Africa CDC Database

Technical Handbook

Dr Jonathan Epstein   Ava Sullivan   Shannon Ball

01 December 2024

Table of contents

1 Introduction

This is a handbook created as a how-to manual for the development and maintenance of Africa CDC's Research Database for the African continent.

1.1 About the database

EcoHealth Alliance is supporting the creation of a database of active infectious disease research activities and research scientists in Africa. The database contains information about scientific research being conducted on the African continent that has particular relevance to understanding, detecting and responding to zoonotic pathogens. EHA is building a detailed, searchable, and visual database (e.g. via a dashboard) populated with information about major (e.g. multi-year) and active One Health research projects that includes subject matter, duration, geographical locations, key personnel, and links to publicly available reports, preprints and publications relevant to zoonoses. Examples of research areas may include epidemiological (syndrome based or disease specific), ecological (studies looking at potential reservoirs or animal hosts for zoonotic pathogens and spillover into humans or livestock); basic science (e.g. virology, bacteriology; serology); clinical (vaccine or therapeutic trials); or sociological (e.g. behavioral risk assessments or behavioral intervention studies) or any other preliminary public health findings. A particularly important view within the database includes a directory of subject matter experts associated with research across the continent. This roster view can be used by public health practitioners to engage expert consultations as needed to support training activities, surveillance or outbreak response, for example.

Part I

Tools

2 Dolt

Dolt is a SQL database you can fork, clone, branch, merge, push and pull just like a Git repository. Connect to Dolt just like any MySQL database to run SQL queries. Use the command line interface to import CSV files, commit your changes, push them to a remote, or merge your teammate's changes.

All the commands you know from Git work exactly the same in Dolt. Git versions files, Dolt versions tables. It's like Git and MySQL had a baby.

Dolt is a version controlled database. Dolt is Git for Data. Dolt is a Versioned MySQL Replica.

2.1 Version Controlled Database

Dolt is a version controlled SQL database. Connect to Dolt just like any MySQL database to run SQL queries. Use Dolt system tables, functions, or procedures to access version control information and features.

2.2 Git for Data

Dolt is Git for data. Dolt matches the Git CLI exactly. When you would have run `git add`, you run `dolt add`. When you would have run `git commit`, you run `dolt commit`.

2.3 Versioned MySQL Replica

Dolt can be deployed as a Versioned MySQL Replica. Because Dolt is MySQL compatible, Dolt can be configured just like any other MySQL replica. A Dolt replica gives you features of a version controlled database without migrating from MySQL.

2.4 Installing Dolt

Dolt is extremely simple to install. Dolt is a single ~100 megabyte program. To install it, you download or compile that program and put it on your PATH. To install in specific operating systems, follow the instructions below:

2.4.1 Windows

winget

`winget` `install dolt`

Chocolatey

`choco` `install dolt`

Both `.msi` files and `.zip` files are available.

Scoop

`scoop` `install dolt`

MSI Files

The easiest way to install Dolt on Windows is to use the MSI files that are provided with each release. They can be found in the Assets section of every release. Grab the latest here.

`.zip` Archive

For those preferring to install Dolt manually a zipped archive is provided with the requisite executables. It can be found in assets along with the latest release.

2.4.2 macOS

Install Script

The download script for Linux can be used, as OSX is a `*nix` system. It will download the appropriate binary, and place it in `/usr/local/bin`:

```
sudo bash -c 'curl -L https://github.com/dolthub/dolt/releases/latest/download/install.sh | b
```

Homebrew

A Homebrew formula is available with with every release, so Mac users using Homebrew for package management can build Dolt from source with a single command:

```
$ brew install dolt
==> Downloading https://homebrew.bintray.com/bottles/dolt-0.18.3.catalina.bottle.tar.gz
==> Downloading from https://d29vzk4ow07wi7.cloudfront.net/c03cc532d5045fa090cb4e0f141883685c
######################################################################### 100.0%
==> Pouring dolt-0.18.3.catalina.bottle.tar.gz
  /usr/local/Cellar/dolt/0.18.3: 7 files, 56.9MB
```

which will install Dolt as follows:

```
$ ls -ltr $(which dolt)
lrwxr-xr-x  1 oscarbatori  admin  30 Aug 26 16:49 /usr/local/bin/dolt -> ../Cellar/dolt/0.18
```

MacPorts

On macOS, Dolt can also be installed via a community-managed port via MacPorts:

```
sudo port install dolt
```

2.4.3 Linux

For Linux users, an installation script is available that will detect your architecture, download the appropriate binary, and place in `/usr/local/bin`:

```
sudo bash -c 'curl -L https://github.com/dolthub/dolt/releases/latest/download/install.sh | s
```

The use of `sudo` is required to ensure the binary lands in your path.

3 DoltHub

DoltHub is a place to share Dolt databases. We host public data for free! DoltHub adds a modern, secure, always on database management web GUI to the Dolt ecosystem. Edit your database on the web, have another person review it via a pull request, and have the production database pull it to deploy.

3.1 What is DoltHub

DoltHub is GitHub for Dolt. DoltHub acts as a Dolt remote you can clone, push, pull and fetch from. DoltHub adds permissions, pull requests, issues, and forks to the Dolt ecosystem. Additionally, DoltHub has a modern SQL workbench built in so you can explore and change databases on the web.

3.2 Getting Started

DoltHub has many uses. We recommend getting started by sharing a database.

This documentation will walk you through discovering data on DoltHub, cloning a copy locally, making a change on a fork, and submitting a pull request to the original database.

3.3 DoltHub API

DoltHub has an API you can script against. The documentation covers:

1. Authentication
2. SQL API - Used to make read or write SQL queries to a DoltHub database
3. CSV API - Used to download CSV format files of DoltHub tables
4. Database API - Used to interact with DoltHub databases and pull requests
5. Hooks - Used to receive change events to your DoltHub databases

4 R

R is a free and open-source programming language and software environment for statistical computing and graphics. It was first developed in the early 1990s by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is now widely used by statisticians, data analysts, and researchers across various fields.

R provides a wide range of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and more. It also has a large and active community of users and developers who contribute to the development of new packages and extensions for the language.

R is popular in the field of data science, as it provides a powerful and flexible platform for analyzing and visualizing data. It can be used in conjuction with other tools such as Airtable through community-developed packages making it particularly well-suited for the current use-case of the RIG.

5 Airtable

> **i** Airtable not anymore used to host the database
>
> As of May 2024, the Africa CDC Database backend has shifted from Airtable to DoltHub. This chapter has been kept here to document how the database started and to show the history of its development.
> To read about DoltHub, see Chapter 2. To learn more about the use of dolt and DoltHub as the current backend for the database, read Chapter 6.

Airtable is a cloud-based software platform that allows users to create and manage databases, spreadsheets, and other types of organizational tools. It can be used for a variety of purposes, including project management, customer relationship management, inventory tracking, event planning, and much more.

One of the key features of Airtable is its flexible and customizable nature. Users can create and customize their own database structures, and can choose from a wide range of data types, including text, attachments, checkboxes, and more. This allows for a high degree of customization and adaptability to different use cases and workflows.

Airtable also offers a variety of collaboration features, including real-time syncing and commenting, as well as integrations with other popular tools such as Slack, Google Drive, and Trello. Additionally, Airtable has a robust API that allows developers to build custom integrations and applications on top of the platform.

Airtable's flexibility, customizability, and features that support collaboration along with its spreadsheet-style interface that is familiar to most that have used other spreadsheet software such as Microsoft Excel or Google Sheets are the key reasons why it was chosen to be the database tool for RIG.

Chapter 5 provides an overview of Airtable and good practices for designing data models in relational databases.

5.1 Key Terms

- **Workspace** - A collection of bases
- **Base** - A database. Each is identified by the Airtable API via a `base id`

- **Table** - A tabular data set within a base. Each table is identified by the Airtable API via the name of the table e.g. "Demo Table"
- **Record** - An individual cell within a table. Each record is identified by the Airtable API via a `record id`
- **Field** - A property of data in a table
- **Views** - A specific way of displaying a table. Default is grid.
- **Entity** - Something that either physically or logically exists whose properties are typically stored in a table and composed of data elements.
- **Element** - an attribute of a Entity (a field)

5.2 Security and Access Control

Airtable maintains physical and technological security as part of its ISO IEC 27001:2013 and SOC 2 compliance measures. Data are 256-bit encrypted when storing on the server and also when transferring data over the internet. To find vulnerabilities in their software, they run daily, weekly, and monthly scans on different components of their system and regularly commission external penetration tests. They also run a bug bounty program to help identify issues. Their data centers have fire detection and suppression systems, redundant power systems, and strict control for physical access. Because Airtable relies on Amazon Web Services (AWS) for its cloud infrastructure (the same providers used by previous EHA projects), data are geo-redundantly replicated in backups across multiple zones to increase data durability. They also have a team monitoring services at all times. Airtable employees are thoroughly vetted before hiring and continually trained on data protection best practices. Their workstations are secured by using full-disk encryption, automatic locking, and strong password requirements.

