Incoming Datasets Questionnaire

Outbreak Data Analysis Platform

Table of contents

1 Introduction

This document is intended for gathering metadata for new datasets to be added to ODAP.

2 Questionnaire

2.1 Summary

2.1.1 Please insert the title of the Data Set:

2.1.2 Please include an abstract for the dataset:

(The abstract should provide a clear and brief descriptive signpost for researchers who are searching for data that may be relevant to their research, and should allow the reader to determine the scope of the data collection and accurately summarise its content.)

2.1.3 Please list any keywords you would like to be used when describing your dataset:

(This is to help researchers locate your specific dataset within the collection of ODAP datasets.)

2.1.4 Please provide the existing DOI for the dataset, if available:

2.1.5 If able, please provide the DOIs of other datasets that have previously been linked to this dataset and their availability:

2.2 Coverage

2.2.1 What is the geographic area covered by the dataset?

2.2.2 What is the sample size in the Data Sets?

2.2.3 if known, Please provide the typical time span that a patient appears in the dataset?

2.2.4 How does the data-set align with the patient pathway, and are there limitations that the dataset may have with respect to pathway coverage?

(This could include if the dataset is from a single specialty or area, a single tier of care, linked across two tiers (e.g. primary and secondary care), or an integrated care record covering the whole patient pathway.)

2.2.5 Are the following categories of data included in your data set?

- ☐ Personally Identifiable Data
- ☐ De-personalised Data
- ☐ Anonymous Data
- ☐ Other – please specify:

2.2.6 Are the following types of data included in the Data Sets?

- ☐ Clinical records
- ☐ Electronic patient records
- ☐ Clinical registries
- ☐ Clinical trial records
- ☐ Health system operational data
- ☐ Digital device data (e.g., app, sensor, wearable)

- ☐ Environmental monitoring data
- ☐ Administrative and socio-economic data
- ☐ Genomics or genetic data
- ☐ Imaging data
- ☐ Geospatial data
- ☐ Government or National statistics
- ☐ Patient or public survey data
- ☐ Unconsented data
- ☐ Other – please specify:

2.3 Provenance

2.3.1 What was the original purpose for gathering this data-set?

- ☐ Study, data collected for a specific research study.
- ☐ Disease registry, data collected as part of a disease registry.
- ☐ Trial, data collected for as part of a clinical trial.
- ☐ Care, data collected as part of routine clinical care.
- ☐ Audit, data collected as part of an audit programme
- ☐ Administrative, data collected for administrative and management information purposes
- ☐ Financial, data collected either for payments or for billing
- ☐ Other, data collected for other purpose

2.3.2 What was the original source for this data-set?

- ☐ EPR, data extracted from electronic patient record
- ☐ Electronic survey, data has been extracted from electronic surveys
- ☐ LIMS, data has been extracted from a laboratory information management system
- ☐ Paper based, data has been extracted from paper forms
- ☐ Free text NLP, data has been extracted from unstructured free text using natural language processing
- ☐ Machine generated, data has been machine generated i.e. imaging
- ☐ Other, data has been extracted.

2.3.3 Where was the original collection situation for this data-set?

- ☐ Clinic, specific clinic such as antenatal clinic
- ☐ Primary care, general medical practitioner practice
- ☐ Accident and emergency, accident emergency department
- ☐ Outpatients, outpatient care
- ☐ In-patients, in-patient care
- ☐ Services, services such as drug misuse or blood transfusion
- ☐ Community, community settings
- ☐ Home, home setting
- ☐ Private, private medical clinic
- ☐ Pharmacy, pharmacy
- ☐ Other, other setting

2.3.4 Is the dataset is Continuous with no known end date to the time period it covers?

- ☐ Yes, and periodic updates will be provided to ODAP (skip questions 3.22 & 3.23)
- ☐ Yes, but only a snapshot of the data is being provided to ODAP (skip questions 3.20 & 3.21)
- ☐ No, the dataset has a fixed end date (skip questions 3.20 & 3.21)

2.3.5 What is the start of the time period that the dataset provides coverage for?

(If there are multiple cohorts in the dataset with varying start dates, please provide the earliest date and use the description or the media attribute to provide more information.)

