# CPRD GOLD Data Specification

**Version 2.0**

**Date: 1st September 2017**

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 | Summary of Change | Prepared By | Date | Reviewed By | Date |
|---------|-------------------|-------------|------|-------------|------|
| 1.0 | Initial Draft | | | | |
| 1.1 | Modified | Shivani Padmanabhan | 01/06/2009 | Nick Wilson | 22/07/2009 |
| 1.2 | Modified | Shivani Padmanabhan | 28/07/2009 | Arlene Gallagher | 30/07/2009 |
| 1.3 | Modified | Shivani Padmanabhan | 06/01/2011 | Nick Wilson | 07/01/2011 |
| 1.4 | Modified | Shivani Padmanabhan | 11/01/2013 | Nick Wilson | 11/01/2013 |
| 1.5 | Modified | Shivani Padmanabhan | 31/07/2013 | Nick Wilson | 03/08/2013 |
| 1.6 | Formatted | Arlene Gallagher | 11/12/2013 | Shivani Padmanabhan | 30/12/2013 |
| 1.7 | Modified | Jennifer Chapman | 19/11/2014 | Darren Lunn | 20/11/2014 |
| 1.8 | Modified | Mark Hobbs | 06/05/2015 | Jennifer Chapman | 12/06/2015 |
| 1.9 | Formatted | Grant Lee | 02/07/2015 | Jennifer Chapman | 03/07/2015 |
| 2.0 | Modified | Shivani Padmanabhan | 01/09/2017 | Darren Lunn | 01/09/2017 |

**Summary of Changes**

Version 1.1
  o   Refined wordings

Version 1.2
  o   Acceptable field in Patient file equals 1 if patient is acceptable, else 0 (Lookup reference incorrectly labelled as Y_N in previous versions)
  o   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)
  o   The 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)
  o   Descriptions of all fields have been revised for clarity

Version 1.3
  o   Field name 'attendtype' in Referral table modified to 'attendance'

Version 1.4
  o   Reference to Multilex product code system has been changed to Gemscript

Version 1.5
  o   Added batch number as a field in immunisation
  o   Description and Mapping for 'data8' in Test file has been amended

Version 1.6
- o   New CPRD branding and formatting

Version 1.7
- o   Mapping of 'data8' field in Test file has been amended to remove reference to GEN_SDC
- o   Description of date formatting has been updated to reflect new format
- o   Description of entity types has been updated to include date values

Version 1.8
- o   Merged information from Spec 1.5 and Spec 1.7
- o   All description tables now list Column Name, Field Name, Description, Mapping, Type, Format

Version 1.9
- o   Minor formatting of document

Version 2.0
- o   Removed textid fields from the clinical, immunisation, test and referral tables
- o   Removed textid and ndd fields, and added dosageid as a field in the therapy table
- o   Removed ses field from the patient table

*Dataset Format*

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, including region and collection information.

3. The **Staff** file (Staff*NNN*.txt) contains practice staff details, with one record per member of staff.

4. The **Consultation** file (Consultation*NNN*.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 (Clinical*NNN*.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 (Additional*NNN*.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 (Referral*NNN*.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 (Immunisation*NNN*.txt) contains details of immunisation records on the GP system.

9. The **Test** file (Test*NNN*.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 (Therapy*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. Patients may have more than one row of data. Drug products and appliances are recorded by the GP using the Gemscript 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 | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| VAMP Identifier | vmid | Old VM id for the patient when the practice was using the VAMP system | None | INTEGER | 20 |
| Patient Gender | gender | Patient's gender | Lookup SEX | INTEGER | 1 |
| Birth Year | yob | Patient's year of birth | Value + 1800 | INTEGER | 4 |
| Birth Month | mob | Patient's month of birth (for those aged under 16). 0 indicates no month set | None | INTEGER | 2 |
| Marital Status | marital | Patient's current marital status | Lookup MAR | INTEGER | 3 |
| Family Number | famnum | Family ID number | None | INTEGER | 20 |
| CHS Registered | chsreg | Value to indicate whether the patient is registered with Child Health Surveillance | Lookup Y_N | INTEGER | 1 |
| CHS Registration Date | chsdate | Date of registration with Child Health Surveillance | dd/mm/yyyy [1] | DATE | dd/mm/yyyy |
| Prescription Exemption | prescr | Type of prescribing exemption the patient has currently (e.g. medical / maternity) | Lookup PEX | INTEGER | 3 |
| Capitation Supplement | capsup | Level of capitation supplement the patient has currently (e.g. low, medium, high) | Lookup CAP | INTEGER | 3 |
| 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 | dd/mm/yyyy [1] | DATE | dd/mm/yyyy |

