|  |  |  |  |
| --- | --- | --- | --- |
| Document filename: | **GP Data Extract Supplier Implementation Guidance** | | |
| Project / Programme | **GPDfSU** |
| Document Reference | **GPDFSU-GPDATA-EIS-2** | | |
| Project Manager | **Andrew Thorne Marsh** | Status | **Published** |
| Owner | Andrew Thorne Marsh | Version | **0.76** |
| Author | **David McAvenue** | Version issue date | **17/08/2020** |

**GP Data Extract Supplier Implementation Guidance**

Document management

Revision History

|  |  |  |
| --- | --- | --- |
| Version | Date | Summary of Changes |
| 0.1 | 25/09/18 | First draft |
| 0.2 | 15/10/18 | Revised following adoption of ‘snapshot only’ approach |
| 0.3 | 29/11/18 | Minor Updates |
| 0.4 | 15/01/19 | Make Effective Date optional – avoid null Flavor dates, update diagram for Problem-Links |
| 0.7 | 08/03/19 | Added validation guidance, removal of condition, renaming of elements/tables |
| 0.71 | 13/05/19 | Minor update |
| 0.72 | 13/05/19 | Minor updates |
| 0.73 | 14/05/19 | Minor updates |
| 0.74 | 05/02/20 | Updates for schema 0.74 and extract pilot early implementation lessons learned |
| 0.75 | 21/02/20 | Corrected incorrect Conceptual Data Model Diagram included in version 0.74 |
| 0.76 | 17/08/20 | Up-versioned to Published status – no content changes over 0.75 |

Reviewers

This document must be reviewed by the following people:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Approved by

This document must be approved by the following people:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Signature | Title | Date | Version |
| Andrew Thorne-Marsh |  | GPDfSU Programme Manager | 17/08/20 | 0.76 |
|  |  |  |  |  |

References

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | Name | Reference | Date | Version |
| 1 | Primary Care Standard Dataset Specification |  |  |  |
| 2 | GP Data Technical Output Specification |  |  |  |
| 4 | GP Data Extract Validation Products |  |  |  |
| 5 | GP Data Extract Implementation Guide |  |  |  |
| 6 | GP Data De-Identification Toolkit Implementation Guidance | TBD |  |  |
| 7 | GPES Interoperability Standard | PD0009 | 14th Aug 2012 | 5.0 |
| 8 | GP Data Supplier Requirements |  |  |  |
| 9 | MESH Information | https://digital.nhs.uk/services/message-exchange-for-social-care-and-health-mesh#section-summary |  |  |
| 10 | MESH API | <https://nhsconnect.github.io/spine-mesh/develop_mesh_overview.html> |  |  |
| 11 | GP Appointments Data Collection (in support of Winter Pressures) Specification and Requirements | WPAD\_SPEC\_REQS\_1.0 | 7th Nov 2017 | 3.0 |

Glossary of Terms

|  |  |
| --- | --- |
| Term / Abbreviation | What it stands for |
| GPDfSU | GP Data for Secondary Uses |
| GPSS | General Practice Software Supplier |
| GPCS | General Practice Clinical System |
| GPES | General Practice Extraction Service |
| GPES-I | GPES Interoperability Standard |
| DQ | Data Quality |
| MESH | Message Exchange for Social Care and Health |
| MOLES | MESH On-line enquiry service |

Document Control:

The controlled copy of this document is maintained in the NHS Digital corporate network. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

Contents

[1. Introduction 5](#_Toc31802886)

[1.1. Background 5](#_Toc31802887)

[1.2. Document Scope and Related Documents 5](#_Toc31802888)

[2. Message Specification 6](#_Toc31802889)

[2.1. GP Data Extract Message 6](#_Toc31802890)

# Introduction

## Background

The GP Data for Secondary Uses (GPDfSU) requirements establishes a new technical approach for the extraction, transfer and controlled utilisation of GP Data for secondary uses via the NHS Digital Data Processing Services (DPS).