2.3.6 Please indicate the frequency that updated versions of the dataset will be provided to ODAP?

2.3.7 Please indicate the typical time-lag between an event and the data for that event appearing in the dataset?

- ☐ Less 1 week, typical time lag of less than a week
- ☐ 1-2 weeks, typical time lag of one to two weeks
- ☐ 2-4 weeks, typical time lag of two to four weeks
- ☐ 1-2 months, typical time lag of one to two months
- ☐ 2-6 months, typical time lag of two to six months
- ☐ 6 months plus, typical time lag of more than six months
- ☐ Variable, variable time lag
- ☐ Other, other time lag

2.3.8 Please indicate the release date for the specific version of the dataset being provided?

(If there are multiple cohorts in the dataset with varying start dates, please provide the earliest date and use the description or the media attribute to provide more information.)

2.3.9 What is the end of the time period that the dataset provides coverage for?

(If there are multiple cohorts in the dataset with varying end dates, please provide the latest date.)

2.3.10 Please explain the legal basis on which the Data Sets were collected:

2.3.11 Was ethical review required under applicable laws, rules, regulations or guidance for the collection of the data Sets?

- ☐ No
- ☐ Yes, but review incomplete
- ☐ Yes, and review complete – please provide details for who carried out the ethical review:

2.3.12 Is ethical review required under applicable laws, rules, regulations or guidance for the Research Purpose?

- ☐ No
- ☐ Yes, but review incomplete
- ☐ Yes, and review complete – please provide details for who carried out the ethical review:

2.3.13 Did you obtain consent from the data subjects to collect their data and use the Data Sets for the Research Purposes?

☐ Yes
☐ No

2.3.14 Are there any other ethical issues we should be aware of in relation to the Data Sets?

☐ No
☐ Yes – please provide further details:

2.3.15 Has there been any public, patient or lay person input or representation related to the collection or use of the data Sets?

☐ No
☐ Yes – please briefly describe the process:

2.4 Accessibility

2.4.1 Please provide an indication of the permissions use for this dataset:

- ☐ No restriction, there are no restriction on use.
- ☐ General research use, allowed for general research use for any research purpose
- ☐ Research use only, use is limited to research purposes (e.g., does not include its use in clinical care).
- ☐ Research specific restrictions, use is limited to studies of a certain research type.
- ☐ Genetic studies only, use is limited to genetic studies only (i.e., no phenotype-only research)
- ☐ No general methods research, methods development research only within the bounds of other use limitations (e.g., development of software or algorithms).
- ☐ No commercial use, there is a restriction on providing this dataset to commercial entities or for use in commercial research
- ☐ No linkage, there is a restriction on linking to any other datasets
- ☐ Collaboration required, new requestors must agree to collaboration with the primary study investigator(s).
- ☐ Ethics approval required, new requestors must provide documentation of local IRB/ERB approval.
- ☐ Geographical restrictions, use is limited to within a specific geographic region.
- ☐ Institution specific restrictions, use is limited to use within an approved institution.
- ☐ Not for profit use, use of the data is limited to not-for-profit organizations and not-for-profit use, non-commercial use.
- ☐ Project specific restrictions, use is limited to use within an approved project.
- ☐ Publication moratorium, new requestors agree not to publish results of studies until a specific date.
- ☐ Publication required, new requestors agree to make results of studies using the data available to the larger scientific community.
- ☐ Publication restrictions, there are restrictions on publication of research outcomes specific to this dataset.
- ☐ Return to database or resource, new requestors must return derived/enriched data to the database/resource.
- ☐ Time limit on use, use is approved for a specific number of months.
- ☐ Active trial data, there are restrictions on the use of this data arising from active trials.
- ☐ Restricted variables, there are specifically restricted variables within the dataset.
- ☐ User specific restriction, use is limited to use by approved users.
- ☐ Other, if there are other limitations on the use of this dataset please provide further details

2.4.2 Please list any languages used in the data-set or accompanying metadata:

2.4.3 If multiple formats are available, please specify. Text, CSV, Audio, image, message, model, multipart, video, - etc.:

2.5 Pre-existing associated data, documents, and tools

2.5.1 Please indicate if any derived datasets or predefined extracts are provided with this dataset, and the type of derivation available (if any):

2.5.2 Please indicate if any analysis tools or models that have been created for this dataset are being available for further use, and where they are located:

2.5.3 Please provide the filenames and locations for any pre-existing observations for this dataset such as number of persons, events, restricted variables, or findings:

2.5.4 Please provide the filenames and locations for any available data dictionaries with named tables and variables available for this dataset:

2.5.5 Please provide the filenames and locations for any available publication or attribution documents: