|  |  |  |  |
| --- | --- | --- | --- |
| Document filename: | **GPDfSU GP Data Interface Specification** | | |
| Project / Programme | **GPDfSU** |
| Document Reference | **GPDFSU-GPDATA-EIS-1** | | |
| Project Manager | **Andrew Thorne Marsh** | Status | **Published** |
| Owner | Andrew Thorne Marsh | Version | **1.1** |
| Author | **David McAvenue** | Version issue date | **17/08/2020** |

**GPDfSU GP Data Interface Specification**

Document management

Revision History

|  |  |  |
| --- | --- | --- |
| Version | Date | Summary of Changes |
| 0.1 | 20/04/18 | First draft |
| 0.94 | 16/10/18 | Updated for snapshot only approach |
| 0.95 | 21/11/18 | Updated to remove MPS match for MPS |
| 0.96 | 21/01/19 | Renamed DPS |
| 0.97 | 08/03/19 | Added outbound validation pipeline |
| 0.98 | 14/05/19 | Minor updates |
| 1.0 | 05/02/20 | Updated references  Added S3 transport  Updated acknowledgement example XML to include schema  Removed MESH application level retry behaviour  Removed transmission checksum  MESH Metadata (Subject) changed to included effective date  Updated reliability and recovery approaches  Added potential approach to delta registration type checking  Added Extract Report  Update sequence validation to monitoring  Removed patient extract sequencing |
| 1.1 | 17/08/20 | Text updates to published status – no content changes |

Reviewers

This document must be reviewed by the following people:

|  |  |  |  |
| --- | --- | --- | --- |
| Reviewer name | Title / Responsibility | Date | Version |
| Andrew Thorne-Marsh | GPDfSU Programme Manager | 05/02/20 | 1.0 |
| Philip Gaertner | GPDfSU Technical Architect | 05/02/20 | 1.0 |
| Jon Smith | DPS Technical Architect | 05/02/20 | 1.0 |
| John MacDonald | GPDfSU Business Analyst | 05/02/20 | 1.0 |
| Richard McEwan | GP IT Futures Lead Architect | 05/02/20 | 1.0 |

Approved by

This document must be approved by the following people:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Signature | Title | Date | Version |
| Andrew Thorne-Marsh |  | Programme Manager | 17/08/20 | 1.1 |
|  |  |  |  |  |

References

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | Name | Reference | Date | Version |
| 1 | GPES Interoperability Standard | PD0009 | 14th Aug 2012 | 5.0 |
| 2 | MESH Information | https://digital.nhs.uk/services/message-exchange-for-social-care-and-health-mesh#section-summary | n/a | n/a |
| 3 | MESH API | <https://nhsconnect.github.io/spine-mesh/develop_mesh_overview.html> | n/a | n/a |
| 4 | 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](#_Toc26356952)

[1.1. Background 5](#_Toc26356953)

[1.2. Document Scope and Related Documents 5](#_Toc26356954)

[2. Out of Scope 6](#_Toc26356955)

[3. Context 7](#_Toc26356956)

[3.1. Overview 7](#_Toc26356957)

[3.2. Key Assumptions 11](#_Toc26356958)

[3.3. Operating Principles 12](#_Toc26356959)

[4. Extract Transmission 22](#_Toc26356960)

[4.1. Principles 22](#_Toc26356961)

[4.2. Message Sizing 23](#_Toc26356962)

[4.3. Compression 23](#_Toc26356963)

[4.4. Networking 23](#_Toc26356964)

[4.5. MESH Addressing 23](#_Toc26356965)

[4.6. Workflow Ids 24](#_Toc26356966)

[5. Message Specification 27](#_Toc26356967)

[5.1. GP Data Extract 27](#_Toc26356968)

[5.2. GP Data Ack Message 27](#_Toc26356969)

# 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).[[1]](#footnote-2)

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 defines the technical interface and contract between the GP data extraction Subsystem within source systems and the Landing Platform, the component of DPS responsible for receiving, validating and processing GP data extracts prior to utilisation.

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

* **GPDfSU GP Dataset Supplier Requirements**

The primary specification of supplier functional and non-functional requirements

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

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

* **GP Data Extract Implementation Guidance**

Provides implementation guidance for producers and consumers of the GP data extract file.

