



CPRD *Gold* Data Specification

Version 1.3

Author: Shivani Padmanabhan, 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	Affected Areas Summary of Change	Prepared By	Date	Reviewed By	Date
1.0	Initial Draft				
1.1	Modified	Shivani Padmanabhan	01/06/09	Nick Wilson	22/07/09
1.2	Modified	Shivani Padmanabhan	28/07/09	Arlene Gallagher	30/07/09
1.3	Modified	Shivani Padmanabhan	06/01/11	Nick Wilson	07/01/11

SUMMARY OF CHANGES

Version 1.1

- Refined wordings

Version 1.2

- Acceptable field in Patient file equals 1 if patient is acceptable, else 0 (Lookup reference incorrectly labelled as Y_N in previous versions)
- UTS field in Practice file has been derived using a CPRD algorithm that looks at death recording at the practice, and gaps in the data (prior to August 2009, this field was populated with the practice UTS date as was generated in the old FF-CPRD system)
- Ndd field in the Therapy file has been populated for the most common occurring dosage strings in the data (field was set to '0' prior to August 2009)
- Descriptions of all fields have been revised, for clarity

Version 1.3

- Field name 'attendtype' in Referral table modified to 'attendance'

Table of contents

<i>Dataset Format</i>	4
<i>Field descriptions</i>	5

Dataset Format

1. The **Patient** file (PatientNNN.txt) contains basic patient demographics and patient registration details for the patients.
2. The **Practice** file (Practice001.txt) contains details of each practice, including region and collection information.
3. The **Staff** file (StaffNNN.txt) contains practice staff details, with one record per member of staff.
4. The **Consultation** file (ConsultationNNN.txt) contains information relating to the type of consultation as entered by the GP from a pre-determined list. Consultations can be linked to the events that occur as part of the consultation via the consultation identifier (consid).
5. The **Clinical** file (ClinicalNNN.txt) contains medical history events. This file contains all the medical history data entered on the GP system, including symptoms, signs and diagnoses. This can be used to identify any clinical diagnoses, and deaths. Patients may have more than one row of data. The data is coded using Read codes, which allow linkage of codes to the medical terms provided.
6. The **Additional Clinical Details** file (AdditionalNNN.txt) contains information entered in the structured data areas in the GP's software. Patients may have more than one row of data. Data in this file is linked to events in the clinical file through the additional details identifier (adid).
7. The **Referral** file (ReferralNNN.txt) contains referral details recorded on the GP system. These files contain information involving patient referrals to external care centres (normally to secondary care locations such as hospitals for inpatient or outpatient care), and include speciality and referral type.
8. The **Immunisation** file (ImmunisationNNN.txt) contains details of immunisation records on the GP system.
9. The **Test** file (TestNNN.txt) contains records of test data on the GP system. The data is coded using a Read code, chosen by the GP, which will generally identify the type of test used. The test name is identified via the *Entity Type*, a numerical code, which is determined by the test result item chosen by the GP at source. There are three types of test records, involving 4, 7 or 8 data fields (data1 - data8). The data must be managed according to which sort of test record it is. Data can denote either qualitative text based results (for example 'Normal' or Abnormal') or quantitative results involving a numeric value.
10. The **Therapy** file (TherapyNNN.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. Patients may have more than one row of data. Drug products and appliances are recorded by the GP using the Multilex product code system.

Field descriptions

Full descriptions of fields in each file are provided in the tables below. All files can be linked using the encrypted patient identifier (patid). The last three digits of the patient identifier (patid), and staff identifier (staffid) denote the identifier of the practice (pracid) that the patient, or staff belongs to. The mapping column references information relating to the use of data in the field. It specifies lookup references, linkages to other tables, and information on decoding numerical values. A mapping of 'None' indicates the existence of raw data in the field.

1. Patient

Column name	Field name	Description	Mapping
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
VAMP Identifier	vmid	Old VM id for the patient when the practice was using the VAMP system	None
Patient Gender	gender	Patient's gender	Lookup SEX
Birth Year	yob	Patient's year of birth	Value + 1800
Birth Month	mob	Patient's month of birth (for those aged under 16). 0 indicates no month set	None
Marital Status	marital	Patient's current marital status	Lookup MAR
Family Number	famnum	Family ID number	None
CHS Registered	chsreg	Value to indicate whether the patient is registered with Child Health Surveillance	Lookup Y_N
CHS Registration Date	chsdate	Date of registration with Child Health Surveillance	dd/mm/yyyy ¹
Prescription Exemption	prescr	Type of prescribing exemption the patient has currently (e.g. medical or maternity)	Lookup PEX
Capitation Supplement	capsup	Level of capitation supplement the patient has currently (e.g. low, medium, or high)	Lookup CAP
Socio-Economic Status	ses	Patient's socio-economic status. Currently 0; to be populated in future builds	PAT_SES ²
First Registration Date	frd	Date the patient first registered with the practice. If patient only has 'temporary' records, the date is the first encounter with the practice; if patient has 'permanent' records it is the date of the first 'permanent' record (excluding preceding temporary records)	dd/mm/yyyy
Current Registration Date	crd	Date the patient's current period of registration with the practice began (date of the first 'permanent' record after the latest transferred out period). If there are no 'transferred out periods', the date is equal to 'frd'	dd/mm/yyyy
Registration Status	regstat	Status of registration detailing gaps and temporary patients	PAT_STAT ³
Registration Gaps	reggap	Number of days missing in the patients registration details	PAT_GAP ⁴
Internal Transfer	internal	Number of internal transfer out periods, in the patient's registration details	None
Transfer Out Date	tod	Date the patient transferred out of the practice, if relevant. Empty for patients who have not transferred out	dd/mm/yyyy
Transfer Out Reason	toreason	Reason the patient transferred out of the practice. Includes 'Death' as an option	Lookup TRA
Death Date	deathdate	Date of death of patient – derived using a CPRD algorithm	dd/mm/yyyy
Acceptable Patient Flag	accept	Flag to indicate whether the patient has met certain quality standards: 1 = acceptable, 0 = unacceptable	Boolean

2. Practice

<i>Column name</i>	<i>Field name</i>	<i>Description</i>	<i>Mapping</i>
Practice identifier	pracid	Encrypted unique identifier given to a specific practice in CPRD GOLD	None
Region	region	Value to indicate where in the UK the practice is based. The region denotes the Strategic Health Authority for practices within England, and the country i.e. Wales, Scotland, or Northern Ireland for the rest	Lookup PRG
Last Collection Date	lcd	Date of the last collection for the practice	dd/mm/yyyy
Up To Standard Date	uts	Date at which the practice data is deemed to be of research quality. Derived using a CPRD algorithm that primarily looks at practice death recording and gaps in the data	dd/mm/yyyy

3. Staff

<i>Column name</i>	<i>Field name</i>	<i>Description</i>	<i>Mapping</i>
Staff Identifier	staffid	Encrypted unique identifier given to the practice staff member entering the data	None
Staff Gender	gender	Staff's gender	Lookup SEX
Staff Role	role	Role of the member of staff who created the event	Lookup ROL

4. Consultation

<i>Column name</i>	<i>Field name</i>	<i>Description</i>	<i>Mapping</i>
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
Event Date	eventdate	Date associated with the event, as entered by the GP	dd/mm/yyyy
System Date	sysdate	Date the event was entered into Vision	dd/mm/yyyy
Consultation Type	constype	Type of consultation (e.g. Surgery Consultation, Night Visit, Emergency etc)	Lookup COT
Consultation Identifier	consid	The consultation identifier linking events at the same consultation, when used in combination with pracid	Link Event tables
Staff Identifier	staffid	The identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown	Link Staff table
Duration	duration	The length of time (minutes) between the opening, and closing of the consultation record	None

5. Clinical

<i>Column name</i>	<i>Field name</i>	<i>Description</i>	<i>Mapping</i>
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
Event Date	eventdate	Date associated with the event, as entered by the GP	dd/mm/yyyy
System Date	sysdate	Date the event was entered into Vision	dd/mm/yyyy
Consultation Type	constype	Code for the category of event recorded within the GP system (e.g. diagnosis or symptom)	Lookup SED
Consultation Identifier	consid	Identifier that allows information about the consultation to be retrieved, when used in combination with pracid	Link Consultation table
Medical Code	medcode	CPRD unique code for the medical term selected by the GP	Lookup Medical Dictionary
Staff Identifier	staffid	Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown	Link Staff table
Text Identifier	textid	Identifier that allows freetext information on the event to be retrieved, when used in combination with pracid and event type 'Clinical'. A value of 0 indicates that there is no freetext information for this event	Link Freetext
Episode	episode	Episode type for a specific clinical event	Lookup EPI
Entity Type	enttype	Identifier that represents the structured data area in Vision where the data was entered	Lookup Entity
Additional Details Identifier	adid	Identifier that allows additional information to be retrieved for this event, when used in combination with pracid. A value of 0 signifies that there is no additional information associated with the event.	Link Additional Clinical Details table

6. Additional Clinical Details

<i>Column name</i>	<i>Field name</i>	<i>Description</i>	<i>Mapping</i>
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
Entity Type	enttype	Identifier that represents the structured data area in Vision where the data was entered	Lookup Entity
Additional Details Identifier	adid	Identifier that allows information about the original clinical event to be retrieved, when used in combination with pracid	Link Clinical table
Data 1	data1	Depends on Entity Type ♦	Lookup Entity
Data 2	data2	Depends on Entity Type	Lookup Entity
Data 3	data3	Depends on Entity Type	Lookup Entity
Data 4	data4	Depends on Entity Type	Lookup Entity
Data 5	data5	Depends on Entity Type	Lookup Entity
Data 6	data6	Depends on Entity Type	Lookup Entity
Data 7	data7	Depends on Entity Type	Lookup Entity

♦ Each entity type may be associated with up to seven data fields. Content of each data field is dependent on the entity type – the fields may contain raw data values, or may be encoded values that represent dates, read codes, text etc. The file Entity.xls contains information on all entity types, and provides the number of data fields associated with the entity, description of the data in each field, and details of the lookups needed to decode the data.

7. Referral

<i>Column name</i>	<i>Field name</i>	<i>Description</i>	<i>Mapping</i>
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
Event Date	eventdate	Date associated with the event, as entered by the GP	dd/mm/yyyy
System Date	sysdate	Date the event was entered into Vision	dd/mm/yyyy
Consultation Type	constype	Code for the category of event recorded within the GP system (e.g. management or administration)	Lookup SED
Consultation Identifier	consid	Identifier that allows information about the consultation to be retrieved, when used in combination with pracid	Link Consultation table
Medical Code	medcode	CPRD unique code for the medical term selected by the GP	Lookup Medical Dictionary
Staff Identifier	staffid	Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown	Link Staff table
Text Identifier	textid	Identifier that allows freetext information on the event to be retrieved, when used in combination with pracid and event type 'Referral'. A value of 0 indicates that there is no freetext information for this event	Link Freetext
Source	source	Classification of the source of the referral e.g. GP, Self	Lookup SOU
NHS Speciality	nhsspec	Referral speciality according to the National Health Service (NHS) classification	Lookup DEP
FHSA Speciality	fhsaspec	Referral speciality according to the Family Health Services Authority (FHSA) classification	Lookup SPE
In Patient	inpatient	Classification of the type of referral, e.g. Day case, In patient	Lookup RFT
Attendance Type	attendance	Category describing whether the referral event is the first visit, a follow-up etc	Lookup ATT
Urgency	urgency	Classification of the urgency of the referral e.g. Routine, Urgent	Lookup URG

8. Immunisation

<i>Column name</i>	<i>Field name</i>	<i>Description</i>	<i>Mapping</i>
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
Event Date	eventdate	Date associated with the event, as entered by the GP	dd/mm/yyyy
System Date	sysdate	Date the event was entered into Vision	dd/mm/yyyy
Consultation Type	constype	Code for the category of event recorded within the GP system (e.g. intervention)	Lookup SED
Consultation Identifier	consid	Identifier that allows information about the consultation to be retrieved, when used in combination with pracid	Link Consultation table
Medical Code	medcode	CPRD unique code for the medical term selected by the GP	Lookup Medical Dictionary
Staff Identifier	staffid	Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown	Link Staff table
Text Identifier	textid	Identifier that allows freetext information on the event to be retrieved, when used in combination with pracid and event type 'Immunisation'. A value of 0 indicates that there is no freetext information for this event	Link Freetext
Type	immstype	Individual components of an immunisation, e.g. Mumps, Rubella, Measles	Lookup IMT
Stage	stage	Stage of the immunisation given, e.g. 1, 2, B2	Lookup IST
Status	status	Status of the immunisation e.g. Advised, Given, Refusal	Lookup IMM
Compound	compound	Immunisation compound administered – may be a single or multi-component preparation, e.g. MMR	Lookup IMC
Source	source	Location where the immunisation was administered, e.g. In this practice	Lookup INP
Reason	reason	Reason for administering the immunisation, e.g. Routine measure	Lookup RIN
Method	method	Route of administration for the immunisation, e.g. Oral, Intramuscular	Lookup IME

