Research Information Gateway Handbook

EcoHealth Alliance

2023-02-17

Contents

1	Abo	About 5			
2	Intr	Introduction 7			
3	Tools				
	3.1	Airtable	9		
	3.2	$R \ldots \ldots$	9		
4	Dat	Database 11			
5	Sources				
	5.1	UKRI	13		
	5.2	Wellcome Trust	14		
	5.3	National Insitutes of Health	14		
	5.4	National Science Foundation	14		
	5.5	Defense Advanced Research Projects Agency	14		
6	Updates 15				
	6.1^{-}	Wellcome Trust Grant Funding Data	15		
	6.2	ClinicalTrials.gov	16		
	6.3	Public journal/research databases	21		
	6.4	GEPRIS	22		
	6.5	EDCTP	25		
7	Sur	Survey 29			

4 CONTENTS

About

This is a handbook created as a how-to manual for the development and maintenance of a One Health Research Information Gateway for the African continent.

Introduction

EcoHelth 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.

Tools

The Research Information Gateway (RIG) is currently using the following set of tools for database development and maintenance.

3.1 Airtable

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.

3.2 R

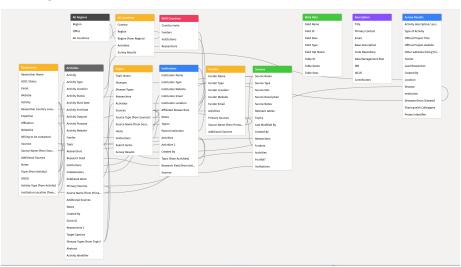
R is a free and open-source programming language and software environment for statistical computing and graphics. It was first developed in the early 1990s by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is now widely used by statisticians, data analysts, and researchers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

Database

The current database is built using Airtable. The current database has the following schema:



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.

5.1 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.

5.2 Wellcome Trust

The Wellcome Trust is a charitable foundation focused on health research based in London, in the United Kingdom. It was established in 1936 with legacies from the pharmaceutical magnate Henry Wellcome to fund research to improve human and animal health. The aim of the Trust is to "support science to solve the urgent health challenges facing everyone.". In 2012, the Wellcome Trust was described as the United Kingdom's largest provider of non-governmental funding for scientific research, and one of the largest providers in the world.

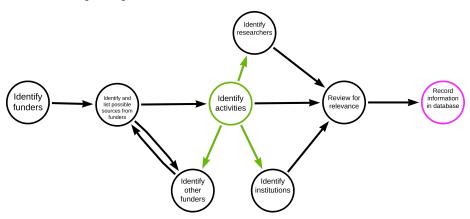
5.3 National Institutes of Health

5.4 National Science Foundation

5.5 Defense Advanced Research Projects Agency

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.

6.1 Wellcome Trust Grant Funding Data

6.1.1 General Information

The Wellcome Trust is very transparent about its funding efforts and makes information about its funded projects available in several ways including on their own website and the World RePORT: (https://worldreport.nih.gov/wrapp/#/search?searchId=62a70c4bfe43c863cea97f29). I found the downloadable spreadsheet of funds awarded between 01st of October 2005 and 4th of May 2022 the most useful: https://cms.wellcome.org/sites/default/files/2022-05/Wellcome-grants-awarded-1-October-2005-to-04-05-2022.xlsx

6.1.2 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.

- 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

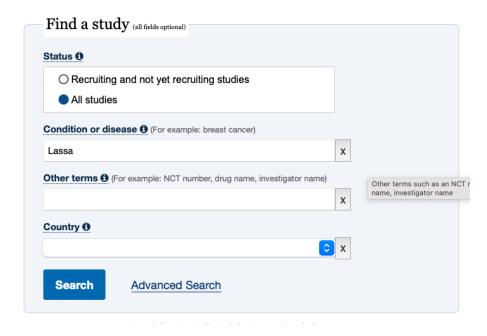
Note: This method is only able to detect projects/activities where an African organisation itself holds the grant. It does not detect projects where African researchers are involved as collaborators. The spreadsheet does not list collaborators on projects, so it's yet to be determined how we will identify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

6.2 ClinicalTrials.gov

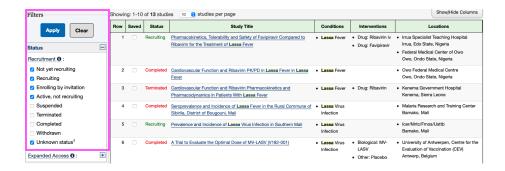
6.2.1 How to update

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

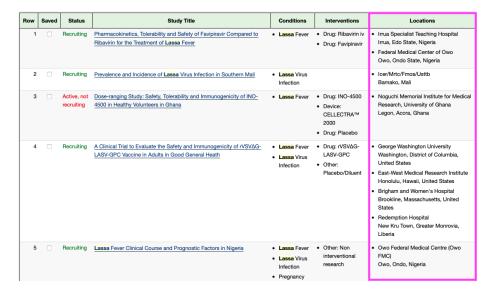
1. Start with searching a disease/topic of interest



