Cornell, K.N. Althoff, P.F. Rebeiro, M. Egger, IeDEA–WHO Research–Policy Collaboration: contributing real-world evidence to HIV progress reporting and guideline development, *J. Virus Erad.* 4 (2018) 9.
- [25] A. IeDEA, IeDEA Network Contributes Data Analyses to WHO for HIV Progress Report, *IeDEA Int. Epidemiol. Databases Eval. AIDS.* (2016). <https://www.iedea.org/2016/02/19/iedea-network-contributes-data-analyses-to-who-for-hiv-progress-report/> (accessed January 3, 2020).
- [26] N. Ford, M. Penazzato, M. Vitoria, M. Doherty, M.-A. Davies, E. Zaniewski, O. Tymejczyk, M. Egger, D. Nash, The contribution of observational studies in supporting the WHO “treat all” recommendation for HIV/AIDS, *J. Virus Erad.* 4 (2018) 5–8.
- [27] Collaborative Initiative for Paediatric HIV Education and Research (CIPHER) Global Cohort Collaboration, Incidence of switching to second-line antiretroviral therapy and associated factors in children with HIV: an international cohort collaboration, *Lancet HIV.* 6 (2019) e105–e115. [https://doi.org/10.1016/S2352-3018\(18\)30319-9](https://doi.org/10.1016/S2352-3018(18)30319-9).
- [28] Collaborative Initiative for Paediatric HIV Education and Research (CIPHER) Global Cohort Collaboration, The epidemiology of adolescents living with perinatally acquired HIV: A cross-region global cohort analysis, *PLoS Med.* 15 (2018) e1002514. <https://doi.org/10.1371/journal.pmed.1002514>.
- [29] CIPHER Global Cohort Collaboration, Inequality in outcomes for adolescents living with perinatally acquired HIV in sub-Saharan Africa: a Collaborative Initiative for Paediatric HIV Education and Research (CIPHER) Cohort Collaboration analysis, *J. Int. AIDS Soc.* 21 Suppl 1 (2018). <https://doi.org/10.1002/jia2.25044>.
- [30] J. Del Amo, I. Jarrín, IeDEA and COHERE in EuroCoord Cohort Collaboration, All-cause mortality after ART initiation in HIV-positive women from Europe, Sub-Saharan Africa and the Americas, *AIDS Lond. Engl.* (2019). <https://doi.org/10.1097/QAD.0000000000002399>.
- [31] IeDEA and COHERE Cohort Collaborations, Global Trends in CD4 Cell Count at the Start of Antiretroviral Therapy: Collaborative Study of Treatment Programs., *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* (2018). <https://doi.org/10.1093/cid/cix915>.
- [32] G. Reniers, M. Wamukoya, M. Urassa, A. Nyaguara, J. Nakiyingi-Miiro, T. Lutalo, V. Hosegood, S. Gregson, X. Gómez-Olivé, E. Geubbels, A.C. Crampin, A. Wringe, L. Waswa, S. Tollman, J. Todd, E. Slaymaker, D. Serwadda, A. Price, S. Oti, M.J. Nyirenda, D. Nabukalu, C. Nyamukapa, F. Nalugoda, O. Mugurungi, B. Mtenga, L. Mills, D. Michael, E. McLean, N. McGrath, E. Martin, M. Marston, S. Maquins, F. Levira, C. Kyobutungi, D. Kwaro, I. Kasamba, C. Kanjala, K. Kahn, C. Kabudula, K. Herbst, D. Gareta, J.W. Eaton, S.J. Clark, K. Church, M. Chihana, C. Calvert, D. Beguy, G. Asiki, S. Amri, R. Abdul, B. Zaba, Data Resource Profile: Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA Network), *Int. J. Epidemiol.* 45 (2016) 83–93. <https://doi.org/10.1093/ije/dyv343>.
- [33] J. Jesson, S. Desmonde, C.T. Yiannoutsos, G. Patten, K. Malateste, S.N. Duda, N. Kumarasamy, M. Yotebieng, M.-A. Davies, B. Musick, V. Leroy, A. Ciaranello, IeDEA, Weight-for-age distributions among children with HIV on antiretroviral therapy in the International epidemiology Databases to Evaluate AIDS (IeDEA) multiregional consortium, *BMC Res. Notes.* 13 (2020) 249. <https://doi.org/10.1186/s13104-020-05081-7>.
