The Pandemic PACT Dataset

A Review

Ernest Guevarra

20 November 2024

Contents

List of Figures

List of Tables

1 Introduction

This document provides a review of the Pandemic PACT dataset. Specifically, this document:

1. Gives a brief description of what the Pandemic PACT dataset is about; and,

2. Reviews the fields available in the Pandemic PACT dataset and matches it to the fields available in EcoHealth Alliance's Research Information Gateway database.

2 Pandemic PACT Overview

The **Pandemic PACT** project is an initiative led by the Pandemic Sciences Institute at the University of Oxford. It aims to monitor and analyse global research efforts in pandemic preparedness, focusing on the development of diagnostics, therapeutics, vaccines, and other countermeasures for infectious diseases.

The project curates and tracks an extensive dataset, which include research projects, funding, and outputs from international studies. The objective is to provide insights into how resources are allocated and to identify gaps in global health security efforts. It emphasises equitable access to the benefits of research and innovation, particularly for vulnerable populations in lower-income regions.

2.1 The dataset

The data and tools developed through the Pandemic PACT project are hosted on platforms like Figshare[1] and on its website[2], enabling open access for researchers to analyse trends and contribute to pandemic readiness globally.

---

[1] Data available from Figshare is more raw data in wide data format/structure.

[2] Data available from website is derived from the raw datasets from Figshare but is already structured in a more accessible structure for general users.

The Pandemic PACT dataset is distributed with an open license under the

2.1.1 Curation process

The process of curating the Pandemic PACT dataset is described in their protocol and is summarised in Figure 1.
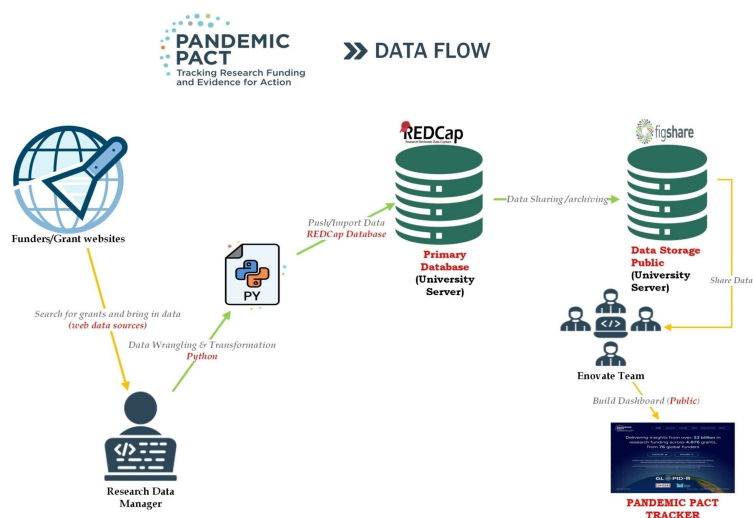
Figure 1: Pandemic PACT Data Flow
Source: https://www.pandemicpact.org/about/our-data

2.2 Data governance

In the Pandemic PACT's published protocol, it states its data governance procedures:

> For the purposes of the Pandemic PACT project, we have collected data on researchers and their research outputs using names and Open Researcher and Contributor IDs (ORCIDs) ("personal" data). The University of Oxford are the 'data controller' for these data, which means we decide how to use it and are responsible for looking after it in accordance with the UK General Data Protection Regulation and

associated data protection legislation. We share data with anyone who wishes to download and re-use the information under a CC-BY licence. We will only retain data for as long as we need it to meet our purposes, including any relating to legal, accounting, or reporting requirements. Data will be held securely in accordance with the University's policies and procedures. Further information is available on the University's Information Security website where information on rights in relation to personal data are explained.

2.2.1 Sources

The Pandemic PACT lists the following funders as their sources of information found in the dataset (see Figure 2).

