Akrivia Health Data Dictionary

March 2026

This data dictionary is provided to prospective clients and partners for evaluation purposes only. It describes the structure, provenance, and scope of Akrivia Health's **anonymised, research-ready dataset** derived from secondary care electronic health records (EHRs).

To make further enquiries, including to access the data or procure our related services, please contact contact@akriviahealth.com.

Introduction

Terminology

Term	Meaning
akr_id	An anonymised, stable patient identifier used across all tables. This is hashed at a project level when provisioned to researchers.
EHR	Electronic health record.
NLP	Natural language processing.
OMOP Mapped	Indicates where fields have been mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model.
SAS	Structured-at-source.

SQL Category & Type

SQL Category	SQL Type	Description
Numeric	int	From -2147483648 to 2147483647
	int8	From -9223372036854775808 to +9223372036854775807
	int4	From -2147483648 to +2147483647

Date/Time	date	Stores date in the format YYYY-MM-DD
	datetime	Stores date and time information in the format YYYY-MM-DD hh:mm:ss
Character / String	varchar	Variable length storage with maximum length of 8,000 characters
	text	Variable unlimited length
UUID	uuid	Unique 128-bit (anonymised) identifier

Output Data Type

Output Data Type	Source Data Type	Processing Method
SAS	Numeric / categorical fields in the source EHR	Identification and harmonisation of similar source data tables across different health care organisations and EHR systems
NLP	Free-text fields in the source EHR	Development of natural language processing models to extract and classify features of interest from free-text clinical notes

Data Availability

Direct data access is provisioned on a **project-by-project basis**. All projects are subject to approval by our **Data Access Committee**.

Some data fields may be subject to **additional safeguards** for information governance reasons. This is most common where a higher degree of specificity in time and space is required for a project. In all cases, patient-level data is only provisioned within our tightly controlled **secure data environment (SDE)**.

For more information, contact contact@akriviahealth.com.

Table Index

Table 1: Patient Details

Column Name	SQL Type	Output Data Type	Description / Comments
match_exists_in_biobank	boolean	SAS	Patients with a matching record in UK Biobank: [v] = true, [] = false
gender_value	text	SAS	Gender: Male, Female, Other/Non-binary, Not known
date_of_birth	date	SAS	Date of birth [YYYY/MM/DD] aggregated to month of birth (all day values set to 1st of the month)
date_of_death	date	SAS	Date of death [YYYY/MM/DD] aggregated to month of birth (all day values set to 1st of the month)
date_of_registration	date	SAS	Date of death [YYYY/MM/DD] aggregated to month of birth (all day values set to 1st of the month)
bt_18_65_flag	boolean	SAS	Patients aged between 18 and 65: [v] = true, [] = false
lt_18_flag	boolean	SAS	Patients aged 18 and younger: [v] = true, [] = false
gt_65_flag	boolean	SAS	Patients aged 65 and older: [v] = true, [] = false
marital_status_value	text	SAS	Marital status: Married/Civil Partnership, Separated, Single, Divorced/Civil Partnership Dissolved, Cohabitating, Widowed, Not known
ethnicity_value	text	SAS	Ethnicity: White, Asian, Black, Mixed, Other, Not known
created_date	date	SAS	Date of patient record creation [YYYY/MM/DD]
age_range	text	SAS	10-year age range e.g., 45-50, 50-55, 55-60 etc
age_min	text	SAS	Minimum age in 10-year range e.g., 45, 50, 55
age_max	text	SAS	Maximum age in 10-year range e.g., 50, 55, 60
akr_id	text	SAS	Unique ID
opc_match	text	SAS	primary care linkage data available for healthcare organisations who have signed [ ] - False, there's no primary care linkage data available [v] - True, there is primary care linkage data available [NULL] - not refreshed primary care linkage for that healthcare organisation

Table 2: Referrals

Column Name	SQL Type	Output Data Type	Description / Comments
referral_received_date	date	SAS	Date on which the referral was received by the Health Care Organisation [YYYY/MM/DD]
referral_accepted_date	date	SAS	Date on which the referral was accepted by the Health Care Organisation [YYYY/MM/DD]
referral_discharge_date	date	SAS	Date on which the patient was discharged by the Health Care Organisation [YYYY/MM/DD]
service_type	text	SAS	464 types of service a patient is referred to (for full list, contact contact@akriviahealth.com)
akr_id	text	SAS	Unique ID