- [34] A. Ciaranello, Z. Lu, S. Ayaya, E. Losina, B. Musick, R. Vreeman, K.A. Freedberg, E.J. Abrams, L. Dillabaugh, K. Doherty, J. Ssali, C.T. Yiannoutsos, K. Wools-Kaloustian, Incidence of World Health Organization stage 3 and 4 events, tuberculosis and mortality in untreated, HIV-infected children enrolling in care before 1 year of age: an IeDEA (International Epidemiologic Databases To Evaluate AIDS) East Africa regional analysis, *Pediatr. Infect. Dis. J.* 33 (2014) 623–629. <https://doi.org/10.1097/INF.0000000000000223>.
- [35] J. Stover, K. Andreev, E. Slaymaker, C. Gopalappa, K. Sabin, C. Velasquez, J. Nakiyingi-Miiro, A. Crampin, T. Lutalo, K. Herbst, S. Gregson, M. Urassa, Updates to the Spectrum model to estimate key HIV indicators for adults and children, *AIDS Lond. Engl.* 28 (2014) S427–S434. <https://doi.org/10.1097/QAD.0000000000000483>.
- [36] C.T. Yiannoutsos, L.F. Johnson, A. Boulle, B.S. Musick, T. Gsponer, E. Balestre, M. Law, B.E. Shepherd, M. Egger, International Epidemiologic Databases to Evaluate AIDS (IeDEA) Collaboration, Estimated mortality of adult HIV-infected patients starting treatment with combination antiretroviral therapy, *Sex. Transm. Infect.* 88 Suppl 2 (2012) i33–43. <https://doi.org/10.1136/sextrans-2012-050658>.

- [37] L.F. Johnson, N. Anderegg, E. Zaniewski, J.W. Eaton, P.F. Rebeiro, G. Carriquiry, D. Nash, M. Yotebieng, D.K. Ekouevi, C.B. Holmes, J.Y. Choi, A. Jiamsakul, G. Bakoyannis, K.N. Althoff, A.H. Sohn, C. Yiannoutsos, M. Egger, International epidemiology Databases to Evaluate AIDS (IeDEA) Collaboration, Global variations in mortality in adults after initiating antiretroviral treatment: an updated analysis of the International epidemiology Databases to Evaluate AIDS cohort collaboration, *AIDS Lond. Engl.* 33 Suppl 3 (2019) S283–S294. <https://doi.org/10.1097/QAD.0000000000002358>.
- [38] S. Desmonde, F. Tanser, R. Vreeman, E. Takassi, A. Edmonds, P. Lumbiganon, J. Pinto, K. Malateste, C. McGowan, A. Kariminia, M. Yotebieng, F. Dicko, C. Yiannoutsos, M. Mubiana-Mbewe, K. Wools-Kaloustian, M.-A. Davies, V. Leroy, International Epidemiology Databases to Evaluate AIDS (IeDEA) Pediatric Working Group, Access to antiretroviral therapy in HIV-infected children aged 0-19 years in the International Epidemiology Databases to Evaluate AIDS (IeDEA) Global Cohort Consortium, 2004-2015: A prospective cohort study, *PLoS Med.* 15 (2018) e1002565. <https://doi.org/10.1371/journal.pmed.1002565>.
- [39] A.L. Ciaranello, Y. Chang, A.V. Margulis, A. Bernstein, I.V. Bassett, E. Losina, R.P. Walensky, Effectiveness of pediatric antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis, *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 49 (2009) 1915–1927. <https://doi.org/10.1086/648079>.
- [40] N. Anderegg, L.F. Johnson, E. Zaniewski, K.N. Althoff, E. Balestre, M. Law, D. Nash, B.E. Shepherd, C.T. Yiannoutsos, M. Egger, IeDEA, MeSH consortia, All-cause mortality in HIV-positive adults starting combination antiretroviral therapy: correcting for loss to follow-up., *AIDS Lond. Engl.* 31 Suppl 1 (2017) S31–S40. <https://doi.org/10.1097/QAD.0000000000001321>.
- [41] B.W. Mamlin, P.G. Biondich, AMPATH Medical Record System (AMRS): Collaborating Toward An EMR for Developing Countries, *AMIA. Annu. Symp. Proc.* 2005 (2005) 490–494.
- [42] W.M. Tierney, E.J. Beck, R.M. Gardner, B. Musick, M. Shields, N.M. Shiyonga, M.H. Spohr, Viewpoint: A Pragmatic Approach to Constructing a Minimum Data Set for Care of Patients with HIV in Developing Countries, *J. Am. Med. Inform. Assoc. JAMIA.* 13 (2006) 253–260. <https://doi.org/10.1197/jamia.M2005>.