5.2.1 User and Administrator Security Features

5.2.1.1 Access Controls

Airtable provides database (referred to as "base") and workspace administrators with granular controls over who can view, edit, comment, or otherwise modify data at the base, table, and field levels. There are four levels of Airtable user permissions:

- **Owner/Creator**: Full administrative control of base

- **Editor**: Sees full base, create and modify records and views, create and modify view share links

- **Commenter**: Sees full base, comment on records

- **Reader**: Sees full base

Direct access to a base or workspace is granted or removed by base owners and creators to Airtable users. Base owners and creators can control who has access to a base and can control any "share" links created for that base. They may also restrict editing of tables or fields within a base. Any collaborator given direct access to a base at any permission level will be able to duplicate that base and share that data further. It is important that direct access to the base is limited to individuals with a need to curate or analyze the data.

5.2.1.2 Share Links and Interfaces

To further restrict access to a base, users can be given indirect access via revocable share links or interfaces.

Share links can be customized to prevent users from seeing the full base, prevent duplicating the base, and prevent copying data from the base. The ability to use the link can be password protected, restricted to people with certain email domains, and may be revoked at any time. If there are concerns about data leaks via base or table duplication, inviting people with a need to view the data via share links constrains their ability to extract data from the base.

Interfaces are dynamic dashboards built on a limited set of data in an airtable base. Users can explore or even edit data based on the permissions provided by the interface creator. Access can be further tuned by setting up a "current-user filter". See this guide for more information.

5.3 Data in Airtable

See Airtable plan comparison for more information on the size of bases and features available. Information in this section pertains to all plans unless specified.

5.3.1 Workspaces, Bases, Tables, Fields, Records

Airtable uses workspaces, bases, tables, and fields to manage data. A workspace generally pertains to a particular project and contains all bases relevant to that project. Sharing a workspace with someone allows them to see all bases within that workspace.

Bases are equivalent to databases. They consist of a set of tables that can be linked and allow you to perform some task (e.g. IRB tracking, capturing research data, etc.). Bases can be duplicated and shared across your workspaces and you can share bases with other users.

Tables are where most of the action happens. Data is entered in tables, tables can be transformed via views into calendars, dashboards, or galleries, and tables can be manipulated via the API. They describe a data entity and are composed of fields. In the bat sampling example, each circle would be a table in the bat sampling base (see Figure 5.1).
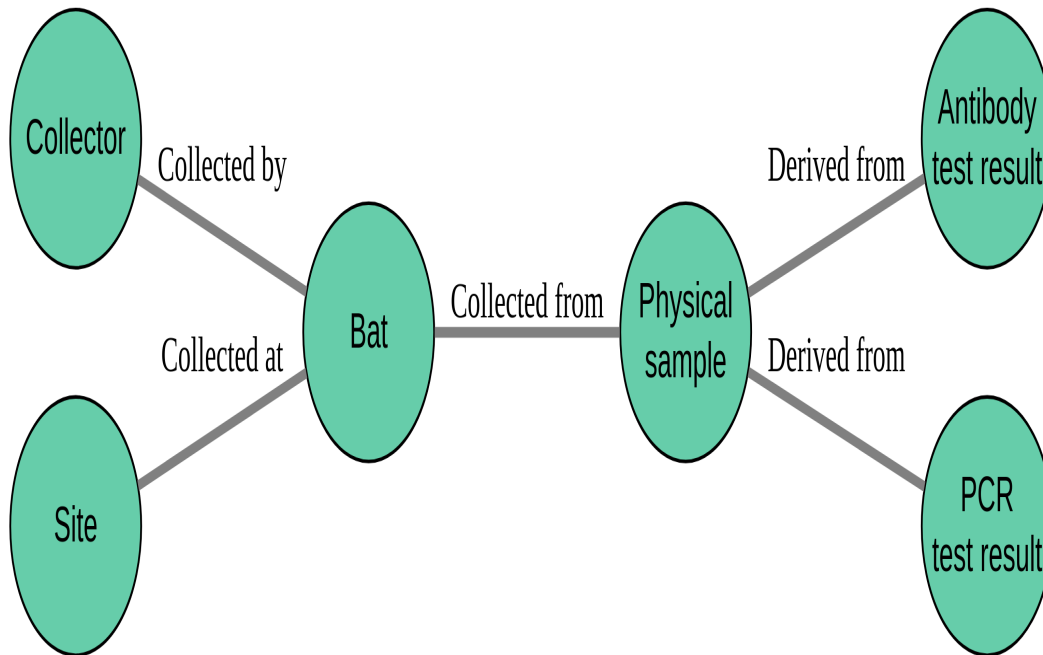
Figure 5.1: Data model for bat sampling example

Fields represent properties of a data entity. In a spreadsheet view, fields are columns. Airtable allows you to control field types (date, number, text, file attachments, logical, etc.) and paid plans allows you to control who can edit fields. Fields can be linked between tables creating links.

Records are the individual data points in a table. In a spreadsheet they would be the rows. Records are shaped by the structure you have created in tables and fields. Each record has a URL that can be used to access it programmatically or share it.

5.4 Views

In Airtable, tables can be displayed in different views to emphasis different components of the data. Views are great for creating concise presentations of data, especially in sprawling tables. The default view in an table is the grid (spreadsheet) view. All other views will derive from the data entered in this view.

For more on views, see the guide to views.

5.5 Internal Backups: Record History, Base Snapshots

Airtable has system for tracking changes to a base. They provide revision history for individual records (how long those histories are stored varies by plan). Any comments made in the revision history will be stored for the life of the record (on any plan). The current state of an entire base may be captured through snapshots. Should a systemic issue arise, the base can be restored to a snapshot at a later date. Restoring from a snapshot will remove the revision history for a record but comments on that record will be maintained.

As revision histories are maintained, "deleted" records may be retrieved. In the event of the need to permanently remove data, the revision history of the base may be removed.

Airtable allows data to be exported as CSVs from individual tables. At this time there is not an Airtable supported base export function.

5.6 Importing data

Data can be imported to Airtable from a number of sources including CSV files, excel, Google sheets, XML, and via copy paste. Airtable will guess what the most appropriate field type is, so make sure the field type is appropriate for the data (e.g. convert from text to date type fields). Certain sources, like Google sheets, can be imported as bases.

For more on importing data see: Importing and Adding data

5.7 Base Design

Having a good base design will make using your data easier. Generally, the process looks like this:

1. Describe what the database will do and collect use cases
2. Determine the roles of various stakeholders
3. List out the entities in the database and define their properties
4. Map out how the entities fit together (which properties link them)
5. Check that the mapping meets the use cases
6. Build base in Airtable
7. Check that the base meets the use cases

If you are migrating from spreadsheets, you likely already have an idea of what you need the base to do and a collection of data properties. It is still a good idea to follow the steps outlined above for mapping out entities. You may find that entire sheets can be replaced by views or that the data in one sheet should actually be stored in two different tables.

5.8 Automating Airtable

Airtable has five main routes for automating processes.

1. Automations - a drag and drop visual programming tool
2. Extensions - pre-built applications that perform some task
3. Scripting - use JavaScript to automate tasks within Airtable
4. Blocks - use JavaScript to create custom applications
5. REST API - use whatever programming language you like to automate processes

5.8.1 Automations with Drag and Drop Programming (pro and above)

The automations feature within Airtable allows you to visually program routines. Each automation has three basic components (see Figure 5.2):

- Status - controls whether or not the automation will run when the trigger condition is met
- Triggers - condition for automation to run: the creation or change of a record, a scheduled time, or some other external action
- Actions - what the automation will do

Figure 5.2: Actions diagram

Automations are commonly used to augment the synced tables feature, send notifications, check data quality, and manipulate or create records.