Funder	COVID-19					Other New Disease				
	2020	2021	2022	2023	2024	2020	2021	2022	2023	2024
African Academy of Sciences (AAS)	No	Yes	No	No	No	No	No	No	No	No
Agence nationale de recherche sur le sida et les hépatites virale National Agency for AIDS Research (ANRS)	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No
Bill & Melinda Gates Foundation	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
British Academy	Yes	Yes	No	No	No	No	No	No	No	No
Bundesministerium für Bildung und Forschung - German Federal Ministry of Education and Research (BMBF)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Canadian Institutes of Health Research (CIHR)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Coalition for Epidemic Preparedness Innovations (CEPI)	Yes	No	No	No	No	Yes	No	No	Yes	Yes
Consejo Nacional de Ciencia y Tecnología (CONACYT)	Yes	No	No	No	No	No	No	No	No	No
Department for Environment, Food and Rural Affairs (DEFRA)	Yes	Yes	No	No	No	No	No	Yes	No	No
Department of Health and Social Care / National Institute for Health and Care Research (DHSC-NIHR)	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
European Commission	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign, Commonwealth & Development Office (FCDO)	Yes	No	No	No	No	No	No	No	No	No
Fundação de Amparo à Pesquisa do Estado de São Paulo - São Paulo Research Foundation (FAPESP)	Yes	Yes	Yes	No	No	Yes	Yes	No	No	No
Global Alliance for Vaccines and Immunization (GAVI)	No	Yes	No	No	No	No	No	No	No	No
Indian Council of Medical Research (ICMR)	Yes	Yes	No	No	No	No	No	No	No	No
Institut Pasteur International Network (IPIN)	Yes	No	No	No	No	No	No	No	No	No
International Development Research Centre (IDRC)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Japan Agency for Medical Research and Development (AMED)	Yes	No	No	No	No	Yes	No	No	No	No
Ministerio de Ciencia, Tecnología e Innovación Ministry of Science, Technology and Innovation (MINCYT)	Yes	No	No	No	No	No	No	No	No	No
Ministry of Health – Italy	Yes	No	No	No	No	No	No	No	No	No
National Council for Science and Technology (NCST) Rwanda	No	Yes	No	No	No	No	No	No	No	No
National Health and Medical Research Council (NHMRC)	Yes	Yes	No	No	No	Yes	Yes	No	No	No
National Institute of Health Carlos III [El Instituto de Salud Carlos III (ISCIII)	Yes	No	No	No	No	No	No	No	No	No
National Institutes of Health (NIH)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
National Research Foundation of Korea (NRF)	Yes	No	No	No	No	No	No	No	No	No
Netherlands Organisation for Health Research and Development (ZonMW)	Yes	Yes	No	No	No	No	No	No	No	No
Research Council of Norway (RCN)	Yes	Yes	No	No	No	No	No	No	No	No
South African Medical Research Council (SAMRC)	No	Yes	No	No	No	No	Yes	No	No	No
Swiss National Science Foundation (SNSF)	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	No
UK Research and Innovation (UKRI)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
WHO	Yes	Yes	No	No	No	No	No	No	No	No
Wellcome Trust	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No

Figure 2: Pandemic PACT data sources
Source: https://www.pandemicpact.org/about/our-data

3 Pandemic PACT dataset fields

As of 2024-11-20, the Pandemic PACT dataset contains **19172** grant records of various research activities relevant to diseases with pandemic potential.

3.1 Description of Pandemic PACT dataset fields compared with EHA database

Table 1 presents the various fields in the Pandemic PACT dataset[3]. A column labelled *Similar/related fields to the EHA database* has been added to indicate whether the particular Pandemic PACT field has a related/similar/same field in the EHA database.

Table 1: Pandemic PACT dataset fields compared to EHA database fields

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
PACTID	string	Non-standard, assigned internally	A combination of a letter character and numbers	Pandemic PACT-specific but EHA database has its own unique identifier	

---

[3]As presented in the Pandemic PACT protocol - https://wellcomeopenresearch.org/articles/9-156