---

[1] **dd/mm/yyyy**: Valid dates are in this format. Missing dates are NULL

| | | | | | |
|---|---|---|---|---|---|
| | | 'permanent' records it is the date of the first 'permanent' record (excluding preceding temporary records) | | | |
| 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 [1] | DATE | dd/mm/yyyy |
| Registration Status | regstat | Status of registration detailing gaps and temporary patients | PAT_STAT [2] | INTEGER | 2 |
| Registration Gaps | reggap | Number of days missing in the patients registration details | PAT_GAP [3] | INTEGER | 5 |
| Internal Transfer | internal | Number of internal transfer out periods, in the patient's registration details | None | INTEGER | 2 |
| 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 [1] | DATE | dd/mm/yyyy |
| Transfer Out Reason | toreason | Reason the patient transferred out of the practice. Includes 'Death' as an option | Lookup TRA | INTEGER | 3 |
| Death Date | deathdate | Date of death of patient – derived using a CPRD algorithm | dd/mm/yyyy [1] | DATE | dd/mm/yyyy |
| Acceptable Patient Flag | accept | Flag to indicate whether the patient has met certain quality standards: 1 = acceptable, 0 = unacceptable | Boolean | INTEGER | 1 |

---

[2] **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.

[3] **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.

## 2. Practice

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Practice identifier | pracid | Encrypted unique identifier given to a specific practice in CPRD GOLD | None | INTEGER | 3 |
| 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 | INTEGER | 3 |
| Last Collection Date | lcd | Date of the last collection for the practice | dd/mm/yyyy **1** | DATE | 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 **1** | DATE | dd/mm/yyyy |

## 3. Staff

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Staff Identifier | staffid | Encrypted unique identifier given to the practice staff member entering the data | None | INTEGER | 20 |
| Staff Gender | gender | Staff's gender | Lookup SEX | INTEGER | 1 |
| Staff Role | role | Role of the member of staff who created the event | Lookup ROL | INTEGER | 3 |

## 4. Consultation

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| Event Date | eventdate | Date associated with the event, as entered by the GP | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| System Date | sysdate | Date the event was entered into Vision | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| Consultation Type | constype | Type of consultation (e.g. Surgery Consultation, Night Visit, Emergency etc.) | Lookup COT | INTEGER | 3 |
| Consultation Identifier | consid | The consultation identifier linking events at the same consultation, when used in combination with pracid | Link **Event** tables | INTEGER | 20 |
| 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 | INTEGER | 20 |
| Duration | duration | The length of time (minutes) between the opening, and closing of the consultation record | None | INTEGER | 10 |

## 5. Clinical

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| Event Date | eventdate | Date associated with the event, as entered by the GP | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| System Date | sysdate | Date the event was entered into Vision | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| Consultation Type | constype | Code for the category of event recorded within the GP system (e.g. diagnosis or symptom) | Lookup SED | INTEGER | 3 |
| Consultation Identifier | consid | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid | Link **Consultation** table | INTEGER | 20 |
| Medical Code | medcode | CPRD unique code for the medical term selected by the GP | Lookup Medical Dictionary | INTEGER | 20 |
| 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 | INTEGER | 20 |
| Episode | episode | Episode type for a specific clinical event | Lookup EPI | INTEGER | 3 |
| Entity Type | enttype | Identifier that represents the structured data area in Vision where the data was entered | Lookup Entity | INTEGER | 5 |
| 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 | INTEGER | 20 |