This automation (see Figure 5.3) creates a weekly summary of applicants who applied for a position. Its trigger is time based, then it finds records that match a condition, and finally it generates an email from those records and sends to the appropriate recipients.
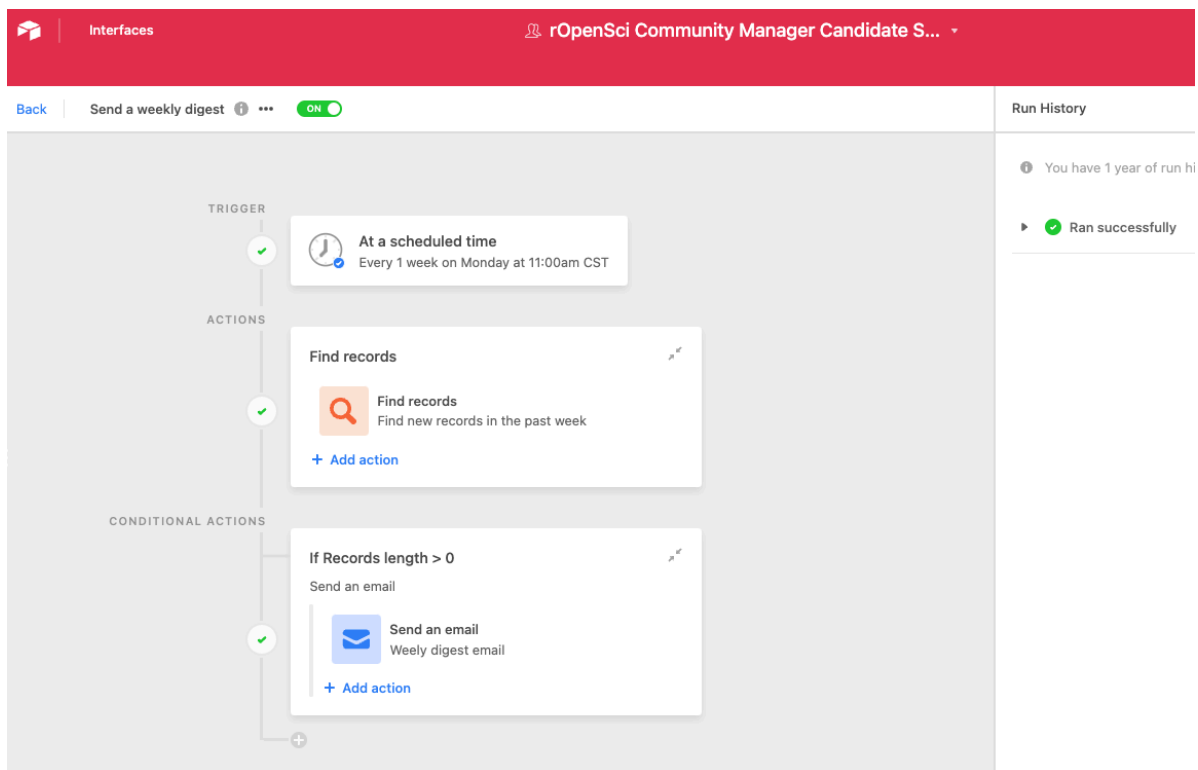
Figure 5.3: Airtable weekly summary

5.8.2 Scripting and Blocks (pro plan and above)

Scripting uses JavaScript to manipulate a base from within an application in Airtable. Scripting is flexible but has a steep learning curve. The scripting environment Airtable provides can be helpful as it provides code linting, direct access to documentation, and example scripts to build from. The major drawback to scripting is that scripts live in the base and files are not version controlled.

- There are a LOT of pre-written scripts, in a shared "marketplace" to perform automatic actions. Search the marketplace before you start writing bespoke code.

Blocks are custom applications built in JavaScript and node.js that add to base functionality. They are created in a development environment outside of Airtable then brought back into platform. There are number of tutorials for getting started with blocks.

5.9 Using the REST API

All Airtable bases are automatically accessible to authorized users via a REST API. The list of API accessible bases you have access to can be found here: https://airtable.com/api. By clicking on a base you will be able to see the full API documentation for that base.

5.9.1 Scoped Tokens

Airtable is moving to a scoped tokens based approach to api access. Scoped personal access tokens allow you to create a token for a specific base with specific permissions - e.g. token has read-only access to a bat sampling base. Using scoped tokens in this way means that if the token is compromised (leaked, stolen, accidentally committed unencrypted to a github repo, etc), you can delete that token to remove any access it might have had and the limited scope means that you know exactly what a person would have been able to access.

To create a personal access token go here: https://airtable.com/create/tokens

> ⚠️ Warning
>
> Remember to save the token in a secure place. Do not store unencrypted tokens on the web (e.g. pushing them to GitHub).

Airtable has also deployed OAuth tokens for all users.

5.9.2 Airtable and R

The Airtable REST API can be used via R with the airtabler package. EHA has started a fork of the package that has additional functionality so it is recommended to use that version. The original package design works well for exploring the data. Our extension adds additional functionality to help use `airtabler` in automation via continuous integration such as GitHub Actions.

```r
devtools::install_github("ecohealthalliance/airtabler")
```

The Airtable API serves up data as JSON, which has a hierarchical structure similar to a list in R. To handle JSON, `airtabler` uses the `jsonlite` package. Its helpful to understand how `jsonlite` handles different JSON structures when working with more complicated Airtable data. See the `jsonlite` quick-start guide for a basic overview. The purrr package is extremely helpful when dealing with data objects derived from JSON because it facilitates navigating nested data structures.

The `airtabler` package provides instructions for setting up access to the Airtable API. You will need to follow those instructions for Example 5.1 and Example 5.2 to work.

Example 5.1.

```r
library(airtabler)

table1  <-  fetch_all(base = "app49bbyLczZxX9PM",table_name = "One To One")

table2 <- fetch_all(base = "app49bbyLczZxX9PM",table_name = "Table 2")

## Linked records are stored as JSON arrays so they become lists
## when part of a data frame. Because each array has a length of 1
## we can safely unlist the arrays and add them back as to the
## data frame.

table1$LinkedRec <- unlist(table1$LinkedRec)

joinedTables <- dplyr::left_join(table1,table2, by = c("LinkedRec"="id"))

recordKey <- joinedTables[c("Name","LinkedRec","Number")]

recordKey
```

One to One join

Example 5.2.

```r
library(airtabler)
library(dplyr)
library(purrr)

oneToMany <- fetch_all(base = "app49bbyLczZxX9PM", table_name = "One To Many")

table2 <- fetch_all(base = "app49bbyLczZxX9PM",table_name = "Table 2")

# Depending on how you want to work with the data joins are a
# little trickier here.

## to replace the values but keep the structure

oneToMany$LinkedRecReplace <- purrr::map(oneToMany$LinkedRec, function(x){
  table2 %>%
    filter(id %in% x) %>%
    select(c("id","Number")) %>%
    pull("Number")
})

data.frame(
  id = unlist(oneToMany$LinkedRec),
  label = unlist(oneToMany$LinkedRecReplace)
)
```

One to Many Join

5.10 Data Management

Because of its flexibility and ease of use, it is extremely important that data management for airtable be taken seriously. Unlike most other relational databases, the fundamental properties of your base can be changed easily by multiple users without warning! Documenting the structure and purpose of your base, as well as creating regular external backups could save you from catastrophe.

5.10.1 Metadata

- **Structural metadata** - Information about a resource that tells you how its put together - e.g. the relationship between a table, field, and automation. With the exception of users on the enterprise plan, you must be deliberate about creating and maintaining structural metadata for your base.

Example of Airtable Structural Metadata:

- **Descriptive metadata** - Information about a resource that makes it easier to find and attribute data. This includes things attributes like author, title, description, keywords, etc. This table becomes especially important when data are transitioned out of airtable, at the end of a project, or if the base will be shared broadly with collaborators.

Example of Airtable Descriptive Metadata:

To see the template base, follow this [link](#)

5.10.1.1 Metadata for automations and extensions

Automations and extensions live entirely inside airtable. As of August 2022, Airtable introduced a "Manage Fields" tab that provides metadata for a specific table, including any automation dependencies. It is not possible to export those dependency tables or the code used in automations automatically. This makes it extremely important to document any automations and extensions outside your base.

5.10.2 External Backups

It is a good idea to create regular external backups of airtable data in the event that something catastrophic happens to your base or you simply decide you no longer wish to use airtable as your data store.Unfortunately, airtable does not provide an off the shelf solution for this. Below are three options for extracting all data from your base.

5.10.2.1 Using airtabler

Using `airtabler::air_dump()` and `airtabler::air_dump_to_csv()` functions, you can export all tables to R then create a versioned folder of CSVs. See the `air_dump_to_csv` help page for more information.

5.10.2.2 Other Open Source Projects

UnlyEd - use this template to push airtable backups to AWS S3.

- https://github.com/UnlyEd/airtable-backups-boilerplate

5.10.2.3 Paid services

Sequin - replicate airtable to postGres database (https://www.sequin.io/sources/airtable)

Part II

Database

6 Database - DoltHub

7 Database

> **i** Airtable not anymore used to host the database
>
> As of May 2024, the Africa CDC Database backend has shifted from Airtable to DoltHub. This chapter has been kept here to document how the database started and to show the history of its development.
> To read about DoltHub, see Chapter 2. To learn more about the use of dolt and DoltHub as the current backend for the database, read Chapter 6.