The GP data collection requirements provides a new standard for primary care data extraction following ‘extract once utilise many’ principles’. The ‘extract once utilise many’ approach allows a single extraction to satisfy multiple existing use cases without the need for burdensome ongoing development of additional extracts on source systems.

Data is de-identified at source and flows between source systems and the NHS Digital Data Processing Services (DPS).

Within DPS, landed data is processed into one or more segregated and secure utilisation domains capable of satisfying multiple utilisation use cases.

GP data collected through this service refers to existing GP data collections, with any potential new collections being subject to usual independent processes.

## Document Scope and Related Documents

This document provides technical guidance for suppliers generating and populating the GP data extract message and counterpart organisations receiving and processing GP data extract messages.

This document should be read in conjunction with other specification documents which provide the full technical specification for GP data.

* **GP Data Technical Output Specification (ToS)**

The Technical Output Specification provides the detailed specification for the GP data extract file, an XML transmission file which transports data from source system to landing platform.

* **GP Data Extract Interface Specification**

The Interface Specification defines the interface/contract between sending systems and the Landing Platform component of receiving datastores. This guidance document and Technical Output Specification specify the population of the extract payload whereas the interface specification describes the transmission of the payload between source system and receiver and provides the overarching principles for the interface.

* **GPDfSU GP Data Validation Products**

Validation products which allow sending and receiving systems to validate conformance of GP Data extract files against the specification.

* **De-Identification Toolkit Implementation Guidance**

Provides guidance to suppliers on de-identification using the NHS Digital supplied de-identification toolkit.

This document is intended to be read by supplier technical and assurance teams responsible for populating the GP Data Extract.

The document is intended to be read by Authority technical and assurance teams responsible for receiving and processing GP data extracts.