## 6. Additional Clinical Details

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| Entity Type | enttype | Identifier that represents the structured data area in Vision where the data was entered | Lookup Entity | INTEGER | 5 |
| 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 | INTEGER | 20 |
| Data 1 | data1 | Depends on Entity Type ♦ | Lookup Entity | NUMERIC DATE | 15.3 dd/mm/yyyy |
| Data 2 | data2 | Depends on Entity Type ♦ | Lookup Entity | NUMERIC DATE | 15.3 dd/mm/yyyy |
| Data 3 | data3 | Depends on Entity Type ♦ | Lookup Entity | NUMERIC DATE | 15.3 dd/mm/yyyy |
| Data 4 | data4 | Depends on Entity Type ♦ | Lookup Entity | INTEGER DATE | 12 dd/mm/yyyy |
| Data 5 | data5 | Depends on Entity Type ♦ | Lookup Entity | INTEGER DATE | 12 dd/mm/yyyy |
| Data 6 | data6 | Depends on Entity Type ♦ | Lookup Entity | INTEGER DATE | 12 dd/mm/yyyy |
| Data 7 | data7 | Depends on Entity Type ♦ | Lookup Entity | INTEGER DATE | 4 dd/mm/yyyy |

---

♦ 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, dates in the form dd/mm/yyyy, or may be encoded values that represent 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

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| Event Date | eventdate | Date associated with the event, as entered by the GP | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| System Date | sysdate | Date the event was entered into Vision | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| Consultation Type | constype | Code for the category of event recorded within the GP system (e.g. management or administration) | Lookup SED | INTEGER | 3 |
| Consultation Identifier | consid | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid | Link **Consultation** table | INTEGER | 20 |
| Medical Code | medcode | CPRD unique code for the medical term selected by the GP | Lookup Medical Dictionary | INTEGER | 20 |
| 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 | INTEGER | 20 |
| Source | source | Classification of the source of the referral e.g. GP, Self | Lookup SOU | INTEGER | 2 |
| NHS Speciality | nhsspec | Referral speciality according to the National Health Service (NHS) classification | Lookup DEP | INTEGER | 3 |
| FHSA Speciality | fhsaspec | Referral speciality according to the Family Health Services Authority (FHSA) classification | Lookup SPE | INTEGER | 3 |
| In Patient | inpatient | Classification of the type of referral, e.g. Day case, In patient | Lookup RFT | INTEGER | 2 |
| Attendance Type | attendance | Category describing whether the referral event is the first visit, a follow-up etc. | Lookup ATT | INTEGER | 2 |
| Urgency | urgency | Classification of the urgency of the referral e.g. Routine, Urgent | Lookup URG | INTEGER | 2 |

## 8. Immunisation

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| Event Date | eventdate | Date associated with the event, as entered by the GP | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| System Date | sysdate | Date the event was entered into Vision | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| Consultation Type | constype | Code for the category of event recorded within the GP system (e.g. intervention) | Lookup SED | INTEGER | 3 |
| Consultation Identifier | consid | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid | Link **Consultation** table | INTEGER | 20 |
| Medical Code | medcode | CPRD unique code for the medical term selected by the GP | Lookup Medical Dictionary | INTEGER | 20 |
| 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 | INTEGER | 20 |
| Type | immstype | Individual components of an immunisation, e.g. Mumps, Rubella, Measles | Lookup IMT | INTEGER | 4 |
| Stage | stage | Stage of the immunisation given, e.g. 1, 2, B2 | Lookup IST | INTEGER | 2 |
| Status | status | Status of the immunisation e.g. Advised, Given, Refusal | Lookup IMM | INTEGER | 3 |
| Compound | compound | Immunisation compound administered – may be a single or multi-component preparation, e.g. MMR | Lookup IMC | INTEGER | 4 |
| Source | source | Location where the immunisation was administered, e.g. In this practice | Lookup INP | INTEGER | 3 |
| Reason | reason | Reason for administering the immunisation, e.g. Routine measure | Lookup RIN | INTEGER | 3 |
| Method | method | Route of administration for the immunisation, e.g. Oral, Intramuscular | Lookup IME | INTEGER | 3 |
| Batch Number | batch | Immunisation batch number | Lookup BatchNumber | INTEGER | 20 |