The current database is built using Airtable. The current database has the following schema:
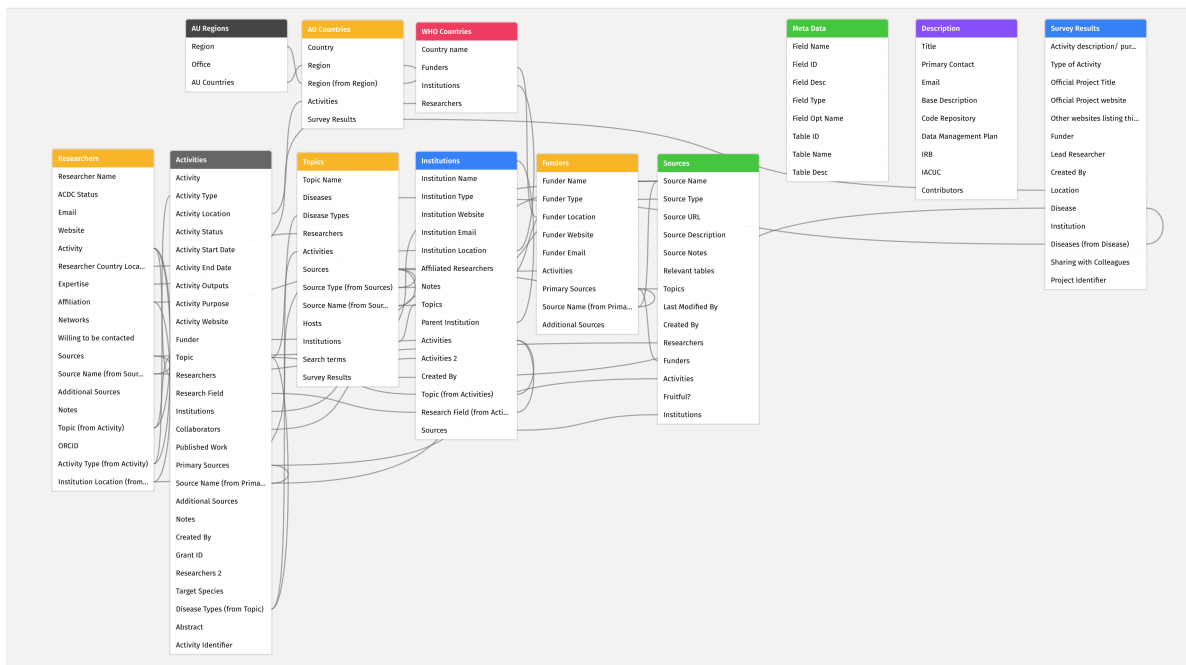
This schema can be viewed interactively from here - https://airtable.com/appAL7fJUpBPYtOq4/tblt9ott045tWl

7.1 Activities Table

7.1.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Activities") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Activities pertain to the formal/official title of the project that has been identified during the database search or survey.

Following are the different fields within the Activities table with their description.

7.1.2 Fields details

```
create_data_dictionary(metadata = airtable_metadata, table = "Activities") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description
Activity	Activity name
Source Name (from Primary Sources)	Name of primary source
Notes	Internal notes for maintenance
Topic	Activity Topic
Grant ID	Grand Identifier
Collaborators	Activity collaborators
Researchers	Researchers involved in activity
Published Work	Links to published work
Activity Location alt	Activity location
UN Stats Sub-region Name (from Activity Location alt)	NA
Activity Outputs	Activity Outputs
Activity Status	Activity Status
Activity Website	Activity Website
Activity Identifier	Activity identifier
Primary Sources	Primary source of information
Date validated	Date when record was validated.

Validated	Indicate whether this record has been validated.
Additional Sources	Additional sources
Continent Name (from Activity Location alt)	NA
AU Region Name (from Activity Location alt)	NA

7.2 Countries Table

7.2.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "AU Countries") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

List of countries used to indicate the location information for the researchers, topics, funders, sources, institutions, and activities tables

Following are the different fields within the Countries table with their description.

7.2.2 Fields details

```
create_data_dictionary(metadata = airtable_metadata, table = "Countries") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description
Researchers	NA
Funders	NA
AU Region Name	NA
UN Stats Sub-region Name	Name of further sub-regional grouping to which country is usually classified wit

7.3 Description Table

7.3.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Description") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

The Description table provides overall description of this entire database.

Following are the different fields within the Description table with their description.

7.3.2 Fields details

```
create_data_dictionary(metadata = airtable_metadata, table = "Description") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description	Field Type	Field Valu
Email	Email for primary contact	email	NA
Primary Contact	Person to correspond with about the base	singleLineText	NA
IRB	Optional: Link to IRB	url	NA
Base Description	Characterization of the base. What is the base describing?	multilineText	NA

7.4 Funder Table

7.4.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Funders") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Funders pertain to the entities funding each activity.

The funders' information is retrieved from declared funders for the activities identified during the database search or the survey.

Following are the different fields within the Funders table with their description.

7.4.2 Fields details

```
create_data_dictionary(metadata = airtable_metadata, table = "Funders") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description	Field Type
Funder Type	Type of funder	singleSelect
Funder Website	Website of funder	url
Source Name (from Primary Sources)	Source from where funder was identified	multipleLookupV
Activities	Activities supported by funder	multipleRecordL
WHO Region Name (from Funder Location alt)	NA	multipleLookupV
Continent Name (from Funder Location alt)	NA	multipleLookupV
Primary Sources	Source from where funder was identified	multipleRecordL

7.5 Institutions Table

7.5.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Institutions") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Institutions pertain to the entities directly related to or implementing the activity and/or the affiliation of the researcher implementing the activity.

Following are the different fields within the Institutions table with their description.

7.5.2 Fields details

```
create_data_dictionary(metadata = airtable_metadata, table = "Institutions") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description	Field Type	Field Value
Institution Email	Email of institution	email	NA
Institution Website	Institution Website	url	NA
Activities	Name of activity identified with institution	multipleRecordLinks	Activities
Sources	Name of primary source	multipleRecordLinks	Sources
Institution Type	Institution Type	singleSelect	Research Ins
Institution Location	Institution location	multipleRecordLinks	WHO Count
Notes	Internal notes (for maintenance)	singleLineText	NA
Topic (from Activities)	Name of topic identified with the institution	multipleLookupValues	NA

7.6 Researchers Table

7.6.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Researchers") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Researchers pertains to the people involved in the activities contained in this database, and maybe involved in a variety of activity types. including but not limited to, research.

Researchers identified during the database search. Information on researchers are primarily retrieved once activities have been identified from the database search process. It is possible that researchers are identified apriori and from which information on topics, funders, sources, institutions, and activities may be identified relevant to the specific researcher. This may happen when the planned/proposed survey is implemented and sent to known researchers who may or may not be in the database to begin with.

Following are the different fields within the Researchers table with their description.

7.6.2 Fields details

```
create_data_dictionary(metadata = airtable_metadata, table = "Researchers") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description
WHO Region Name (from Researcher Country Location alt)	NA
ACDC Status	Is researcher from the Africa CDC?
Sources	Name of source
UN Stats Region Name (from Researcher Country Location alt)	NA
Affiliation	Institution researcher is affiliated with. Ch
Institution Location alt (from Affiliation)	NA
Topic (from Activity)	Name of topic
Researcher Name	Full name of researcher identified in the se
Notes	Internal notes (for maintenance)
Source Name (from Sources)	Name of source
Additional Sources	Additional sources
Activity Type (from Activity)	Type of activity
Topic (from Activity) 2	NA
Website	Personal website of researcher (if any)
Activity	Activity/activities researchers is involved i
Institution Location (from Affiliation)	Name of location where institution is based
Willing to be contacted	Is the researcher willing to be contacted?
Continent Name (from Researcher Country Location alt)	NA

7.7 Sources Table

7.7.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Sources") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Sources contain initial set of sources (primarily funders) identified to initiate the database search. These original sources are fully described here - https://ecohealthalliance.github.io/rig-handbook/sources.html.

From the initial search, other sources were identified from which additional streams of searches were performed. Currently, the sources table is updated through a primary search of possible sources in addition to the search performed on the original sources list.

Following are the different fields within the Sources table with their description.

7.7.2 Fields details

```
create_data_dictionary(metadata = airtable_metadata, table = "Sources") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description	Field Type	Field Values
Source Type	Type of primary source	singleSelect	Website text p
Source Notes	Internal notes for primary source (for maintenance)	multilineText	NA
Topics	Topics identified from primary source	multipleRecordLinks	Topics
Relevant tables	Notes on relevant tables found from primary source	multipleSelects	Researchers, T
Funders	Funders identified from primary source	multipleRecordLinks	Funders
Source Name	Name of primary source	singleLineText	NA
Created By	Person who created the entry	createdBy	NA
Researchers	Name of researcher identified from primary source	multipleRecordLinks	Researchers

7.8 Survey Results Table

7.8.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Survey Results") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

NA

Following are the different fields within the Survey Results table with their description.

7.8.2 Field details

```
create_data_dictionary(metadata = airtable_metadata, table = "Survey Results") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description	Field Type	Field V
Activity description/ purpose	Description of activity	multilineText	NA
Sharing with Colleagues	Is respondent going to share with colleagues?	singleLineText	NA
Lead Researcher	Name of lead researcher	singleLineText	NA
Created By	Name of survey respondent	createdBy	NA
Other websites listing this project	Other website listings for this project	singleLineText	NA
Funder	Name of funder for this project	singleLineText	NA
Official Project Title	Official project title	singleLineText	NA
Official Project website	Official project website	singleLineText	NA

7.9 Topics Table

7.9.1 Table details

```
create_data_dictionary(metadata = airtable_metadata, table = "Topics") |>
  (\(x) x$table_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Topics pertain to the disease names and disease types.

The topics table contains disease types/disease entities of interest to the Africa CDC based on their priorities (as described in their Framework document) and is continually updated with whatever disease/disease entity/disease type were deemed relevant by the database contributor.

Following are the different fields within the Topics table with their description.

7.9.2 Fields details