Table 3: Inpatient Stay

Column Name	SQL Type	Output Data Type	Description / Comments
admission_date	date	SAS	Date of admission to inpatient service [YYYY/MM/DD]
admission_source_value	character varying	SAS	Location from which patient was referred to inpatient service e.g., NHS other hospital provider - ward for general patients, Usual place of residence etc)
admission_method_value	character varying	SAS	Method of admission e.g., Emergency - Accident & Emergency Department, Planned, Booked etc
discharge_date	date	SAS	Date of discharge from inpatient service [YYYY/MM/DD]
discharge_destination_value	character varying	SAS	Location to which patient is discharged from inpatient service e.g., Local Authority residential accommocation, NHS run care home, Patient died or stillbirth
stay_duration	interval	SAS	Duration of stay in inpatient service e.g., 2 months 5 days; 17 days; 4 years 5 mons 4 days
akr_id	text	SAS	Unique ID

Table 4: SAS Diagnosis

Column Name	SQL Type	Output Data Type	Description / Comments
icd10_code	text	SAS	Code for diagnosis, either ICD-10 or OPCS4 (see coding_scheme below)
icd10_description	text	SAS	ICD-10 label description for diagnosis, or OPCS4 code
icd11_code	text	SAS	ICD-11 code for diagnosis (mapped from ICD-10 code if available)
icd11_description	text	SAS	ICD-11 label description for diagnosis
diagnosis_start_date	date	SAS	Diagnosis start date [DD/MM/YYYY]
diagnosis_end_date	date	SAS	Diagnosis end date [DD/MM/YYY]
entry_date	date	SAS	Date [DD/MM/YYYY] that a diagnosis was entered into the EHR.
coding_scheme	text	SAS	Diagnostic coding scheme: ICD10 or OPCS4
removal_date	text	SAS	Date [DD/MM/YYYY] that a diagnosis was removed from the EHR.
removal_reason_code	text	SAS	Code denoting reason for diagnosis removal, if available, from the following list: 01, CHG, EIE, ERR, RES, CD, DRR02, 02, DIED, DIE, DRR01, DM
removal_reason_description	text	SAS	Code description for diagnosis removal reason [WORK IN PROGRESS]
struc_diag_duration_days	text	SAS	Duration of diagnosis e.g., 12 months
akr_id	text	SAS	Unique ID

Table 5: Deprivation

Column Name	SQL Type	Output Data Type	Description / Comments
datecode	date	SAS	Date of statistical release of The English Indices of Deprivation 2019 (IoD2019): 2019
measurement	text	SAS	Unit of measurement: decile and quintile
index_of_multiple_deprivation	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the index of multiple deprivation
income_deprivation	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the income domain of deprivation
employment_deprivation	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the employment domain of deprivation
education_skills_and_training	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the education domain of deprivation
health_deprivation_and_disability	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the health domain of deprivation
crime_domain	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the crime domain of deprivation
barriers_to_housing_and_services	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the barriers to housing and services domain of deprivation
living_environment_deprivation	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the living environment domain of deprivation
income_deprivation_affecting_children	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the index of income deprivation affecting children
income_deprivation_affecting_older_people	int2	SAS	Score value of deciles and quintiles ranging from least deprived (decile score 10) to most deprived (decile score 1) on the index of income deprivation affecting older people
akr_id	text	SAS	Unique ID

Table 6: NLP Diagnosis

Column Name	SQL Type	Output Data Type	Description / Comments
extraction_id	bigint	NLP	Unique identifier of an individual NLP-derived extraction
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
icd10_code	text	NLP	ICD-10 code for diagnosis extraction
icd10_label	text	NLP	ICD-10 label description for diagnosis extraction
icd10_extraction_clean	text	NLP	Original text string extracted from a source document by an NLP model
diagnosis_status	text	NLP	NLP-derived classification of diagnosis extraction given semantic and syntactic context of source sentence: has (individual currently has the diagnosis), had (individual had the diagnosis in the past, but no indication from the sentence that individual currently has the diagnosis), does not have (individual does not currently have the diagnosis), could have (sentence neither affirms nor negates the diagnosis)
diagnosis_experiencer	text	NLP	NLP-derived classification of diagnosis extraction given semantic and syntactic context of source sentence: patient (diagnosis refers to the patient), other (diagnosis refers to an individual other than the patient)
akr_id	text	NLP	Unique ID