# Message Specification

## GP Data Extract Message

### Wrapper Structure

A screenshot of a social media post

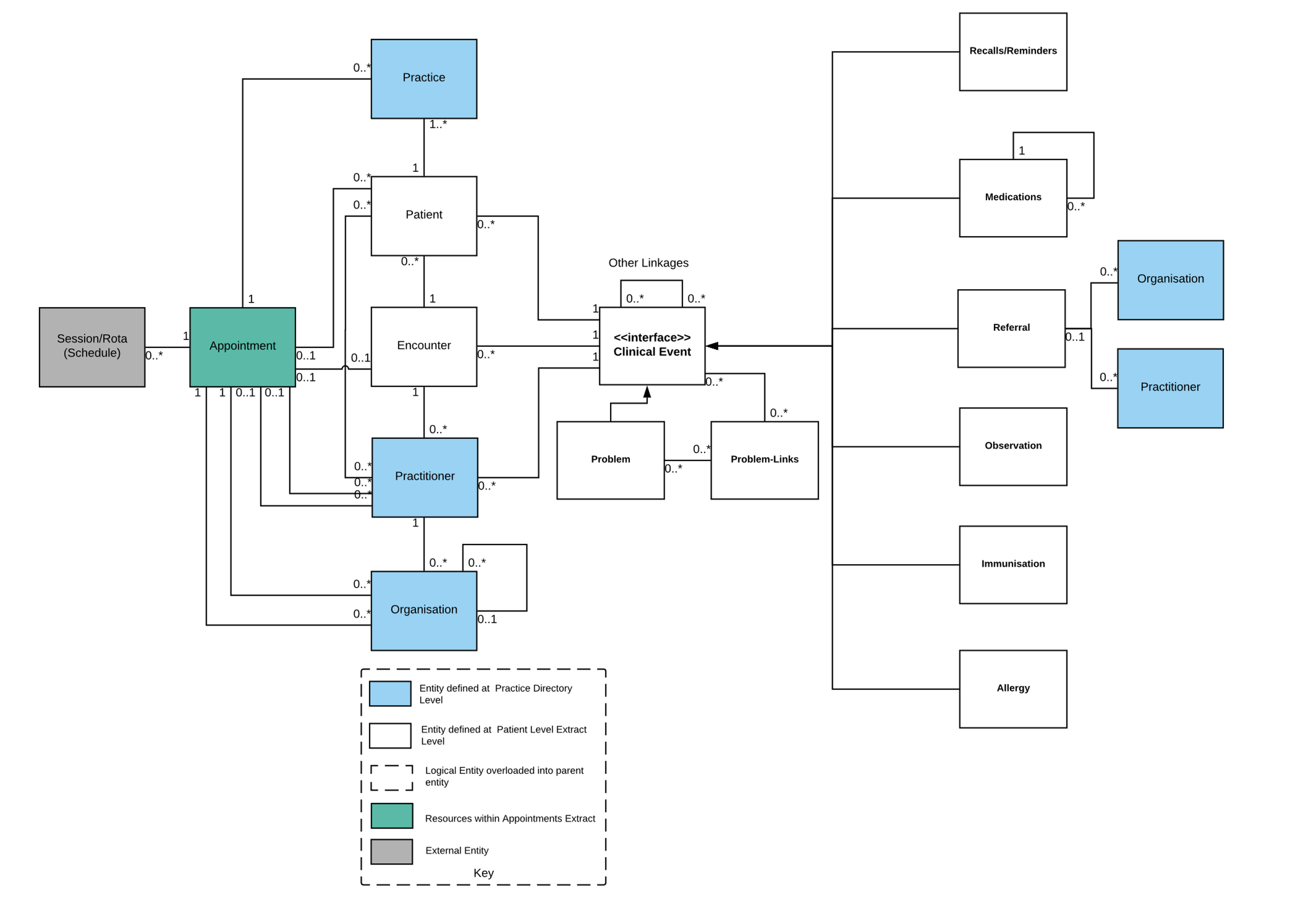
Description automatically generated

Each extract transmission file sent from GPCS to DPS contains a single GP Data Extract.

The extract is organised into a series of hierarchical wrapper elements which provide flexibility over the aggregation and transmission of extracted GP Data e.g. allows single practice/provider extraction or the aggregation of multiple extracts from practices/providers into single transmission files.

* Each GP Data transmission file contains a single GP Data Extract.
* Each GP Data Extract is made up of a single Practice Extracts[[1]](#footnote-2)
* By convention the appointment slot extract and patient level extract flows are segregated (separate flows). A Practice Extract, will contain either an Appointments extract or multiple Patient Level extracts but not both.
* Each Practice Extract contains a Practice Directory which provides information on the organisations and practitioners referenced by the contained patient level or appointment extracts.
* The inclusion of a single Practice Directory avoids redundancy in the case of Patient Level extracts where Patient Level data originating the same practice will tend to reference the same practitioners many times.
* Patient Extracts are made up of multiple resource entries that correspond to record entries in source patient records.
* Appointment extracts may contain supplier provided Appointment reference data that provides supplier and practice specific mappings that will assist in the utilisation and interpretation of extracted appointment data.
* The resources representing patient records, appointments, practitioners and expressed within Patient and Appointment Slot Extracts are modelled according to the Primary Care Data Model (below)

### Primary Care Data Model



Each GP Data Extract is made up of resource records (message elements) which represent the entities and relationships in the Primary Care Data model shown.