```r
create_data_dictionary(metadata = airtable_metadata, table = "Topics") |>
  (\(x) x$field_details)() |>
  knitr::kable() |>
  kableExtra::row_spec(row = 0, bold = TRUE) |>
  kableExtra::kable_paper(lightable_options = "striped")
```

Field Name	Field Description
Sources	Name of Source
Diseases	Disease entity related/relevant to the field/
Activity Location (from Activities)	Name of location of activity in which topic
Survey Results	Results from survey
Search terms	Search terms used to identify topic
Disease Types	Type/category to which disease is grouped
Continent Name (from Activity Location alt) (from Activities)	NA
Activity Location alt (from Activities)	NA
Institutions	Institutions
Region (from Activity Location) (from Activities)	Name of region where topic was identified
AU Region Name (from Activity Location alt) (from Activities)	NA

Part III

Sources

8 Sources

Following is an initial list of sources of information used for the RIG database.

The initial search performed was non-systematic and focused primarily on a known funder of global research related/relevant to the topics of interest for the database . The main aim of focusing first on this limited and focused search was to get a sense of what information is available from such bodies/organisations, and the limitations of the information available. This is based on an initial idea that research funders would tend to have a system of collecting/archiving information on research they have funded. The expectation was that at the minimum, the information available from funders would lead to identifying further sources of information relevant to the ACDC database specifically those of research groups/institutions particularly those based in countries/regions within Africa. This initial search will hopefully inform a more systematic and informed search strategy for the database information.

8.1 UKRI

UK Research and Innovation or UKRI is a non-departmental public body in the United Kingdom that was established in 2018. It brings together the seven UK Research Councils, Innovate UK, and Research England, which were previously separate organizations, to create a single body that oversees research and innovation funding and strategy in the UK.

The seven UK Research Councils are:

1. Arts and Humanities Research Council (AHRC)
2. Biotechnology and Biological Sciences Research Council (BBSRC)
3. Engineering and Physical Sciences Research Council (EPSRC)
4. Economic and Social Research Council (ESRC)
5. Medical Research Council (MRC)
6. Natural Environment Research Council (NERC)
7. Science and Technology Facilities Council (STFC)

Innovate UK is the UK's innovation agency, which provides funding and support for innovative businesses and projects.

Research England is responsible for funding and overseeing research in English universities and higher education institutions.

UKRI's main role is to drive innovation and research in the UK and to support research and development that benefits society and the economy. It funds research projects, provides support to researchers, promotes international collaboration, and works to ensure that research and innovation are integrated with government policies and priorities.

Of these various groups within UKRI, we further focused on the Biotechnology and Biological Sciences Research Council (BBSRC), Medical Research Council (MRC), Science and Technology Facilities Council (STFC), Innovate UK, and Research England.

8.2 Wellcome Trust

The Wellcome Trust is a global charitable foundation based in the UK. It was established in 1936 by Sir Henry Wellcome, a pharmaceutical entrepreneur and philanthropist. The Wellcome Trust is one of the largest charitable organizations in the world, with an endowment of over £29 billion.

The Trust's mission is to improve health by supporting scientists, researchers, and innovators in their work to understand, treat, and prevent disease. The Trust funds research in areas such as neuroscience, genetics, infectious diseases, and global health. It also provides support for public engagement with science, education and training for scientists, and the translation of research into practical applications that benefit patients and communities.

The Wellcome Trust is known for its long-term, strategic approach to funding research, and for its commitment to open science and data sharing. It also operates the Wellcome Collection, a public venue in London that hosts exhibitions and events related to health, medicine, and science.

8.3 National Institutes of Health

The National Institutes of Health (NIH) is a biomedical research agency of the United States federal government. It is the largest biomedical research institution in the world, with its main campus located in Bethesda, Maryland. The NIH is composed of 27 separate institutes and centers, each with a specific research focus, and is responsible for conducting and funding research in a wide range of areas, including cancer, genetics, infectious diseases, and neuroscience.

The NIH was founded in 1887 as the Hygienic Laboratory and was later renamed the National Institutes of Health in 1930. Today, it is one of the world's foremost centers for medical research, with a mission to seek fundamental knowledge about the nature and behavior of living systems and to apply that knowledge to enhance health, lengthen life, and reduce illness and disability. The NIH is funded by the U.S. government through the Department of Health and Human Services and operates under the direction of the Office of the Director.

8.4 National Science Foundation

The National Science Foundation (NSF) is an independent federal agency of the United States government that supports fundamental research and education across all fields of science and engineering. The NSF was established by the National Science Foundation Act of 1950 and has a budget of around $8 billion.

The NSF funds research and education in areas such as mathematics, computer science, physics, chemistry, biology, social sciences, and engineering. It supports individual researchers, small teams, and large interdisciplinary research collaborations through a competitive, merit-based process of proposal submission and review. The NSF also supports the development of science, technology, engineering, and mathematics (STEM) education at all levels, from K-12 through graduate education.

The NSF operates through several directorates and offices, each with a specific research focus or mission, such as the Directorate for Biological Sciences, the Directorate for Social, Behavioral and Economic Sciences, and the Office of Polar Programs. The NSF works to advance scientific discovery, promote science education and outreach, and promote innovation and economic growth through its investments in research and education.

8.5 Defense Advanced Research Projects Agency

The Defense Advanced Research Projects Agency is a research and development agency of the United States Department of Defense that is responsible for the development of emerging technologies for use by the military.

DARPA was established in 1958 in response to the Soviet Union's launch of Sputnik, the first artificial satellite, and has been involved in a number of high-profile technological innovations, including the development of the Internet, GPS, and stealth technology.

DARPA's mission is to maintain the technological superiority of the U.S. military by sponsoring and conducting research in a wide range of fields, including artificial intelligence, robotics, biotechnology, materials science, and aerospace technology. DARPA works with academic researchers, private companies, and other government agencies to develop and test new technologies, and it is known for its high-risk, high-reward approach to research and development.

Some of DARPA's current research initiatives include the development of hypersonic weapons, the creation of autonomous drone swarms, and the development of brain-machine interfaces for use in treating neurological disorders. DARPA's work has had significant impacts on both military and civilian technology, and the agency is seen as a leader in cutting-edge research and development.

8.6 ClinicalTrials.gov

ClinicalTrials.gov is a publicly accessible database of clinical trials that are being conducted worldwide. It is maintained by the National Library of Medicine, a part of the National Institutes of Health (NIH) in the United States.

The database provides information on clinical trials for a wide range of diseases and conditions, including both interventional and observational studies. It includes information about the purpose of the trial, who may participate, where the trial is being conducted, and the status of the trial, such as whether it is recruiting participants or has been completed.

ClinicalTrials.gov was created in response to a 1997 law requiring the registration of clinical trials for certain serious or life-threatening diseases or conditions. Since then, the database has grown to include information on thousands of trials from around the world.

ClinicalTrials.gov is an important resource for researchers, healthcare professionals, and members of the public who are interested in clinical research. It can be used to identify ongoing or completed trials, learn about the purpose and design of a study, and find out how to participate in a trial. It also serves as a platform for researchers to share their results and comply with the requirements of various funding agencies and regulatory bodies.

8.7 GEPRIS

Geförderte Projekte in der Forschung und Entwicklung (Funded Projects in Research and Development) or GEPRIS is an online database of research projects funded by the German Research Foundation (DFG).

The DFG is the largest independent research funding organization in Germany and funds projects across all scientific disciplines, from the humanities and social sciences to the natural and life sciences. GEPRIS provides information about the projects that the DFG has funded, including their aims, methods, and outcomes, as well as the institutions and researchers involved.

Researchers and members of the public can use GEPRIS to search for projects that have been funded by the DFG, and to access information about these projects. The database includes information about ongoing and completed projects, and users can search by various criteria, such as by researcher name, institution, scientific discipline, or project title.

GEPRIS is a valuable tool for researchers to identify potential collaborators, explore research trends, and find information about the funding landscape in their field. It is also useful for members of the public who are interested in learning about the research being conducted in Germany and the impact of this research on society.

8.8 EDCTP

European and Developing Countries Clinical Trials Partnership or EDCTP is a public-public partnership between countries in Europe and sub-Saharan Africa, established in 2003, with the aim of accelerating the development of new clinical interventions to fight infectious diseases that disproportionately affect Africa.

The partnership's mission is to improve the health of people in Africa by supporting the development of new medicines, vaccines, and other health interventions to prevent and treat diseases such as HIV/AIDS, tuberculosis, malaria, and neglected infectious diseases. EDCTP supports collaborative research projects that bring together scientists, institutions, and countries from both regions to conduct clinical trials and other research activities.