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

Table 7: NLP Medication

Column Name	SQL Type	Output Data Type	Description / Comments
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
drug_name	text	NLP	Name of drug e.g., gabapentin, aripriprazole, mirtazapine
drug_concept_id	text	NLP	Dictionary of Medicines and Devices (Dm+d) ID for medication extraction
brand_name	text	NLP	Medication brand name extraction, if available (medications are typically referred to by generic names in NHS services)
rxnorm_drug_id	text	NLP	RxNorm ID for medication brand extraction
rxnorm_extension_drug_id	text	NLP	RxNorm Extension ID for generic and brand medications which exist outside the US
dosage_amount	text	NLP	Dosage amount, if available (medications are often referred to without providing dosages)
dosage_unit	text	NLP	Dosage unit, if available
va_class	text	NLP	Vetrans Affairs (VA) psyhciatric-specific class name from the VA drug classification system
va_class_code	text	NLP	VA psyhciatric-specific class code from the VA drug classification system
va_subclass	text	NLP	VA psyhciatric-specific subclass name from the VA drug classification system
va_subclass_code	text	NLP	VA psyhciatric-specific subclass code from the VA drug classification system
akrivia_subclass	text	NLP	Akrivia-adjusted psychiatric-specific subclass name
medication_status	text	NLP	NLP-derived classification of medication extraction given semantic and syntactic context of source sentence: is on (patient is currently taking the medication), was on (patient was taking the medication in the past, but no indication from the sentence that patient is currently taking the medication), other (no indication that the patient is or has previously been taking the medication)
count	bigint	NLP	Number of mentions for the medication within the same document (with the same classification status and dosage)
akr_id	text	NLP	Unique ID

For more information on variable subclasses, please contact contact@akriviahealth.com.

Table 8: NLP Health Score

Column Name	SQL Type	Output Data Type	Description / Comments
note_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
dateandtime	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
healthscore_name	text	NLP	Health score type, from the following list: ace, bdi, core10, gad, hamd, honos, madrs, mmse, moca, panss, phq, qids, qols, wsas
healthscore_value	bigint	NLP	Score value for the NLP-derived health score. Note these are summary scores only (clinical notes do not generally include subscale information).
akr_id	text	NLP	Unique ID

Table 9: NLP Health Measurement

Column Name	SQL Type	Output Data Type	Description / Comments
extraction_id	bigint	NLP	Unique identifier of an individual NLP-derived extraction
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
health_measurement	text	NLP	Over 80 harmonised health measurement name e.g., BMI, weight, blood glucose level (see below for full list)
harmonised_value	int	NLP	Harmonised health measurement value
harmonised_unit	text	NLP	Harmonised health measurement unit
extraction_name	text	NLP	Unharmonised health measurement name
extraction_value	text	NLP	Unharmonised health measurement value
extraction_unit	text	NLP	Unharmonised health measurement unit
extraction_description	text	NLP	Unharmonised descriptive text associated with the health measurement found within the same sentence context (e.g., "decrease", "stable")

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

For more information on specific health measurements, please contact contact@akriviahealth.com.

Table 10: NLP Psychotherapy

Column Name	SQL Type	Output Data Type	Description / Comments
extraction_id	bigint	NLP	Unique identifier of an individual NLP-derived extraction
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
group_name	text	NLP	Therapy subtype e.g., Psychotherapy, Relationship counselling, Family therapy etc
psychotherapy_status	text	NLP	NLP-derived classification of psychotherapy extraction given semantic and syntactic context of source sentence: is receiving (the sentence indicates that the patient is receiving psychotherapy at time of note creation), was receiving (the sentence indicates that the patient received psychotherapy in the past - note that this does not necessarily mean they have stopped receiving it), other (any other context, including sentences where the experiencer receiving a psychotherapy is someone other than the patient).
akr_id	text	NLP	Unique ID