9. Test

Column name	Field name	Description	Mapping
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
Event Date	eventdate	Date associated with the event, as entered by the GP	dd/mm/yyyy
System Date	sysdate	Date the event was entered into Vision	dd/mm/yyyy
Consultation Type	constype	Code for the category of event recorded within the GP system (e.g. examination)	Lookup SED
Consultation Identifier	consid	Identifier that allows information about the consultation to be retrieved, when used in combination with pracid	Link Consultation table
Medical Code	medcode	CPRD unique code for the medical term selected by the GP	Lookup Medical Dictionary
Staff Identifier	staffid	Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown	Link Staff table
Text Identifier	textid	Identifier that allows freetext information on the event to be retrieved, when used in combination with pracid and event type 'Test'. A value of 0 indicates that there is no freetext information for this event	Link Freetext
Entity Type	enttype	Identifier that represents the structured data area in Vision where the data was entered	Lookup Entity
Data 1	data1	Qualifier	Lookup TQU
Data 2	data2	Normal range from	None
Data 3	data3	Normal range to	None
Data 4	data4	Normal range basis	None

Depending on the Test Entity Type, tests have either 4, 7, or 8 data fields

Data 1	data1	Operator	Lookup OPR
Data 2	data2	Value	None
Data 3	data3	Unit of measure	Lookup SUM
Data 4	data4	Qualifier	Lookup TQU
Data 5	data5	Normal range from	None
Data 6	data6	Normal range to	None
Data 7	data7	Normal range basis (or peak flow device for entity type 311)	Lookup POP (or PFD)

For some test entity types (data 1 to data 6 same as above):

Data 7	data7	Normal range basis	Lookup POP
Data 8	data8	Expected delivery date	GEN_SDC ⁵

10. Therapy

Column name	Field name	Description	Mapping
Patient Identifier	patid	Encrypted unique identifier given to a patient in CPRD GOLD	None
Event Date	eventdate	Date associated with the event, as entered by the GP	dd/mm/yyyy
System Date	sysdate	Date the event was entered into Vision	dd/mm/yyyy
Consultation Identifier	consid	Identifier that allows information about the consultation to be retrieved, when used in combination with pracid	Link Consultation table
Product Code	prodcode	CPRD unique code for the treatment selected by the GP	Lookup Product Dictionary
Staff Identifier	staffid	Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown	Link Staff table
Text Identifier	textid	Identifier that allows freetext information (dosage) on the event to be retrieved, when used in combination with pracid and event type 'Therapy'. A value of 0 indicates that there is no freetext information for the event. Use the Common Dosages Lookup (constituting ~ 95% of dosage strings in data) to interpret values < 100,000	Lookup Common Dosages Link Freetext
BNF Code	bnfcode	Code representing the chapter & section from the British National Formulary for the product selected by GP	Lookup BNFCodes
Total Quantity	qty	Total quantity entered by the GP for the prescribed product	None
Numeric Daily Dose	ndd	Numeric daily dose prescribed for the event. Derived using a CPRD algorithm on common dosage strings (represented by textid < 100,000). Value is set to 0 for all dosage strings represented by textid > 100,000	None
Number of Days	numdays	Number of treatment days prescribed for a specific therapy event	None
Number of Packs	numpacks	Number of individual product packs prescribed for a specific therapy event	None
Pack Type	packtype	Pack size or type of the prescribed product	Lookup PackType
Issue Sequence Number	issueseq	Number to indicate whether the event is associated with a repeat schedule. Value of 0 implies the event is not part of a repeat prescription. A value ≥ 1 denotes the issue number for the prescription within a repeat schedule	None

¹ **dd/mm/yyyy:** Valid dates are in the format DD/MM/YYYY. Missing dates are NULL, and invalid dates are set to 01/01/2500

² **PAT_SES:** The Index of Multiple Deprivation (IMD) socio-economic status (overall measure as a quintile) will be implemented in the 'ses' field of the Patient file for all patients belonging to English practices that have consented to linkage. The SES quintile will be calculated on the basis of lower level super output area (LSOA). Townsend scores (also based on LSOA) will be made available for the same patients (as a lookup file), on request.

³ **PAT_STAT:** Transferred out period is the time between a patient transferring out and re-registering at the same practice. If the patient has transferred out for a period of more than 1 day, and the transfer is not internal, this value is incremented. 0 means continuous registration, 1 means one 'transferred out period', 2 means two periods, etc. If the patient only has 'temporary' records then this value is set to 99.

⁴ **PAT_GAP:** Number of days between patient's transferred out date and re-registration date for the patient's 'transferred out periods', regardless of whether the transfer was internal or not.

⁵ **GEN_SDC:** The date in dd/mm/yyyy format can be obtained as follows:

0 = An invalid/ missing date

2 = A date greater than 31/12/2014

3 = A date earlier than 01/01/1800

All other values = number of days between the date and the 31/12/2014 offset by 10.

Example: A value of 4027 decodes to the date 01/01/2004. $4027 - 10 = 4017$ days prior to the date 31/12/2014 is the date 01/01/2004