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Grant in Scope	binary	Non-standard, assigned internally			This field indicates whether grant is in scope as per Pandemic PACT protocol; no corresponding field in EHA database
Grant Title Original	text	Non-standard			This is the same as the **activity** field in the `activities` table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Grant Number	text	Non-standard		As assigned by a funder	This is the same as the **grant_id** field in the `activities` table of the EHA database
Grant Amount Original	string	Non-standard			Not found in the EHA database
Grant Currency	string	ISO 4217 code			Not found in the EHA database
Currency Exchange Rate USD	numeric	Non-standard		Calculated using API and code	Not found in EHA database
Grant Amount Converted	numeric	Non-standard			Not found in the EHA database
Grant Type	text	Non-standard			Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Abstract Original	text	Non-standard			This is the same as the **abstract** field in the `activities` table in the EHA database
Abstract English	text	Non-standard			This is the same as the **abstract** field in the `activities` table in the EHA database
Lay Summary	text	Non-standard			Not found in the EHA database
ODA funding used	binary	Non-standard, assigned internally		Official Development Assistance (ODA)	Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Grant Start Month	numeric	MM, ISO standard			Not found as is in the EHA database; related to the **activity_start_date** field in the `activities` table of the EHA database
Grant Start Year	numeric	YYYY, ISO standard			Not found as is in the EHA database; related to the **activity_start_date** field in the `activities` table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Grant End Month	numeric	MM, ISO standard			Not found as is in the EHA database; related to the **activity_end_date** in the `activities` table of the EHA database
Grant End Year	numeric	YYYY, ISO standard			Not found as is in the EHA database; related to the **activity_end_date** in the `activities` table of the EHA database
Publication Month of Award	numeric	MM, ISO standard			Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Publication Year of Award	numeric	YYYY, ISO standard			Not found in the EHA database
Grant Type	text	Non-standard	New Grant, Grant Extension		Not found in the EHA database
Study Subject	Text, Boolean	MESH Terms	Animals, bacteria, human populations, disease vectors, viruses, environment, other, unspecified, not applicable		Similar to the **target_species** field in the `activities` table of the EHA database
Ethnicity	text, Boolean	Standard, UK Census	Asian, Black, White, Mixed, other, unspecified, not applicable	Optional field, populate if the grant is for research involving a specific ethnic group	Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Age Groups	Text, Boolean	MESH Terms modified	Adolescent 13–17 yrs; Adults, 18+; Children, 1–12 yrs; Infants, 1mth–1yr; Newborn (<1mth); Older adults, 65+; Unspecified, not applicable	Optional field, populate if the grant is for research involving a specific age group	Not found in the EHA database
Rurality	text, Boolean	MESH terms, modified	Rural population/setting; suburban population/setting; urban population/setting; other, unspecified, not applicable	Optional field, populate if the grant is for research on urban or rural populations or settings	Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Vulnerable Populations	Text, Boolean	MESH Terms, modified	Disabled persons, drug users, Internally Displaced and Migrants, Indigenous People, Sexual and gender minorities, Prisoners, Sex workers, Smokers, Women, Pregnant women, Individuals with multimorbidity, Minority communities unspecified, vulnerable populations unspecified, other, unspecified, not applicable	Optional field, populate if the grant is for research involving a specific vulnerable population group	Not found in the EHA database

13

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Occupational Groups	Text, Boolean	MESH terms modified	Farmers, Emergency Responders, Military Personnel, Social workers, Caregivers, Health Personnel, Hospital personnel, Nurses and Nursing Staff, Physicians, Dentists and dental staff, Vets, Volunteers, other, unspecified, not applicable	Optional field, populate if the grant is for research involving a specific occupational group	Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Study Type	Text, Boolean	Non-standard	Clinical, Non-clinical, other, unspecified, not applicable	If clinical is selected, then there is an option to select a clinical trial phase and design and record this information in a new field. If non-clinical is selected, then there is an option to choose a report or literature review in a new field	Similar to **activity_type** field in the **activities** table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Disease	numeric	Standard, SNOMED code	See the list of diseases at https://termbrowser.nhs.uk/		Similar/related to the **diseases** field in the `topics` table of the EHA database
Pathogen	numeric	Standard, SNOMED code	See the list of diseases at https://termbrowser.nhs.uk		Similar/related to the **topic_name** field of the `topics` table of the EHA database
Funder	text	Standard, Cross-Ref Open Funder Registry	https://www.crossref.org/services/funder-registry/		This is the same as the **funder_name** of the `funders` table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Funder Region	text	Standard, WHO region	https://en.Wikipedia.org/wiki/List_of_WHO_regions	The region was assigned automatically based on the country of the funding organisation as listed in the global standard list	Similar to the **funder_region** field of the `funders` table of the EHA database but instead of WHO regions, the EHA database shows AU regions. However, the `funders` table has a corresponding WHO region classification