### Message Mapping

|  |  |  |  |
| --- | --- | --- | --- |
| **Entity** | **Description** | **GP Data Extract Message Element** | **FHIR Resource/Profile Mapping** |
| n/a | Overall wrapper element for each GP Data Extract. Carries appropriate identifiers and routeing information. | gpdata-extract | n/a |
| n/a | Wrapper for grouping of patient level extracts at practice level | practice-extract | n/a |
| n/a | Wrapper for attributed practitioners and organisations conveyed at practice level | practice-directory | n/a |
| n/a | Wrapper for Patient Extracts | practice-patient-extracts | n/a |
| n/a | Patient level extract – snapshot | patient-extract-snapshot | n/a |
| n/a | Wrapper for practice appointments data carried at practice level | slot-extract | n/a |
| n/a | Supplier and practice specific reference data which may be used to interpret extracted appointment data | appointment-reference-data | n/a |
| Practice | The GP practice corresponding associated with the extract. | organisation-table | CareConnect-Organization-1 |
| Organisation | Represents organisations associated with the patient record e.g. organisations referred to, organisations associated with practitioners involved in the care of the patient | organisation-table | [CareConnect-Organization-1](https://fhir.hl7.org.uk/STU3/StructureDefinition/CareConnect-Organization-1) |
| Practitioner | A practitioner attributed to record entries in the source system. Typically the clinical performer or authoriser attributed by the source record entry or in the case of referrals the referral target. | practitioner-table | CareConnect-GPC-Practitioner-1 |
| Patient | Represents an individual Patient providing an identifier, limited demographic information and information on the Patient’s registration | patient-table | CareConnect-GPC-Patient-1 |
| Encounter | Represents an encounter (consultation) between a care professional and patient. | encounter-table | Encounter |
| Recalls/Reminders | Represents a planned diarised activity or procedure for which a reminder or recall will be generated. | recall-table | CarePlan |
| Medication | Represents the authorisation or issue of a medication | medication-table | CareConnect-GPC-MedicationRequest-1 |
| Referral | Inbound or outbound referral | referral-table | ReferralRequest |
| Problem | Represents a defined problem header in source record | problem-table | CareConnect-GPC-ProblemHeader-1 |
| Observation | Used to represent lifestyle values, test results, signs and symptoms, procedures Utilised as the default record type for source record entries which are not mapped to a more specific type | observation-table | CareConnect-Observation-1 |
| Immunisation | Represents the recording of an Immunisation procedure in source record | immunisation-table | CareConnect-GPC-Immunization-1 |
| Allergy | Represents an allergy, intolerance or adverse reaction recorded in the source record | allergy-table | CareConnect-GPC-AllergyIntolerance-1 |
| Appointment | Represents an appointment slot and appoint booking for that slot | appointment-table | *n/a* |
| Linkages | Each element type may have specific additional attributes that are used to represent linkages that exist in the source record. Examples are the linkage between a medication issue and its authorisation or the linkage between an entry made within a date entry template and the record representing the template header. | n/a | n/a |
| Problem-Links | Provides the explicit linkage between a record entry and a problem header. A record entry may be linked to more than one problem. | problem-link-table | n. |

1. There is no additional entity utilised to represent a binding between a practitioner and the organisation they are acting for (FHIR PractitionerRole equivalent) resulting in potential redundancy where the same individual is present and working in multiple roles for different organisation. This will be a rare occurrence.
2. Currently the primary care record has been partially mapped by GP Connect. The GP Data Extract ToS has been based on published and draft GP Connect profiles. Where no specific FHIR profile is available the base resource has been referenced. These will be updated to specific profiles when these are available from GP Connect.
3. Source system Procedure and Family History record entries are overloaded into the default Observation entity. *This is because the distinction is primarily a terminological distinction on source systems and may therefore the same distinction can be made by receiving data stores. May change approach if GP Connect models and implements these resources.*
4. Candidate FHIR resources for representing appointment data are booking oriented and have limited support for the additional fields relating to capacity/availability of the Winter Pressures dataset. As a result, FHIR resource mappings have been omitted for these entities.
5. Where Family History record entries are extracted as Observation, no additional information on the subject of the family history will be provided beyond any information that may be inferred from the code associated with the record entry.
6. Where an organisation is an organisation with an ODS code only minimal organisational information is carried beyond the organisation code as further organisation information is derivable within the data store from the ODS code or at the point of utilisation.