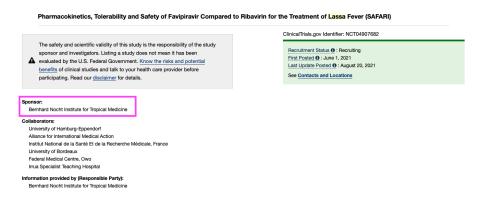
2. On the results page apply the following filters to look for active studies



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



- 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



- 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

Study Design	Go to 🔻
Study Type 0:	Interventional (Clinical Trial)
Estimated Enrollment 6 :	40 participants
Allocation:	Randomized
Intervention Model:	Parallel Assignment
Intervention Model Description:	Exploratory, prospective, controlled, multisite, open label, randomized clinical trial with two arms
Masking:	None (Open Label)
Primary Purpose:	Treatment
Official Title:	Pharmacokinetics, Tolerability and Safety of Favipiravir Compared to Ribavirin for the Treatment of Lassa Fever: A Randomized Controlled Open Label Phase
	Clinical Trial
Actual Study Start Date 6 :	July 30, 2021
Estimated Primary Completion Date 6 :	April 2022
Estimated Study Completion Date 6 :	April 2022

- 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 clinical trials 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

6.3 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

6.3.1 Example of a successful search:

(zoonoses OR zoonotic disease OR zoonotic illness) and (africa*) and (surveillance OR tracking OF

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

6.3.2 Example of PubMed search for surveillance activities for Brucellosis:

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

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

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.

6.3.3 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.

6.3.4 Some relevant journals/databases:

- Zoonoses & Public Health from Wiley Online Library: https://onlinelibrary.wiley.com/action/doSearch?SeriesKey=18632378&sortBy=Earliest
- 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=PLoSNTD

6.4 GEPRIS

6.4.1 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 6.4. GEPRIS 23

for the promotion of science and research in the Federal Republic Germany. In 2019, the DFG had a funding budget of $\in 3.3$ billion.

6.4.2 How to Use:

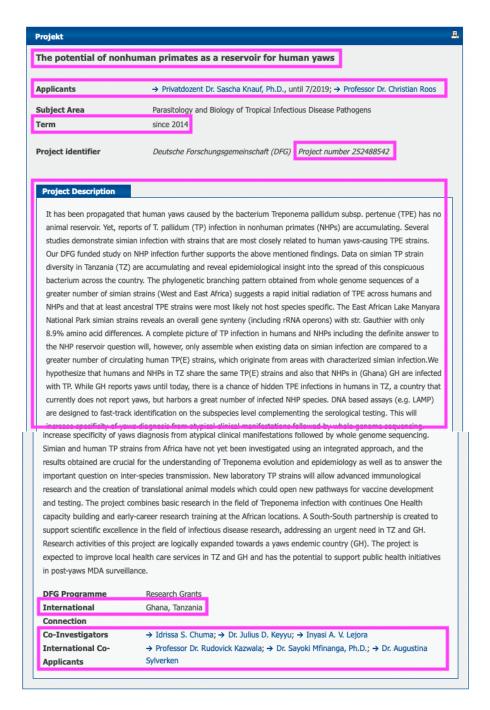
The database can be accessed here: https://gepris.dfg.de/gepris/OCTOPUS?l anguage=en&task=showSearchSimple

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



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)

6.5. EDCTP 25

6.4.3 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.

6.4.4 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.

6.5 EDCTP

6.5.1 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 Activities table for the Project ID, as this might be helpful to identify duplicates and remove them automatically.

6.5.2 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.

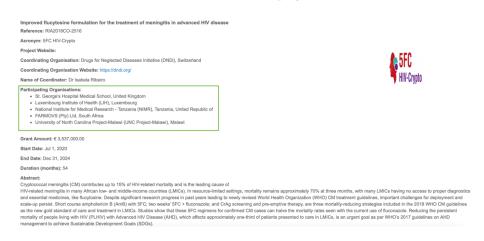
- 1. Go to Status of Project and select the filter Active
- 2. Go to Classification and select one of the following filters:
 - Co-Infections
 - COVID-19
 - Cysticercosis/Taeniasis
 - Diagnostics
 - Diarrhoeal Diseases
 - Drugs
 - Emerging Infections, incl. Ebola, Lassa
 - Epidemiology
 - HIV
 - Human 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

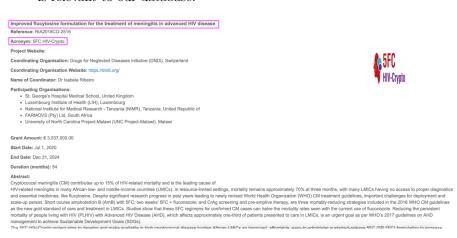
- 3. Once one Classification was selected click on search
- 4. Change from Show Map to Show List -> this makes it easier to systematically look through the projects
- 5. On the results list you can see the location of the coordinating organisation, but even if it is not in an African country, it is worth checking the project

6.5. EDCTP 27

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



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



- 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
- 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

6.5.3 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

6.5.4 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

Survey