* **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 implementing the GP data subsystem within their solutions.

This document is intended to be read by Authority technical and assurance teams responsible for implementing the GP data ‘Landing Platform’ component of DPS.

# Out of Scope

**Information Governance**

GP data contains at source potentially sensitive patient identifiable information and the commissioning and enablement of these data flows in de-identified form requires approval and authorisation by the appropriate statutory, regulatory and professional bodies as well as source system data controllers.

**Service and Operational Management**

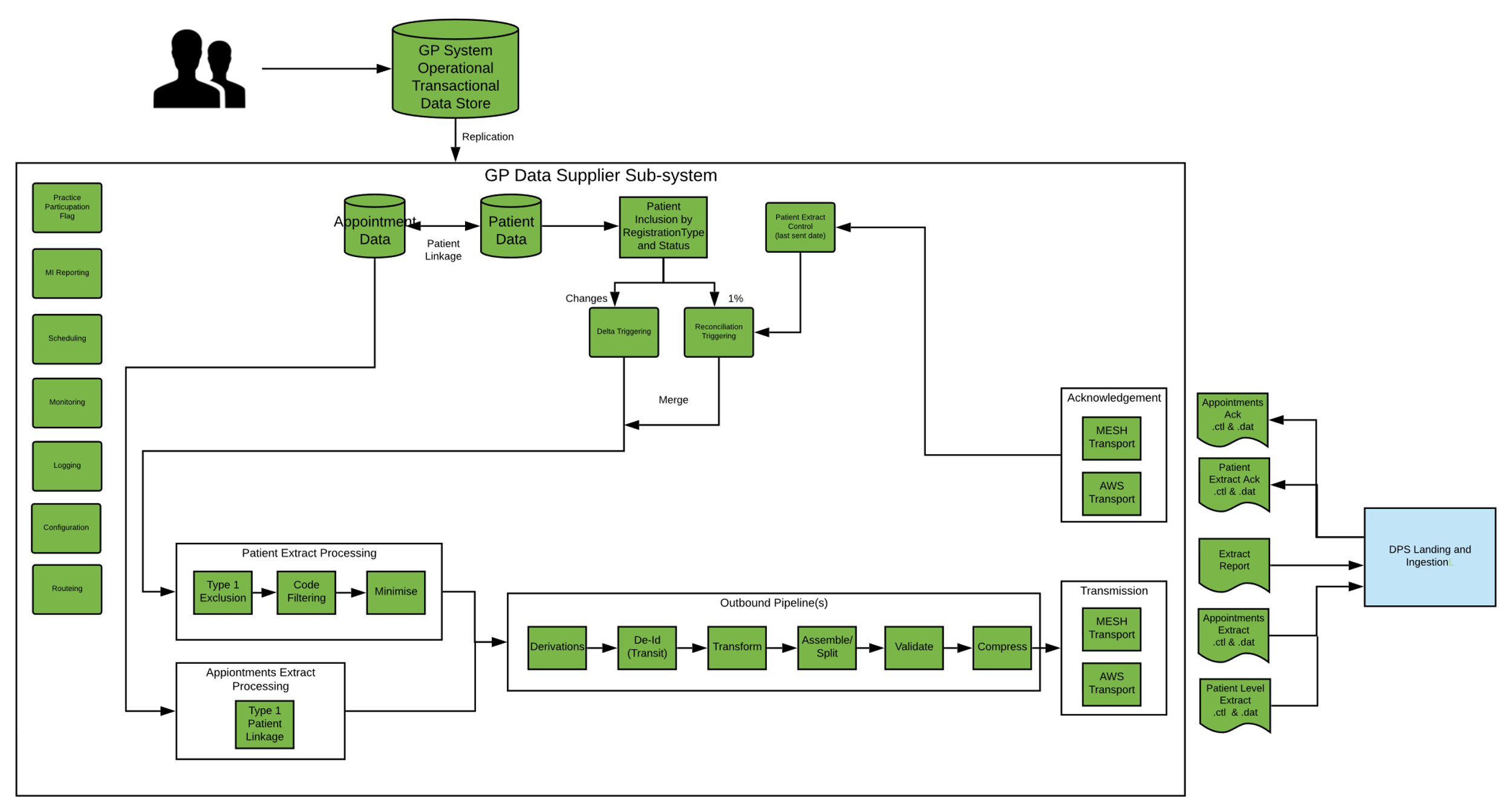
It is assumed that the interface and transmission components described in this document are supported by the appropriate technical and process components to successfully monitor, operate and support the solution. At a minimum, appropriate components to log system activity and exceptions and generate appropriate alerts to initiate operational and service management action by supplier staff should be available and utilised by the solution. The support and operations model for the solution including the responsibilities of supplier and the Authority will be specified separately.