### Message Conventions

#### Practice Directory Inclusion

The content of the practice directory should be limited to those organisations and practitioners that are referenced by record entries within extracted patient records.

#### Resource Mapping and Granularity

More granular representations of the patient record using additional FHIR resource types are possible and envisaged in direct care domains such as GP Connect/Care Connect e.g. introducing FamilyHistory, Procedure etc … Such granular decompositions are challenging to achieve consistently across participating systems and are of limited benefit in a secondary use domain like GP Data where the ‘type’ of the record entry can be inferred from the coding of the record entry. In GP Data therefore record types have been decomposed into resources which have distinct meanings or remove ambiguity about the intent or type of the record entry. For example, medications and allergies have distinct type and purpose, it is also useful to distinguish between elements like recalls (CarePlans) which may be coded as procedures and actual recording of procedures. Where no explicit resource type is specified for a record entry, the Observation resource should be used as the default for all otherwise uncategorised record types in source systems. The overloading of uncategorised ‘journal’ events into Observation does not imply that such resources are coded as ‘Observables’ or that all Observations have values.

#### Element and Attribute Naming

Element and attribute naming follows an XML minimisation approach such that most element and attributes adopt a 2 or 3 character short naming convention. This approach mitigates to a degree the data volume concerns around the use of XML (in uncompressed form) for a data intensive large data volume movement application such as GP Data. For understandability, long names are used throughout the accompanying documentation and the corresponding short names are specified in the Technical Output Specification.[[2]](#footnote-3)

#### GP Connect Implementation Guidance

Unless otherwise specified the mappings and implementation guidance applying to GP Connect apply to the mapping from supplier systems to the extract.

#### Supplier Mapping

The Authority will work with suppliers to agree appropriate map record structure and content appropriately to the specified extract message structure. Such mappings may for example attempt to maximise the proportion of source records that is extractable in coded form. Another goal of the mapping activity will be to maximise semantic consistency across participating systems. This activity for example adopts established conventions from other domains e.g. GP2GP and/or conventions employed in existing GPES extracts/QoF

#### Null or Empty Fields

Null or otherwise empty optional fields should be suppressed to avoid empty attributes appearing in the extract.

If there is potential for mandatory fields to generate null or empty data then appropriate mapping or handling at source should be used to avoid the mandatory field being suppressed or appearing with empty content.

#### Null/Empty Values and De-ID fields

All De-Id fields are optional in the extract.

If a De-ID field is null or empty then no transit token should be generated and the field should not appear in the extract.

#### Implicit Coding

There are instances among participating systems where clinical content is implicitly coded via specific modules or structures. Examples are systems which record immunisations or vaccinations within locally defined immunisation modules without directly visible coding in Snomed or other coding systems. It is expected that where such implicit coding is performed for clinical content that it will be possible to map the content to Snomed CT in at the point of extraction.

Other examples of implicit coding are systems containing complex forms which implicitly contain or express coded concepts that are not explicitly displayed or recorded by the system. An example might be an over-arching ‘Smoking’ form that captures lifestyle values and history about smoking habits. Such a form may contain information like a ‘Date stopped smoking’ field accepting a date. In cases like this it may be appropriate to extract this information in the form of explicit codes that correspond to the implicit coding of information in the form e.g. 160625004 |Date ceased smoking (observable entity)| with the date as a value. This approach may be necessary to allow GP Data to fulfil the same role as existing GPES extractions without complex additional and system specific logic within DPS analytics i.e. if implicit codings are already mapped to external codes and values in existing GPES extractions then it may be appropriate to extract implicitly coded data in the GP Data Extract in a form that respects those existing mappings.