## 9. Test

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| Event Date | eventdate | Date associated with the event, as entered by the GP | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| System Date | sysdate | Date the event was entered into Vision | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| Consultation Type | constype | Code for the category of event recorded within the GP system (e.g. examination) | Lookup SED | INTEGER | 3 |
| Consultation Identifier | consid | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid | Link **Consultation** table | INTEGER | 20 |
| Medical Code | medcode | CPRD unique code for the medical term selected by the GP | Lookup Medical Dictionary | INTEGER | 20 |
| 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 | INTEGER | 20 |
| Entity Type | enttype | Identifier that represents the structured data area in Vision where the data was entered | Lookup Entity | INTEGER | 5 |

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

4 fields:

| Data 1 | data1 | Qualifier | Lookup TQU | INTEGER | 3 |
|---|---|---|---|---|---|
| Data 2 | data2 | Normal range from | None | NUMERIC | 16.3 |
| Data 3 | data3 | Normal range to | None | NUMERIC | 16.3 |
| Data 4 | data4 | Normal range basis | None | NUMERIC | 16.3 |

7 fields:

| Data 1 | data1 | Operator | Lookup OPR | INTEGER | 3 |
|---|---|---|---|---|---|
| Data 2 | data2 | Value | None | NUMERIC | 16.3 |
| Data 3 | data3 | Unit of measure | Lookup SUM | INTEGER | 4 |
| Data 4 | data4 | Qualifier | Lookup TQU | INTEGER | 3 |
| Data 5 | data5 | Normal range from | None | NUMERIC | 16.3 |
| Data 6 | data6 | Normal range to | None | NUMERIC | 16.3 |
| Data 7 | data7 | Normal range basis (or peak flow device for entity type 311) | Lookup POP (or PFD) | INTEGER | 2 |

8 fields:

| Data 1 | data1 | Operator | Lookup OPR | INTEGER | 3 |
|---|---|---|---|---|---|
| Data 2 | data2 | Value | None | NUMERIC | 16.3 |
| Data 3 | data3 | Unit of measure | Lookup SUM | INTEGER | 4 |
| Data 4 | data4 | Qualifier | Lookup TQU | INTEGER | 3 |
| Data 5 | data5 | Normal range from | None | NUMERIC | 16.3 |
| Data 6 | data6 | Normal range to | None | NUMERIC | 16.3 |
| Data 7 | data7 | Normal range basis | Lookup POP | INTEGER | 2 |
| Data 8 | data8 | Expected delivery date (entity type 284) / Weeks (entity type 154) | dd/mm/yyyy None | DATE INTEGER | dd/mm/yyyy 10 |

## 10. Therapy

| Column name | Field name | Description | Mapping | Type | Format |
|---|---|---|---|---|---|
| Patient Identifier | patid | Encrypted unique identifier given to a patient in CPRD GOLD | None | INTEGER | 20 |
| Event Date | eventdate | Date associated with the event, as entered by the GP | dd/mm/yyyy **1** | DATE | dd/mm/yyyy |
| System Date | sysdate | Date the event was entered into Vision | dd/mm/yyyy **1** | DATE | 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 | INTEGER | 20 |
| Product Code | prodcode | CPRD unique code for the treatment selected by the GP | Lookup Product Dictionary | INTEGER | 20 |
| 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 | INTEGER | 20 |
| Dosage Identifier | dosageid | Identifier that allows dosage information on the event to be retrieved. Use the Common Dosages Lookup to obtain the anonymised dosage text, and extracted numerical information such as daily dose. | Lookup Common Dosages | CHAR | 64 |
| BNF Code | bnfcode | Code representing the chapter & section from the British National Formulary for the product selected by GP | Lookup BNFCodes | INTEGER | 5 |
| Total Quantity | qty | Total quantity entered by the GP for the prescribed product | None | INTEGER | 20 |
| Number of Days | numdays | Number of treatment days prescribed for a specific therapy event | None | INTEGER | 20 |
| Number of Packs | numpacks | Number of individual product packs prescribed for a specific therapy event | None | INTEGER | 8 |
| Pack Type | packtype | Pack size or type of the prescribed product | Lookup PackType | INTEGER | 10 |
| 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 $\geq 1$ denotes the issue number for the prescription within a repeat schedule | None | INTEGER | 20 |