**Transactional to Secondary Data Store Replication**

It is expected that data extraction for GP Data will operate on supplier secondary data stores rather than the supplier operational/transactional data store due to the potential negative impact on user experience and response times from extracting high volumes of GP Data from the source system operational data store. The replication mechanism by which the supplier system synchronises the operational data store with the data store from which GP Data is extracted will be supplier specific and outside of the scope of this specification. The characteristics of the replication mechanism in terms of latency, scalability, reliability is however relevant and will be discussed with suppliers as part of the overall design engagement for the solution.

# Context

## Overview



### Extract Scope

1. Two distinct extracts are incorporated within the GP data Standard Extract
   1. Patient level extract – a comprehensive de-identified extract of patient records e.g. Problems, Observations, Immunisations, Referrals and others. Freetext notes and document attachments are specifically excluded.
   2. Appointment extract – an extract of appointment data, with a de-identified patient linkage where applicable.
2. Patient Level Extracts are full point in time (snapshot) extracts of patient records containing all content within the scope of the extract.
3. Appointment extracts are incremental, containing changed appointment data since the last extraction.
4. Scheduling and frequency of appointment and patient level data extraction is independently configurable but operationally will be configured to the extraction schedule specified in the **GPDfSU GP Dataset Supplier Requirements.**
5. Patient level extracts and appointment extracts are segregated and flow to DPS as two distinct flows.
6. All extraction and submission is automated requiring no intervention or action by practices or (in normal operation) system suppliers.
7. Supplier systems signal completion of extraction generation and transmission activity for a given day by a delivering extract report files to NHS Digital. A secondary purpose of these files is to provide the daily management information for GP Data.

### Source Systems

1. Extraction of patient level data is targeted at instances of GP Computing Systems holding the recognised GP patient record
2. Extraction of appointment data is targeted at all systems holding appointment capacity for the delivery of primary care appointments. This includes instances of GP systems utilised by GP practices, community instances of GP systems, GP systems utilised by other providers of appointment capacity such as Extended Access Providers (EAPs) and third-party providers of appointment management applications.
3. Although the GP Data extract schema supports bundling of extracts from multiple provider organisations into a single appointment or patient extract transmission file, GP Data will operationally be based on single extract files per provider such that in any given day and in the absence of exceptions, each practice providing patient extracts and each provider generating appointment extracts will generate a single extract transmission file. This restriction supports simpler tracking and traceability compared with bundling of multiple provider outputs into single transmission files.

### Patient Level Extract

#### Extract Types

1. All patient level extracts are full extracts of the specified record content from the source patient record and structured according to the **GP Data Technical Output Specification (ToS).**
2. Patient level extracts are made up of two components
   * Reconciliation extracts – a synchronisation feed that caters for initial and potentially ongoing synchronisation of unchanged records with DPS
   * Delta extract – a change triggered feed that ensures that the ‘latest’ patient record is synchronised with DPS when relevant information in the patient record changes
3. The reconciliation extract is a bulk synchronisation process that supports initial synchronisation of patient records with DPS.

* The reconciliation extract can be viewed as synchronising the active patient population which is not interacting with their practice and hence with unchanging records that would not otherwise be synchronised with DPS.
  + The reconciliation extract is driven by a configurable ‘activity level’ - the percentage of qualifying patients to be extracted in the nightly extraction window.
  + The candidate patient records to be sent in any execution of the snapshot extract process are determined by the percentage configured and appropriate prioritisation of qualifying patient records in the source system. Prioritisation operates for on a ‘least recently sent’ basis where records where no patient level extract has been sent are prioritised first.
  + At a 1% activity level full synchronisation of source patient records with DPS would be achieved in approximately 100 days.