EDCTP works with a range of partners, including national governments, research institutions, civil society organizations, and the private sector, to support research that is relevant and responsive to the needs of African communities. It also provides training and capacity-building opportunities to support the development of sustainable health research infrastructure and expertise in Africa.

The partnership is funded by the European Union, its member states, and other donors. Since its inception, EDCTP has supported over 100 collaborative research projects and played a key role in advancing the development of new interventions for infectious diseases that affect the people of Africa.

8.9 GLOPID-R

Global Research Collaboration for Infectious Disease Preparedness or GLOPID-R is an international partnership that aims to strengthen global research efforts in the field of infectious disease preparedness. The partnership was established in response to the 2014 Ebola outbreak in West Africa, which highlighted the need for improved global coordination and collaboration in research and development for emerging and re-emerging infectious diseases.

GLOPID-R brings together stakeholders from the global health community, including research funders, policy-makers, researchers, and public health organizations. The partnership aims to promote international cooperation and coordination in research to accelerate the development of new tools and approaches to prevent, detect, and respond to infectious disease outbreaks.

GLOPID-R's main objectives include identifying research priorities for infectious disease preparedness, coordinating research efforts across different regions and countries, and promoting capacity building and knowledge exchange to strengthen global health research infrastructure.

The partnership focuses on a range of infectious diseases, including those caused by emerging and re-emerging pathogens, neglected tropical diseases, and antimicrobial resistance. It works

to support research efforts across the entire spectrum of infectious disease preparedness, from basic research to clinical trials and implementation research.

GLOPID-R is supported by a range of funding agencies and partners from around the world and is seen as an important platform for promoting global cooperation and collaboration in infectious disease research and preparedness.

8.10 NRF South Africa

The National Research Foundation of South Africa (NRF) is an independent organization that promotes and supports research and innovation in all fields of science, engineering, technology, and social sciences in South Africa. The NRF was established in 1999 through the National Research Foundation Act and operates under the jurisdiction of the Department of Science and Innovation.

The NRF provides funding, develops policies, and manages research infrastructure to support South African researchers and institutions. It also fosters international collaboration in research, and supports the training and development of the next generation of researchers through various funding and fellowship schemes.

The NRF provides funding through a number of programs, including competitive grants, fellowships, and research chairs. It also supports the development of research infrastructure and the establishment of research centers of excellence.

In addition to providing funding and support for research, the NRF plays a key role in developing research policies and strategies at the national level. It advises the South African government on research priorities and is involved in various initiatives aimed at promoting science, technology, and innovation in the country.

The NRF is an important organization for the South African research community and has been instrumental in advancing the country's research and innovation capabilities. Its funding and support have contributed to numerous scientific discoveries and innovations in a wide range of fields, including health, energy, and the environment.

Part IV

Maintenance

9 Updates

The current update process of the RIG database is summarised in this workflow:
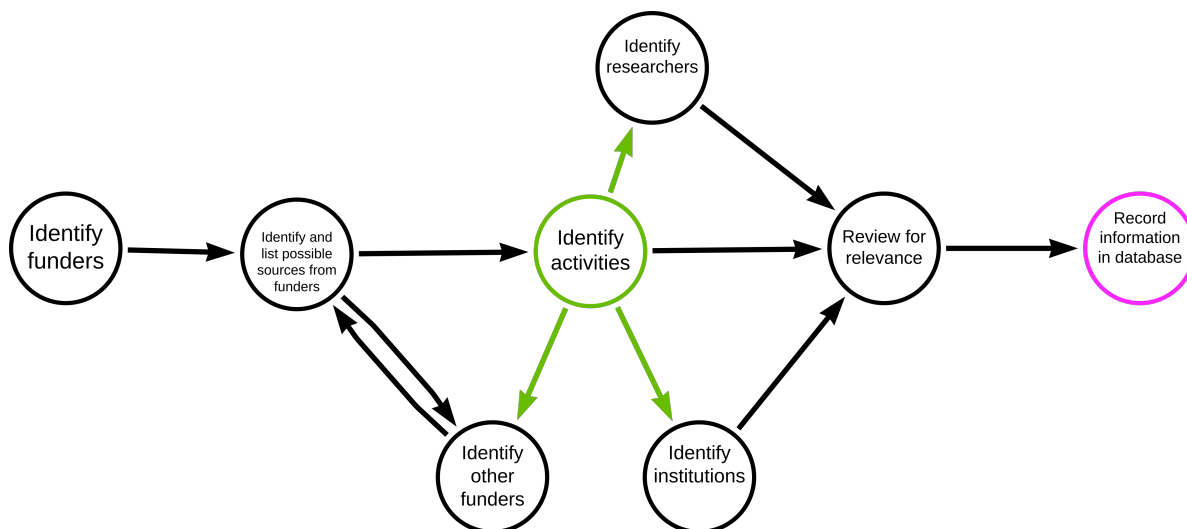
Following are the source-specific process of updating or retrieving information for the RIG database.

9.1 Wellcome Trust Grant Funding Data

9.1.1 General Information

Using the downloadable spreadsheet of funds awarded between 01st of October 2005 and 4th of May 2022 found at https://cms.wellcome.org/sites/default/files/2022-05/Wellcome-grants-awarded-1-October-2005-to-04-05-2022.xlsx, the following steps were taken to retrieve relevant information for the database:

9.1.2 How to update

Please note that the steps below were done using the current available spreadsheet from the Wellcome Trust website and added the relevant projects to the RIG database. These steps

can therefore be used as a guide for how to update the database with new information in the future to when the Wellcome Trust publishes its most up-to-date spreadsheet.

1. Go to column J – Recipient Org:Country -> deselect all and then select all the African countries in the list

2. Go to column N – Planned end date -> select all the years in the future

3. These two steps reduced the list from 19,833 projects to 111 projects

4. Read the project title and decide if the project is relevant for our database or not

5. If unsure, read the abstract – that also helps to identify the keywords to tag the project within our database

6. If the project is relevant transfer all the information into our database

Note: This method is only able to detect projects/activities where an African organisation itself holds the grant. It does not detect projects where African researchers are involved as collaborators. The spreadsheet does not list collaborators on projects, so it's yet to be determined how we will identify projects on which African research institutes collaborate with international organisations being awarded the grant.

9.2 ClinicalTrials.gov

9.2.1 How to update

Following are steps taken to extract data from ClinicalTrials.gov.

1. Start with searching a disease/topic of interest

2. On the results page apply the following filters to look for active studies

3. Click on 'Apply' and then look manually through the column 'Locations' of the list of the results to find studies that take place in African countries

Row	Saved	Status	Study Title	Conditions	Interventions	Locations
1	☐	Recruiting	Pharmacokinetics, Tolerability and Safety of Favipiravir Compared to Ribavirin for the Treatment of **Lassa** Fever	• **Lassa** Fever	• Drug: Ribavirin iv • Drug: Favipiravir	• Irrua Specialist Teaching Hospital Irrua, Edo State, Nigeria • Federal Medical Center of Owo Owo, Ondo State, Nigeria
2	☐	Recruiting	Prevalence and Incidence of **Lassa** Virus Infection in Southern Mali	• **Lassa** Virus Infection		• Icer/Mrtc/Fmos/Usttb Bamako, Mali
3	☐	Active, not recruiting	Dose-ranging Study: Safety, Tolerability and Immunogenicity of INO-4500 in Healthy Volunteers in Ghana	• **Lassa** Fever	• Drug: INO-4500 • Device: CELLECTRA™ 2000 • Drug: Placebo	• Noguchi Memorial Institute for Medical Research, University of Ghana Legon, Accra, Ghana
4	☐	Recruiting	A Clinical Trial to Evaluate the Safety and Immunogenicity of rVSVΔG-LASV-GPC Vaccine in Adults in Good General Heath	• **Lassa** Fever • **Lassa** Virus Infection	• Drug: rVSVΔG-LASV-GPC • Other: Placebo/Diluent	• George Washington University Washington, District of Columbia, United States • East-West Medical Research Institute Honolulu, Hawaii, United States • Brigham and Women's Hospital Brookline, Massachusetts, United States • Redemption Hospital New Kru Town, Greater Monrovia, Liberia
5	☐	Recruiting	**Lassa** Fever Clinical Course and Prognostic Factors in Nigeria	• **Lassa** Fever • **Lassa** Virus Infection • Pregnancy	• Other: Non interventional research	• Owo Federal Medical Centre (Owo FMC) Owo, Ondo, Nigeria

4. Click on the first study to start working your way through the information available

5. The first information provided is the sponsor -> this information should be added to the Funder – column in the Activities table

6. Information about collaborators can be added to Collaborators column

Pharmacokinetics, Tolerability and Safety of Favipiravir Compared to Ribavirin for the Treatment of Lassa Fever (SAFARI)

⚠ The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT04907682

Recruitment Status ❶ : Recruiting
First Posted ❶ : June 1, 2021
Last Update Posted ❶ : August 20, 2021
See Contacts and Locations

Sponsor:
Bernhard Nocht Institute for Tropical Medicine

Collaborators:
University of Hamburg-Eppendorf
Alliance for International Medical Action
Institut National de la Santé Et de la Recherche Médicale, France
University of Bordeaux
Federal Medical Centre, Owo
Irrua Specialist Teaching Hospital

Information provided by (Responsible Party):
Bernhard Nocht Institute for Tropical Medicine

7. Staying in the 'Study Details'-tab, scroll down to 'Study Design'

8. This section contains the official title, which should be used as the name for the Activity

9. Additionally it contains information about the start and end date, which should be copied into the respective fields in the Activities table

10. Scroll further down to 'Contact and Locations'

11. The information given under contacts should be added to the Researcher column in Airtable

12. Switching into the Researcher-table within in Airtable the given contact details should be added to the newly created entries for the involved researchers

13. Also the affiliation to a certain institute can be added based on these information as well as the researcher's location -> is it possible to link the Location with the Affiliation so that the location is automatically added based on the information about the institution the researcher is affiliated with?

14. Switch back into the Activities table and add information about the Locations to the Activity Location and the Institutions columns

Locations

Nigeria

Irrua Specialist Teaching Hospital   **Recruiting**
     Irrua, Edo State, Nigeria
     Contact: Peter Akhideno, Dr   +2348037048831   ehideno@yahoo.co.uk
     Contact: Cyril Erameh, Dr   +2348032413382   cyrilerameh@gmail.com

Federal Medical Center of Owo   **Recruiting**
     Owo, Ondo State, Nigeria
     Contact: Oluwafemi Ayodeji, Dr   femiayodeji@yahoo.com

Sponsors and Collaborators

Bernhard Nocht Institute for Tropical Medicine

University of Hamburg-Eppendorf

Alliance for International Medical Action

Institut National de la Santé Et de la Recherche Médicale, France

University of Bordeaux

Federal Medical Centre, Owo

Irrua Specialist Teaching Hospital

Investigators

Principal Investigator: Peter Akhideno, Dr   ISTH

Principal Investigator: Sylvanus Okogbenin, Prof   ISTH

Principal Investigator: Oluwafemi Ayodeji, Dr   FMCO

15. Scroll up again to the selection of tabs

16. Click on the Results tab -> it's worth checking this tab even when it's called No Results Posted as it might still contain links to publications that are affiliated with the study

17. These links can be copied into the Published Work column in the Activities table

18. I also copied the link of the study page on clinicaltrials.gov into the Activity Website column

19. to determine the Research Field, I had to use my own understanding of the study so I am not sure if this can be automated or rather needs to be done by a database librarian

9.3 Public journal/research databases

In order to aid in automation, maintain a list relevant search terms for each topic of interest (stored in the "Topics" table in Airtable). Even if the terms are not used for the purposes of developing a search strategy, they can be used by those who are not subject matter experts when collection information on a specific topic

9.3.1 Example of a successful search:

```
(zoonoses OR zoonotic disease OR zoonotic illness) and (africa*) and (surveillance OR trackin
```

The majority of results from this search, when conducted in PubMed, appeared relevant to the database (based on title/abstract scanning)

9.3.2 Example of PubMed search for surveillance activities for Brucellosis:

```
("surveillance"[Title/Abstract] OR "prevalence"[Title/Abstract] OR "monitoring"[Title/Abstra
```

This search yielded a large quantity of results, not all of which were relevant. Manual processes are required to validate results.

Including terms to filter the results based on location were helpful, but still included results not located on the African continent. Search term to filter for African countries:

```
(Djibouti[Title/Abstract] OR Seychelles[Title/Abstract] OR DR Congo[Title/Abstract] OR Comor
```

From publications, can extract researchers, institutions, funders, activities. Ideally, researchers, institutions, and funders can be extracted automatically as opposed to manually, but scripts would need to be customized for each journal.

9.3.3 Validation of results

Validation of results can be useful to better understand the overlap between publications and activities and determine the priority of searching through publications vs. navigating to institution sites directly (or other strategies).

After finding a relevant publication, look at the publication's authors and their respective institutions

Navigate to institutions' sites to search for publications or results from research

Are their activities listed on the site? Are those activities explicitly mentioned in the publications? Etc.

9.3.4 Some relevant journals/databases:

- Zoonoses & Public Health from Wiley Online Library : https://onlinelibrary.wiley.com/action/doSearch?Se
- Journal of Public Health in Africa: https://www.publichealthinafrica.org/jphia/issue/view/30
- PLoS Journal of Neglected Tropical Diseases: https://journals.plos.org/plosntds/search?filterJournals=PL

9.4 GEPRIS

9.4.1 General Information:

GEPRIS is a database listing all projects funded by the German Research Foundation (German: Deutsche Forschungsgemeinschaft; abbr. DFG) The DFG is a research funding organisation, which functions as a self-governing institution for the promotion of science and research in the Federal Republic Germany. In 2019, the DFG had a funding budget of €3.3 billion.

9.4.2 How to Use:

The database can be accessed here: https://gepris.dfg.de/gepris/OCTOPUS?language=en&task=showSearchSir

This link should directly lead to the English version of the website, otherwise the language can be changed by clicking on English in the top right corner.

- In the database one can search for Projects, People, or Institutions – for our purpose the project option is the most relevant
- One can either search for keywords or filter for different criteria – for a systematic approach I found using the filtering options easier than going through all our

1. On the search start site stay in the Projects tab.
2. Click on Show extended search.
3. Under Subject Area select one of the following:

   - Agriculture, Forestry and Veterinary Medicine
   - Basic Research in Biology and Medicine
   - Medicine
   - Microbiology, Virology, and Immunology
   - Social Sciences
   - Water Research

- Zoology

Note: After working through all these subject areas, any relevant project in the field of One Health should be picked up by the searches

4. Leave everything under DFG Programme as it is

5. Move on to Funding and change Status to Current

6. Move on to International and change Continent to Africa

7. Click on Find

8. Read through the project titles on the results page to identify relevant projects

9. Import all the relevant project information (as highlighted on the screenshots) into the Africa CDC database

Projekt

The potential of nonhuman primates as a reservoir for human yaws

Applicants	→ Privatdozent Dr. Sascha Knauf, Ph.D., until 7/2019; → Professor Dr. Christian Roos
Subject Area	Parasitology and Biology of Tropical Infectious Disease Pathogens
Term	since 2014
Project identifier	Deutsche Forschungsgemeinschaft (DFG) – Project number 252488542

Project Description

It has been propagated that human yaws caused by the bacterium Treponema pallidum subsp. pertenue (TPE) has no animal reservoir. Yet, reports of T. pallidum (TP) infection in nonhuman primates (NHPs) are accumulating. Several studies demonstrate simian infection with strains that are most closely related to human yaws-causing TPE strains. Our DFG funded study on NHP infection further supports the above mentioned findings. Data on simian TP strain diversity in Tanzania (TZ) are accumulating and reveal epidemiological insight into the spread of this conspicuous bacterium across the country. The phylogenetic branching pattern obtained from whole genome sequences of a greater number of simian strains (West and East Africa) suggests a rapid initial radiation of TPE across humans and NHPs and that at least ancestral TPE strains were most likely not host species specific. The East African Lake Manyara National Park simian strains reveals an overall gene synteny (including rRNA operons) with str. Gauthier with only 8.9% amino acid differences. A complete picture of TP infection in humans and NHPs including the definite answer to the NHP reservoir question will, however, only assemble when existing data on simian infection are compared to a greater number of circulating human TP(E) strains, which originate from areas with characterized simian infection. We hypothesize that humans and NHPs in TZ share the same TP(E) strains and also that NHPs in (Ghana) GH are infected with TP. While GH reports yaws until today, there is a chance of hidden TPE infections in humans in TZ, a country that currently does not report yaws, but harbors a great number of infected NHP species. DNA based assays (e.g. LAMP) are designed to fast-track identification on the subspecies level complementing the serological testing. This will increase specificity of yaws diagnosis from atypical clinical manifestations followed by whole genome sequencing.

increase specificity of yaws diagnosis from atypical clinical manifestations followed by whole genome sequencing. Simian and human TP strains from Africa have not yet been investigated using an integrated approach, and the results obtained are crucial for the understanding of Treponema evolution and epidemiology as well as to answer the important question on inter-species transmission. New laboratory TP strains will allow advanced immunological research and the creation of translational animal models which could open new pathways for vaccine development and testing. The project combines basic research in the field of Treponema infection with continues One Health capacity building and early-career research training at the African locations. A South-South partnership is created to support scientific excellence in the field of infectious disease research, addressing an urgent need in TZ and GH. Research activities of this project are logically expanded towards a yaws endemic country (GH). The project is expected to improve local health care services in TZ and GH and has the potential to support public health initiatives in post-yaws MDA surveillance.

DFG Programme	Research Grants
International Connection	Ghana, Tanzania
Co-Investigators	→ Idrissa S. Chuma; → Dr. Julius D. Keyyu; → Inyasi A. V. Lejora
International Co-Applicants	→ Professor Dr. Rudovick Kazwala; → Dr. Sayoki Mfinanga, Ph.D.; → Dr. Augustina Sylverken

10. To identify the research institutes that are involved in the project one has to click on the researchers names and extract that information from their profile (their affiliation with a research institute is listed there)

9.4.3 Positive aspects of this source:

The filtering options allow to filter for several criteria which are crucial for the relevance of a project to our database. That removes a lot of irrelevant projects from the results pages. The project pages list almost all the information we are interested in.

9.4.4 Downsides of this source:

The project page doesn't list the anticipated end date of a project.

One has to click on the link to the researcher's profile to identify the participating organisations.

Even using all the different filtering options not all resulting hits are relevant for our database, so I don't think the process can be fully automated or at least requires a subsequent manual validation or clean up step to remove irrelevant projects.

9.5 EDCTP

9.5.1 General Information

The European & Developing Countries Clinical Trials Partnership (EDCTP) is a non-profit organisation with a European office in The Hague, The Netherlands and an African office in Cape Town, South Africa. EDCTP is a partnership between European Union (EU), Norway, Switzerland, and African countries to accelerate the development of new clinical interventions such as drugs, vaccines, microbicides, and diagnostics against poverty-related diseases in Africa. The organisation supports clinical trials, capacity strengthening and networking in Africa and Europe. Funding comes from the EU, member states, pharmaceutical industry and private organisations and charities like The Wellcome Trust and The Bill & Melinda Gates foundation.

Note: Since funding comes from several sources that we also list as sources for populating the database such as the European Union (European Commission), The Wellcome Trust and The Bill & Melinda Gates Foundation, there is the possibility that downloading project information from all these sources into our database could lead to duplicate entries. I created a column in the Actvities table for the Project ID, as this might be helpful to identify duplicates and remove them automatically.

9.5.2 How to use:

The database of funded project can be accessed here: https://www.edctp.org/edctp2-project-portal/

There is the option to download the list of projects as a PDF, CSV or XLXS file. Personally, I did not find that helpful for manually adding projects to the database, but it might be useful for an automated process.

1. Go to Status of Project and select the filter Active

2. Go to Classification and select one of the following filters:

   - Co-Infections
   - COVID-19
   - Cysticercosis/Taeniasis

   - Diagnostics
   - Diarrhoeal Diseases
   - Drugs
   - Emerging Infections, incl. Ebola, Lassa
   - Epidemiology
   - HIV

- Human African trypanosomiasis (sleeping sickness)

- Implementation Research

- Leishmaniases
- Leprosy (Hansen disease)
- Lower respiratory infections

- Lymphatic filariasis
- Malaria
- Microbicides
- Onchocerciasis (river blindness)
- Rabies
- Schistosomiasis
- Soil-transmitted helminthiasis

- Social Science
- Tuberculosis
- Vaccines
- Yaws
- Yellow Fever

Note: Only one Classification at a time can be selected

3. Once one Classification was selected click on search

4. Change from Show Map to Show List -> this makes it easier to systematically look through the projects

5. On the results list you can see the location of the coordinating organisation, but even if it is not in an African country, it is worth checking the project details for the Participating Organisations. So read the project title and decide whether this could be a relevant project, if so, click on View details

Improved flucytosine formulation for the treatment of meningitis in advanced HIV disease
Reference: RIA2018CO-2516

Acronym: 5FC HIV-Crypto

Project Website:

Coordinating Organisation: Drugs for Neglected Diseases Initiative (DNDi), Switzerland

Coordinating Organisation Website: https://dndi.org/

Name of Coordinator: Dr Isabela Ribeiro

Participating Organisations:
- St. George's Hospital Medical School, United Kingdom
- Luxembourg Institute of Health (LIH), Luxembourg
- National Institute for Medical Research - Tanzania (NIMR), Tanzania, United Republic of
- FARMOVS (Pty) Ltd, South Africa
- University of North Carolina Project-Malawi (UNC Project-Malawi), Malawi

Grant Amount: € 3,537,000.00

Start Date: Jul 1, 2020

End Date: Dec 31, 2024

Duration (months): 54

Abstract:
Cryptococcal meningitis (CM) contributes up to 15% of HIV-related mortality and is the leading cause of HIV-related meningitis in many African low- and middle-income countries (LMICs). In resource-limited settings, mortality remains approximately 70% at three months, with many LMICs having no access to proper diagnostics and essential medicines, like flucytosine. Despite significant research progress in past years leading to newly revised World Health Organization (WHO) CM treatment guidelines, important challenges for deployment and scale-up persist. Short course amphotericin B (AmB) with 5FC; two weeks' 5FC + fluconazole; and CrAg screening and pre-emptive therapy, are three mortality-reducing strategies included in the 2018 WHO CM guidelines as the new gold standard of care and treatment in LMICs. Studies show that these 5FC regimens for confirmed CM cases can halve the mortality rates seen with the current use of fluconazole. Reducing the persistent mortality of people living with HIV (PLHIV) with Advanced HIV Disease (AHD), which affects approximately one-third of patients presented to care in LMICs, is an urgent goal as per WHO's 2017 guidelines on AHD management to achieve Sustainable Development Goals (SDGs).

6. Check the Participating Organisations first to decide whether the project is relevant to our database:

Improved flucytosine formulation for the treatment of meningitis in advanced HIV disease
Reference: RIA2018CO-2516

Acronym: 5FC HIV-Crypto

Project Website:

Coordinating Organisation: Drugs for Neglected Diseases Initiative (DNDi), Switzerland

Coordinating Organisation Website: https://dndi.org/

Name of Coordinator: Dr Isabela Ribeiro

Participating Organisations:
- St. George's Hospital Medical School, United Kingdom
- Luxembourg Institute of Health (LIH), Luxembourg
- National Institute for Medical Research - Tanzania (NIMR), Tanzania, United Republic of
- FARMOVS (Pty) Ltd, South Africa
- University of North Carolina Project-Malawi (UNC Project-Malawi), Malawi

Grant Amount: € 3,537,000.00

Start Date: Jul 1, 2020

End Date: Dec 31, 2024

Duration (months): 54

Abstract:
Cryptococcal meningitis (CM) contributes up to 15% of HIV-related mortality and is the leading cause of HIV-related meningitis in many African low- and middle-income countries (LMICs). In resource-limited settings, mortality remains approximately 70% at three months, with many LMICs having no access to proper diagnostics and essential medicines, like flucytosine. Despite significant research progress in past years leading to newly revised World Health Organization (WHO) CM treatment guidelines, important challenges for deployment and scale-up persist. Short course amphotericin B (AmB) with 5FC; two weeks' 5FC + fluconazole; and CrAg screening and pre-emptive therapy, are three mortality-reducing strategies included in the 2018 WHO CM guidelines as the new gold standard of care and treatment in LMICs. Studies show that these 5FC regimens for confirmed CM cases can halve the mortality rates seen with the current use of fluconazole. Reducing the persistent mortality of people living with HIV (PLHIV) with Advanced HIV Disease (AHD), which affects approximately one-third of patients presented to care in LMICs, is an urgent goal as per WHO's 2017 guidelines on AHD management to achieve Sustainable Development Goals (SDGs).
The 5FC HIV-Crypto project aims to develop and make available in high cryptococcal disease burden African LMICs an improved, affordable, easy-to-administer sustained-release 5FC (SR 5FC) formulation to increase

7. If the project is relevant use the Project Name + Acronym for the Activity Column

8. Transfer all the relevant information to our database, including Project ID, Start and End date, Participating Organisations and corresponding locations, Project Website, Coordinating Organisation and Coordinating Researcher

9. Tag the project with the correct keywords in the Activity Type, Activity Outputs, Target Species, Topic, and Research Field columns based on your understanding of the project abstract

9.5.3 Benefits of this Source:

The projects can be easily filtered for currently active projects.

The EDCTP is specifically focused on projects in Africa or Europe with African collaborators so most projects fulfil at least one of our selection criteria

A lot of the projects (not all) are also topic-wise relevant for our database

The database lists most of the information that we are interested in for our database

9.5.4 Shortcomings of this source:

Only one disease/topic can be selected at a time so sequential searches are necessary

Project details only list the Coordinating researcher but no other lead researchers at the participating organisations

10 Survey

11 References