#### Blood Pressures

There are a variety of blood pressure concepts in use across participating systems e.g. Sitting, Standing, Target, Average, Pre-Treatment etc …

Although the Observation record type in the GP Data extract has support for two values that could be used to express all blood pressures in a rolled up consolidated form that would involve an unacceptable loss of information about the original coding of the blood pressure and also potentially require elimination of the systolic and diastolic component that were rolled up into the Observation. For that reason the GP Data approach for blood pressures is to extract them in a componentised rather than rolled up form such that a parent Observation code representing the blood pressure reading has a whole is linked to Observations carrying the Systolic and Diastolic values, with each Observation appropriately coded for the blood pressure concept recorded on the system. Linkage and grouping between the Observations is performed via the template\_id/@tid linkage.

#### Proprietary Qualifiers and Other Structure

The GP Data and underlying FHIR Resource models from which GP Data is derived, do not fully model all of the data elements or structure associated with record entries on source systems. Examples of additional structure are additional flags, qualifiers and other information that are routinely recorded as part of the structure of source systems e.g. a system that records blood pressures with additional information like Korotkoff sounds, cuff size or posture. If this proprietary structure is already available in over secondary use extracts then it may also be appropriate for this information to be extracted in a structured firm. The mechanism to do this would be to encode the proprietary structure within the other\_expression (@ote) field of the GP Data Codeable Concept type. Alternative approaches such as extracting as separate appropriately coded Observation records are also possible (see ***Implicit Coding*** above).

#### Persistent Identifier

A globally unique persistent event identifier (Uuid) is utilised as the primary identifier for record entries represented conveyed in GP Data Extracts. Whenever the same source record entry is represented in a GP Data extract the same identifier will be used. This means that the same identifier can be used to link each occurrence of a record entry (creation, modification, logical deletion) in receiving systems.

The scope of the persistent identifier is dependent on the architecture of source systems e.g. may have practice scope or in the case of an enterprise single record system its scope may span the lifetime of the patient within that system boundary.

The identifiers used should be distinct from the unique statement identifiers utilised in other domains (GP2GP, GP Connect).

#### PMIP and Test Requests

The extraction of information relating to the original pathology reports providing the test results that are filed within patient records is out of scope. Where test results are exported in the form of observation-table records then these reflect the ‘filed’ results only.

Where Test Requests are recorded in the patient record as coded journal events then these will be extracted like any other record entry.

#### Attachments

Although the specification specifically excludes freetext and attachments, the coded element of attachments which appears as journal entries in participating systems should be extracted.

#### Optionality of Patient Resource

Patient extracts always contain a single Patient resource (patient-table)

#### Effective Date

A single consistent datetime attribute (effective\_date/@ed) is associated with most resource entries. This provides the most clinically significant, user modifiable date associated with the source record. In most situations there will be a single date available in the source system which meets these criteria. On some systems there may be more than one candidate date available in relation to specific concepts e.g. information about the Last epileptic seizure suffered by a patient, may have an availability date (when the information was made available to the recorder) and the date of the Last epileptic seizure itself. In the GP Data extract it is the date of the last seizure that should be supplied as the effective date.

Where no clinically effective date is available then the effective\_date attribute should not be supplied.

#### Dates and Times

Although it is possible on most participating systems to record record entries with partial dates (year only or year and month) these should be padded default in extraction to provide full datetimes to at least second precision. The ToS provides a full definition of the allowed DateTime formats. Padding will pad year only dates to 1st January and year and month dates to the first of the month (time component will be midnight).

It is expected that effective dates will always be interpreted as a local time i.e. the time displayed to users even if a time zone components is supplied.

Recorded dates (audit trail) and other time stamp fields (e.g. file timestamps) may be interpreted using any supplied time zone information.

#### Attribution

The primary attribution of a record entry to a responsible clinician or performer is extracted and expressed via the practitioner\_id/@pid attribute. This may be distinct from the user who may have actually entered the record on the source system.

Where no clinical attribution is available for a record entry then the Practitioner will still be used to represent the attributed use but the HCP Type will be coded as ‘D’ Other Practice Staff.

