CPRD Aurum Data Specification

Version 1.2

Date: 29 January 2018

Author:

Helen Booth and Daniel Dedman, CPRD, UK

Documentation Control Sheet

During the course of the project it may be necessary to issue amendments or clarifications to parts of this document. This form must be updated whenever changes are made and should be filed inside the front cover of the new or amended document.

Version	Summary of Change	Prepared By	Date	Reviewed By	Date
1.0		Helen Booth	06/12/2017		
1.1	Update fields	Helen Booth	11/01/2018		
1.2	Update fields	Helen Booth	29/01/2018		

Summary of Changes

- Primary keys in tables now indicated. Additional fields in the product dictionary
- Change to fields in the observation table to reflect decimal values in the database

Data Format

The data are available to researchers as eight files in text format as listed below. The data will be released as study specific datasets.

1. The **Patient** file (Patient*NNN*.txt) contains basic patient demographics and patient registration details for the patients.

2. The **Practice** file (Practice001.txt) contains details of each practice. Currently this is limited to the practice identifier and the last collection date, but derived variables are due to be added in Spring 2018.

3. The **Staff** file (Staff*NNN*.txt) contains practice staff details.

4. The **Consultation** file (Consultation*NNN*.txt) contains information relating to the type of consultation as entered by the GP. Consultations can be linked to observations that occur during the consultation via the consultation identifier (consid).

5. The **Observation** file (Observation*NNN*.txt) contains medical history events. This file contains all the medical history data entered on the GP system including symptoms, measurements and diagnoses. This can also be used to identify demographic information and deaths that has been recorded as a clinical code e.g. patient ethnicity.

6. The **Referral** file (Referral*NNN*.txt) contains referral details recorded on the GP system. These files contain information involving both inbound and outbound patient referrals to or from external care centres (normally to secondary care locations such as hospitals for inpatient or outpatient care). Referrals should be linked to the Observation file using the observation identifier (obsid) to identify the reason for the referral and the date.

7. The **Problem** file (ProblemNNN.txt) contains details of the patient's medical history that have been defined by the GP as a 'problem'. Information on identifying associated problems, the significance of the problem and its expected duration can be found in this table. Problems can be linked to an observation record (which would contain the clinical code for the problem) using the observation identifier (obsid). GPs may use 'problems' to manage chronic conditions as it would allow them to group clinical events (including drug prescriptions, measurements, symptom recording) by problem rather than chronologically.

8. The **Drug issue** file (DrugIssue*NNN*.txt) contains details of all prescriptions on the GP system. This file contains data relating to all prescriptions (for drugs and appliances) issued by the GP. Prescriptions may be linked to problems via the Observation file, using the observation identifier (obsid).

Data dictionaries

The CPRD Aurum dictionaries will be provided as text files that can be imported into standard statistical software to enable code searching. The dictionaries will also be available through the CPRD code browser. The CPRD code browser and a user guide can be requested by contacting enquiries@cprd.com. If you are already using the code browser to search the CPRD GOLD dictionaries you will still need to contact us to download the browser containing the CPRD Aurum dictionaries. For advice on producing code lists please refer to the CPRD Aurum FAQ documentation.

1. The **Medical** Dictionary contains information on all medical history observations that have been recorded in EMIS Web. Observations are coded using a combination of SNOMED, Read and local EMIS codes. Further information is provided in later sections of this document.

2. The **Drug** Dictionary contains information on drug and appliance prescriptions recorded in EMIS Web. This information is coded using the Dictionary of Medicines and Devices (DM+D). Further information is provided in later sections of this document.

Field descriptions

Full descriptions of fields in each data file are provided in the tables below. Most files can be linked using the encrypted patient identifier (patid). The last five digits of the patient identifier (patid) denote the identifier of the practice (pracid) that the patient belongs to. The mapping column lists lookup files with further information on decoding numerical values.

1. Patient

Column name	Field name	Description	Mapping	Type	Format
Patient identifier	patid	Encrypted unique identifier given to a patient in CPRD Aurum. The patient identifier is unique to CPRD Aurum, and may represent a different patient in the CPRD GOLD database. This is the primary key for this table.		INTEGER	19
CPRD practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum		INTEGER	10
Usual GP	usualgpstaffid	The GP that the patient is nominally registered with. To be used with the Staff table for reference	Lookup: Staff table	INTEGER	5
Gender	gender	Patient's gender	Lookup: Gender.txt	INTEGER	1
Year of birth	yob	Patient's year of birth. This is actual year of birth e.g. 1984.		INTEGER	4
Month of birth	mob	Patient's month of birth (for those aged under 16).		INTEGER	2
Date of death	emis_ddate	Date of death as recorded in the EMIS software. Researchers are advised to treat the *emis_ddate* with great caution as CPRD researchers have observed inaccuracies in its recording. A CPRD death algorithm will be developed to provide a more realistic estimate of death date.		DATE	DD/MM/YYYY
Registration start date	regstartdate	The date that the patient registered with the CPRD contributing practice. Most recent date the patient is recorded as having registered at the practice. If a patient deregistered for a period of time and returned, the return date would be recorded.		DATE	DD/MM/YYYY
Patient type	patienttypeid	The category that the patient has been assigned to e.g. private, regular, temporary.	Lookup: PatientType.txt	INTEGER	1
Registration end date	regenddate	Date the patient's registration at the practice ended. This may represent a transfer-out date or death date.		DATE	1
Acceptable flag	acceptable	The acceptable patient flag will be added to the database in 2018. In the interim this field will be empty. Please see the CPRD Aurum FAQ		INTEGER	1

Column name	Field name	Description	Mapping	Type	Format
		document for advice on a temporary solution in the absence of an acceptable flag.			
CPRD death date	cprd_ddate	CPRD is developing an algorithm to assign a likely date of death in CPRD Aurum. Until this is added this field will be empty. There is an 'emis_ddate' but we advise users to treat this variable with great caution.		DATE	DD/MM/YYYY

2. Practice

Column name	Field name	Description	Mapping	Type	Format
CPRD practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum. This is the primary key for this table.		INTEGER	10
Last Collection Date	lcd	Date of the most recent CPRD data collection for the practice.		DATE	DD/MM/YYYY
Up-to-standard date	uts	This field will be empty while a CPRD algorithm is developed for the CPRD Aurum database. Once filled, it will represent the date at which the practice data is deemed to be of research quality.		DATE	DD/MM/YYYY
Region	region	Value to indicate where in the UK the practice is based. The region denotes the Strategic Health Authority for English practices.	Lookup: Region.txt [not included in initial release]	INTEGER	3

3. Staff

Column name	Field name	Description	Mapping	Type	Format
Staff identifier	staffid	Encrypted unique identifier given to the practice staff member in CPRD Aurum. This is the primary key for this table.		INTEGER	5
CPRD practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum		INTEGER	10
Job category	jobcategoryid	Job category of the staff member who created the event	Lookup JobCat.txt	INTEGER	2

4. Consultation

Column name	Field name	Description	Mapping	Type	Format
Patient identifier	patid	Encrypted unique identifier given to a patient in CPRD Aurum. The patient identifier is unique to CPRD Aurum, and may represent a different patient in the CPRD GOLD database.		INTEGER	19
Consultation identifier	consid	Unique identifier given to the consultation. This is the primary key for this table.		INTEGER	
CPRD Practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum		INTEGER	10
Event date	consdate	Date associated with the event		DATE	DD/MM/YYYY
Entered date	enterdate	Date the event was entered into the practice system		DATE	DD/MM/YYYY
Staff identifier	staffid	Encrypted unique identifier given to the practice staff member who took the consultation in CPRD Aurum		INTEGER	5
EMIS consultation source identifier	conssourceid	Source of the consultation as recorded in the EMIS software. Many of the fields in this column are generic e.g. GP surgery, telephone, but many contain identifiers such as practice name. Until a process for anonymising and cleaning this variable is established a lookup will not be provided.	Lookup: ConSource.txt [not included in initial release]	INTEGER	1
CPRD consultation source identifier	cprdconstype	Source of the consultation e.g. telephone, main surgery. This lookup will be a compressed version of the 'conssourceid' i.e. where practice names have been recorded they will be compressed to 'GP practice'.	Lookup: cprdconstype.txt [not included in initial release]	INTEGER	1
Consultation source code identifier	consmedcodeid	Source of the consultation from EMIS software. This is a medical code that can be used with the medical dictionary. It may contain information similar to the consultation source identifiers, but is available for use now. Some of the codes may not be interpretable e.g. Awaiting clinical code migration to EMIS Web.	Medical dictionary. Maps to medcodeid	INTEGER	19

5. Observation

Column name	Field name	Description	Mapping	Type	Format
Patient identifier	patid	Encrypted unique identifier given to a patient in CPRD Aurum. The patient identifier is unique to CPRD Aurum, and may represent a different patient in the CPRD GOLD database.		INTEGER	19
Consultation identifier	consid	Unique identifier given to the consultation. In EMIS Web it is not necessary to enter observations within a consultation, so this identifier may be missing.		INTEGER	
CPRD Practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum		INTEGER	10
Observation identifier	obsid	Unique identifier given to the observation. This is the primary key for this table.		INTEGER	
Event date	obsdate	Date associated with the event		DATE	DD/MM/YYYY
Entered date	enterdate	Date the event was entered into EMIS Web		DATE	DD/MM/YYYY
Staff identifier	staffid	Encrypted unique identifier given to the practice staff member who took the consultation in CPRD Aurum		INTEGER	
Parent observation identifier	parentobsid	Observation identifier (obsid) that is the parent to the observation. This enables grouping of multiple observations, such as systolic and diastolic blood pressure, or blood test results.	Observation table	INTEGER	
Medical code	medcodeid	CPRD unique code for the medical term selected by the GP	Lookup: Medical dictionary	INTEGER	19
Value	value	Measurement or test value		NUMERIC	
Numeric unit identifer	numunitid	Unit of measurement	Lookup: NumUnit.txt	INTEGER	4
Observation type identifier	obstypeid	Type of observation (allergy, family history, observation)	Lookup: ObsType.txt	INTEGER	2
Numeric range low	numrangelow	Value representing the low boundary of the normal range for this measurement		NUMERIC	
Numeric range high	numrangehigh	Value representing the high boundary of the normal range for this measurement		NUMERIC	
Problem observation identifier	probobsid	Observation identifier (obsid) of any problem that an observation is associated with. An example of this might be an overarching condition that the current observation is associated with such as 'wheezing' with		INTEGER	

Column name	Field name	Description	Mapping	Type	Format
		the problem observation identifier that links to an observation of 'asthma', that in turn contains information in the problem table.			

6. Referral

Column name	Field name	Description	Mapping	Type	Format
Patient identifier	patid	Encrypted unique identifier given to a patient in CPRD Aurum. The patient identifier is unique to CPRD Aurum, and may represent a different patient in the CPRD GOLD database.		INTEGER	19
Observation identifier	obsid	Unique identifier given to the observation		INTEGER	
CPRD Practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum		INTEGER	10
Referral source organisation identifier	refsourceorgid	Source organisation of the referral. Organisations are identified by an ID number and each organisation has a type (e.g hospital inpatient department, community clinic). Both the organisation table and the OrgType lookup are required.	Lookups: Organisation.txt *and* OrgType.txt [not included in initial release]	INTEGER	
Referral target organisation identifier	reftargetorgid	Source organisation of the referral. Organisations are identified by an ID number and each organisation has a type (e.g hospital inpatient department, community clinic). Both the organisation table and the OrgType lookup are required.	Lookups: Organisation.txt *and* OrgType.txt [not included in initial release]	INTEGER	
Referral urgency identifier	refurgencyid	Urgency of the referral e.g. routine, urgent, dated	Lookup: RefUrgency.txt	INTEGER	1
Referral service type identifier	refservicetypeid	Type of service the referral relates to e.g. assessment, management, investigation	Lookup: RefServiceType.txt	INTEGER	2
Referral mode identifier	refmodeid	Mode by which the referral was made e.g. telephone, written	Lookup: RefMode.txt	INTEGER	1

7. Problem

Column name	Field name	Description	Mapping	Type	Format
Patient identifier	patid	Encrypted unique identifier given to a patient in CPRD Aurum. The patient identifier is unique to CPRD Aurum, and may represent a different patient in the CPRD GOLD database.		INTEGER	19
Observation identifier	obsid	Unique identifier given to the observation.		INTEGER	
CPRD Practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum		INTEGER	10
Parent problem observation identifier	parentprobobsid	Observation identifier for the parent observation of the problem. This can be used to group problems via the Observation table.		INTEGER	
Problem end date	probenddate	Date the problem ended		DATE	DD/MM/YYYY
Expected duration	expduration	Expected duration of the problem in days		INTEGER	
Last review date	lastrevdate	Date the problem was last reviewed		DATE	DD/MM/YYYY
Last review staff identifier	lastrevstaffid	Staff member who last reviewed the problem		INTEGER	10
Parent problem relationship identifier	parentprobrelid	Description of the relationship of the problem to another problem e.g. the problem may have evolved or been merged with another problem as the problem, or the GP's understanding of the problem, changes	Lookup ParentProblemRelationship.txt	INTEGER	1
Problem status identifier	probstatusid	Status of the problem (active, past)	Lookup: ProbStatus.txt	INTEGER	1
Significance	signid	Significance of the problem (minor, significant)	Lookup: Sign.txt	INTEGER	1

8. Drug issue

Column name	Field name	Description	Mapping	Type	Format
Patient identifier	patid	Encrypted unique identifier given to a patient in CPRD Aurum. The patient identifier is unique to CPRD Aurum, and may represent a different patient in the CPRD GOLD database.		INTEGER	19
Issue record identifier	issueid	Unique identifier given to the issue record. This is the primary key for this table.		INTEGER	
CPRD Practice identifier	pracid	Encrypted unique identifier given to a practice in CPRD Aurum		INTEGER	10
Problem observation identifier	obsid	Unique identifier for the observation that links to a problem under which this prescription was issued. This refers to an 'obsid' in the Observation table which, in turn, can be linked to a record in the Problem table using 'obsid'.		INTEGER	
Drug record identifier	drugrecid	Unique identifier to a drug template record, which is not currently for release. At present this may be used to group repeat prescriptions from the same template.		INTEGER	
Event date	issuedate	Date associated with the event		DATE	DD/MM/YYYY
Entered date	enterdate	Date the event was entered into EMIS		DATE	DD/MM/YYYY
Staff identifier	staffid	Encrypted unique identifier given to the practice staff member issued the treatment in CPRD Aurum		INTEGER	5
Drug code identifier	prodcodeid	Unique CPRD code for the treatment selected by the GP	Lookup: Drug dictionary	INTEGER	
Dosage identifier	dosageid	Identifier that allows dosage information on the event to be retrieved. Not included in first release	Lookup: tbc	INTEGER	10
Quantity	quantity	Total quantity entered by the GP for the prescribed treatment		INTEGER	9
Quantity unit identifier	quantunitid	Unit of the treatment (capsule, tablet)	Lookup: QuantUnit.txt	INTEGER	2
Course duration in days	duration	Duration of the treatment in days		INTEGER	4
Estimated NHS cost	estnhscost	Estimated cost of the treatment to the NHS		INTEGER	9

1. Medical dictionary

Column name	Description	Mapping	Type	Format
medcodeid	CPRD code to describe the observation. Links to the observation table		INTEGER	
term	Description of the observation associated with the codeid		STRING	
originalreadcode	The original Read code text as provided in the EMIS dictionary (contains codes which are not valid Read codes)		STRING	
cleansedreadcode	CPRD-cleaned and validated version of the originalreadcode		STRING	
snomedctconceptid	The SNOMED CT concept identifier associated with the observation		INTEGER	
snomedctdescid	The SNOMED CT description identifier associated with the observation		INTEGER	
emiscodecatid	Category identifier in EMIS that describes the observation	Lookup: EMIScodecat.txt	INTEGER	
release	Currently empty			

2. Drug code dictionary

Column name	Description	Mapping	Type	Format
prodcodeid	CPRD code to describe the treatment. Maps to the Issue Record file		INTEGER	
termfromemis	Description of the treatment provided by EMIS associated with the prodcodeid		STRING	
dmdcode	The DM+D code associated with the treatment		INTEGER	
productname	Name of the product		STRING	
formulation	Formulation of the product		STRING	
routeofadministration	Route of administration for the product		STRING	
drugsubstancename	Active ingredient(s) included in the product. For combination therapies, each component is listed, separated by /		STRING	
substancestrength	Strength of each active ingredient listed in the drugsubstancename column, including units. Separated by / if more than 1		STRING	
bnfchapter	BNF chapter to which the product belongs		STRING	
release	Currently empty			