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

Table 11: NLP ECT

Column Name	SQL Type	Output Data Type	Description / Comments
extraction_id	bigint	NLP	Unique identifier of an individual NLP-derived extraction
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
ect_status	text	NLP	NLP-derived classification of ECT extraction given semantic and syntactic context of source sentence: is receiving (the sentence indicates that the patient is receiving ECT at time of note creation), was receiving (the sentence indicates that the patient received ECT in the past - note that this does not necessarily mean they have stopped receiving it), other (any other context, including sentences where the experiencer receiving ECT is someone other than the patient)
akr_id	text	NLP	Unique ID

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

Table 12: NLP Substance Use

Column Name	SQL Type	Output Data Type	Description / Comments
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_id	bigint	NLP	Unique identifier of an individual NLP-derived extraction
substance_type	text	NLP	Substance type, from the following list: Alcohol, Cannabis, Cocaine, Nicotine, Opiates, Other
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
substance_use_status	text	NLP	NLP-derived classification of signs and symptoms extraction given semantic and syntactic context of source sentence: is using (the sentence indicates that the patient is currently using the substance), was using (the sentence indicates that the patient was using the substance - note that this does not necessarily mean they have stopped its usage), is not using (the sentence indicates that the patient is not using the substance), other (any other context)
akr_id	text	NLP	Unique ID

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

Table 13: NLP Signs & Symptoms

Column Name	SQL Type	Output Data Type	Description / Comments
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_id	bigint	NLP	Unique identifier of an individual NLP-derived extraction
concept_id	text	NLP	OMOP concept ID for symptom extraction
snomed_id	text	NLP	SNOMED ID for symptom extraction
snomed_label	text	NLP	SNOMED label description for symptom extraction
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
signs_and_symptoms_status	text	NLP	NLP-derived classification of 420 signs and symptoms extraction given semantic and syntactic context of source sentence: has (individual currently has the sign or symptom), had (individual had the sign or symptom in the past, but no indication from the sentence that individual currently has the sign or symptom), does not have (individual does not currently have the sign or symptom), other (sentence neither affirms nor negates the sign or symptom)
akr_id	text	NLP	Unique ID

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

For more information on our 400+ signs and symptoms, please see our 'Signs & Symptoms Annex' to this data dictionary or contact contact@akriviahealth.com.

Table 14: NLP Unemployment

Column Name	SQL Type	Output Data Type	Description / Comments
extraction_id	int8	NLP	Unique identifier of an individual NLP-derived extraction
akr_id	uuid	NLP	Unique ID
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
group_name	int4	NLP	Subtype within the concept: Job loss; Unemployed
version	int4	NLP	Version of the model ran

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

Table 15: NLP Homelessness

Column Name	SQL Type	Output Data Type	Description / Comments
extraction_id	int8	NLP	Unique identifier for the NLP extraction
akr_id	uuid	NLP	Unique ID
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
sentence_id	text	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
group_name	int4	NLP	Subtype within the concept: Homeless; Loss of home
version	int4	NLP	Version of NLP model ran

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.

Table 16: NLP Suicidality

Column Name	SQL Type	Output Data Type	Field Description
akr_id	uuid	NLP	Unique ID
document_id	text	NLP	Unique identifier for the document from which an NLP-derived mention was extracted
document_date	date	NLP	Creation date [YYYY/MM/DD] for the document from which an NLP-derived mention was extracted
extraction_id	int8	NLP	Unique identifier of an individual NLP-derived extraction
sentence_id	int8	NLP	Unique identifier for the sentence from which an NLP-derived mention was extracted
extraction_1	text	NLP	Original text string extracted from a source document by an NLP model
version	int4	NLP	version of NLP model ran
suicidality_status	text	NLP	NLP-derived classification of suicidality extraction given semantic and syntactic context of source sentence: has (individual currently has the sign or symptom), had (individual had the sign or symptom in the past, but no indication from the sentence that individual currently has the sign or symptom), does not have (individual does not currently have the sign or symptom), (sentence neither affirms nor negates the sign or symptom)

Note that extraction fields are 'semi-structured' in that they have been pre-filtered for information governance reasons.