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Funder Country	numeric	ISO 3166-1 numeric	https://www.crossref.org/services/funder-registry/	Country information was pulled from the Cross-Ref Open Funder Registry	The same as the **who_country_id** of the **funders** table of the EHA database
Funder Acronym	text	Standard, Cross-Ref Open Funder Registry		Acronym was pulled from the Cross-Ref Open Funder Registry	The same as the **funder_short_name** field of the **funders** table of the EHA database
Investigator Title	text	Non-standard			Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Investigator First Name	text	Non-standard			Not found as is in the EHA database; related to the **researcher_name** field in the `researchers` table of the EHA database
Investigator Last Name	text	Non-standard			Not found as is in the EHA database; related to the **researcher_name** field in the `researchers` table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Investigator ORCID	string	Standard, ORCID ID number		Optional field. Researchers manually searched and entered the ORCID using the first and last name of the awardee.	This is the same as the **orcid** field in the `researchers` table in the EHA database
ROR ID	string	Standard, ROR ID	https://ror.org/	Research Organisation Registry (ROR ID) for research institution	Not found in the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Institution Name	text	Standard, ROR list of research institutions	https://ror.org/		This is the same as the **institution_name** field in the `institutions` table of the EHA database
Institution Country	text	Standard, ROR list of research institutions	https://ror.org/		This is the same as the **who_country_id** field of the `institutions` table of the EHA database
Institution Country ISO	numeric	ISO 3166-1 numeric	https://www.iso.org/iso-3166-country-codes.html		This is the same as the **country_code** field of the `countries` table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Research Institution Region	text	Standard, WHO region		The region was assigned by a data manager using information from the ROR list	This is the same as the **who_region_name** field in the `countries` table of the EHA database
Partner Organisation Name	text	Non-standard		Information on the partner organisation is added if available in the grant abstract	This is similar to the **institution_name** of the `institutions` table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Research Location Country	text	Non-standard		Information on the location of research is added if available in the grant abstract. Otherwise, we used the country where the Research Institution is based	This is similar to the **country_id** field in the `activities` table of the EHA database
Research Location Country ISO	numeric	Standard, ISO 3166-1 numeric code		This is similar to the **country_id** field in the `activities` table of the EHA database	

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Research Location Region	text	Standard, WHO Region		Assigned based on the location of research is such information is available in the grant. Otherwise, we used the region where the Research Institution is based	This is similar to the **who_region_name** field in the `activities` table of the EHA database
Tags	Text, Boolean	Non-standard	Data Management and Data Sharing, Digital Health, Innovation, Gender	The tags were assigned by researcher who reviewed the grants	Not found as is in the EHA database; related to the fields in the `topics` table of the EHA database

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Research and Policy Roadmaps	Text, Boolean	Non-standard	100 Days Mission, WHO Surveillance, ESSENCE re- for Health	Mapping to selected roadmaps was done by researcher reviewing the grants	Not found in the EHA database
Primary Research Category	string	Non-standard	12 broad research categories, each has a list of subcategories		Researchers reviewed each grant and assigned a broad research category and subcategory. Multiple values permitted

Variable name	Data format	Data Standard	Values	Notes	Similar/related field to EHA database
Secondary Research Category	string	Non-standard	12 broad research categories, each has a list of subcategories		Not found as is in the EHA database; similar to but not quite the same as **acrtivity_outputs** and **activity_purpose** fields of the `activities` table of the EHA database

Out of the 50 fields in the Pandemic PACT dataset, **28** (56%) have similar/related/same field in the EHA database.

The following fields in the EHA database has no similar/related/same field in the Pandemic PACT dataset (see Table 2).

Table 2: EHA database fields compared to Pandemic PACT dataset fields

Field name	Table name	Data format	Data standard	Values
activity_status	activities	text	Non-standard	Active or completed
funder_type	funders	text	Non-standard (ontology created by EHA)	Government, Academic, Pharmaceutical, Foundation, Research Institute, Charity, Inter-governmental, International Organisation, International Non-governmental Organisation, Public-Public Partnership
funder_email	funders	text	Non-standard	
funder_website	funders	text	Non-standard	

Field name	Table name	Data format	Data standard	Values
institution_type	institutions	text	Non-standard (ontology created by EHA)	University, Industry, Government Agency, Hospital, Research Institute, Charity, Partnership of Institutes, Non-Government Institution
institution_email	institutions	text	Non-standard	
institution_website	institutions	text	Non-standard	
email	researchers	text	Non-standard	
website	researchers	text	Non-standard	

Table 3 summarises the per table field coverage of the EHA datasets by the Pandemic PACT dataset.

Table 3: Summary of per table field coverage of the EHA database by the Pandemic PACT dataset

Table	No. of Fields	No. of Fields Covered by Pandemic PACT	%
activities	33	33	100%
au_countries	3	1	33%
countries	9	6	66%
funders	8	5	63%

Table	No. of Fields	No. of Fields Covered by Pandemic PACT	%
institutions	8	5	63%
researchers	19	17	89%
topics	6	6	100%
who_countries	2	2	100%
Overall	88	75	85%