#### Free-text notes

General purpose free-text notes are excluded by design from the specification. Where content that may be sourced from user entered text has been included care has been taken to ensure that source data fields are in normal usage utilised for specific purposes e.g. immunisation batch/lot number or medication dosage. While this minimises the risk of information leakage via freetext occurring it cannot fully prevent it if the fields on the source system have been used incorrectly. Consequently, some monitoring of received extracts may be appropriate to identify whether potential information leakage is occurring.

#### Non-Consultation Record Entries

Some participating systems allow record entries to be made outside of a consultation context (encounter-table/et entity).

Linkage from event record entities to encounters is optional and where a record entry is made outside of a consultation context the encounter\_id/@eid is not supplied.

#### Push Down Attribution

Each record entry level resource will always have an attribution to a practitioner. Where there is no attribution at record entry level the attribution of the associated encounter will be pushed down to the record entry. Where there is already attribution at record entry level this will not be overridden by the attribution at consultation level.

#### Terminology Support

The extract supports multiple terminologies including supplier managed/proprietary terminologies (‘local’ codes).

Local code support is provided to satisfy some existing local rather than national use cases and to allow analysis of concepts which are often coded in a proprietary form e.g. allergies from some systems,

Standalone local codes will only be provided whether there is no alternative code in another recognised terminology.

Source systems may use the user\_selected/@us indicator to identify the ‘user selected code’ i.e. the code system/terminology the code that was originally selected and entered by a user (and related term/rubric) belonged to. The user\_selected attribute should only be supplied when it can be reliably determined that the user selected a code from the specified terminology.

By convention it is expected that codes will be extracted in the terminology that was originally used to record the coded information e.g. legacy Read v2 or CTV3 codes along with a Snomed translation. When information is recorded directly in Snomed CT only the Snomed code will be extracted.

By convention, both the Snomed Concept Id and Description Id are extracted where possible.

#### Expression of term/rubric text

Although the coded information extracted in the GP Data extract is sufficient to support most analytical use cases, the extract also supports extraction of the term text displayed to the user in the form of the rubric (@rb) attribute.

This feature is included because on some systems the term text displayed to the user may be different to the term text corresponding to the coded concept recorded in the underlying terminology.

For consistency, the rubric/@rb is always supplied (redundantly) even if the term text displayed to the user is the same as that indicated by the coding of the record entry.

#### Modifiers

It is recognised that some resources in the FHIR standard and subsequent curation of FHIR resources within CareConnect and GP Connect have in rare cases adopted the use of modifier concepts for particular resources. Examples of modifying concepts within FHIR are ‘notGiven’ and accompany ‘explanation’ reasons which when utilised significantly modify the meaning of the accompanying coded information extracted in GP Data. These features may for example be employed within Immunisations to convey a reason like lack of consent for not administering a vaccine.

Although modifier concepts are rare in primary care records, modifiers **are not** supported in GP Data because of the undesirable potential for a secondary attribute or value to modify the meaning of extracted codes which would then oblige all downstream analysis to be reliably modifier aware.

The authority will work with suppliers during the extract design and implementation phase to identify potential modifiers and adopt different approaches for the GP Data extract.

A typical approach to avoid modifiers is to reverse the coding such that the modifier concept becomes the main coded concept extracted qualified by the main code e.g. in the case of an immunisation procedure code modified to not given by reason of lack of consent the ‘refused consent’ code would become the main code, qualified by the immunisation procedure.

### Technical Output Specification

The Technical Output Specification (TOS) [2] provides the detailed field level description of the extract message and should be read in conjunction with this interface specification.

1. ‘Practice Extract’ in the context of appointment data means a provider of appointment capacity in primary care – it does not necessarily correspond to a GP practice. It may represent a community provider or Extended Access Provider. [↑](#footnote-ref-2)
2. The effect of minimised attribute names on compressed XML is minimal as all commonly occurring strings whether small or large will tend to be efficiently compressed [↑](#footnote-ref-3)