1. The reconciliation process may also be configured to run beyond the period of initial population of DPS proving an ongoing refresh of patient records This ongoing synchronisation means changes to the implementation of the extract, upgrades or changes to behaviour like altering data minimisation rules are handled by routine data refresh rather than requiring specific intervention/management by NHS Digital and suppliers. Whether to keep the reconciliation extract going in this manner is a business rather than technical decision.
2. The Delta extract ensures that changes to the patient record are synchronised with DPS such that DPS has the latest version of the patient record when relevant changes are made. Examples of changes are changes to the patient’s registration or demographic details and changes to the patient’s clinical record such as adding clinically coded information to the patient record or deleting or modifying existing record entries.
3. Delta triggering by registration change (type) ensures that DPS is notified when patient’s move between practices or die. Patient movement triggers an extract at the new practice and an extract from the old practice because there are delta triggering changes in both organisations.
4. Multiple patient record changes result in a single Patient level extract in the next scheduled extraction period and only one patient level extract is generated where the same patient record is flagged for both reconciliation and delta extraction in the same period.
5. It is expected that both bulk reconciliation and delta extracts will be scheduled to execute as a recurring nightly process although both extracts may be scheduled to run independently and at any specified time to support the ability to disable the reconciliation extract without affecting the delivery of change driven extracts.
6. There is no requirement that the reconciliation extraction should have occurred for a patient prior to sending delta extracts.
7. The reconciliation and delta extracts are technically the same in terms of structure and content. A supplier may opt to either bundle these feeds together in the same transmission files or split them. From a receiver’s perspective where reconciliation and delta feeds are segregated, it appears like the sort of file splitting that occurs when extract submission files are split by size i.e. multiple files submitted by a provider in each day.

#### Inclusion Criteria

1. The extraction process targets active patients having a defined set of registration types and statuses e.g. GMS Regular and Temporary, Active.
2. While the static set of registration inclusion criteria is adequate for determining inclusion in the reconciliation feed, it is inadequate for determining inclusion in the delta feed because the delta feed must also trigger when a patient’s registration changes to values that are outside of the inclusion criteria e.g. patient moves or dies. The delta inclusion logic must therefore utilise a patient’s registration history or other appropriate criteria in addition to their current registration type to determine whether a patient should be included.

#### Extraction Process

1. Type 1 Exclusions are applied at source such that records with a current type 1 opt-out are excluded from extracted even if they would otherwise qualify for inclusion in the reconciliation or delta components of the extract.
2. Codes are filtered by a configurable ‘blacklist’ of SNOMED CT Concept Ids. This removes record entries with codes present in the blacklist. The application of the filter is to provide a capability to remove legally sensitive and restrictive codes or apply data minimisation rules to the extract e.g. remove codes that are of limited utility. The precise configuration of the blacklist and whether it is enabled is still to be determined but the capability will be implemented.
3. A minimisation process is applied to snapshot extracts. This longitudinally limits the amount of historic information extracted in relation to specified content types e.g. a 10- year limit on medication history, no limits on diagnoses, problems.

### Appointment Extract

1. The recurring appointment extract is an incremental extract of all changed appointments since the last appointment extract.
2. The initial appointment extract from a system, is an appointment snapshot of all current and future existing appointments.[[2]](#footnote-3)
3. Inclusion by registration type and status is not applicable to the appointment extract but the type 1 status of a patient (where known) determines whether the linkage from an appointment to patient by NHS number is populated or not.

### End of Extraction Reports

When all extraction and transmission of each extract type is completed for a given extract day, the supplier solutions signals completion by sending an extract report file to NHSD.

An extract report file is generated for each extract type, one for appointments and one for the patient extraction process.

The report file acts as a signal to NHS Digital that from a supplier perspective the extraction process for the day is complete and is populated with the data required for management information.

The extract report file is not acknowledged by NHS Digital.

### Appointment and Patient Outbound Pipelines

This section provides a single description for the processing performed in the remainder of the outbound processing pipeline for brevity. There will be distinct pipeline instances for appointment and patient extracts and minor variations between the two pipelines as a result of the different data content in each extract.

1. Primary derivations are applied to the source extracts. These primary derivations are performed on PID data items which are de-identified at source to minimise the need for subsequent re-identification i.e. the less precise derivations of those PID fields can be utilised for analytical purposes instead of re-identifying PID data. Examples are blurring of date of birth to week and year of birth or post code to LSOA.
2. De-identification de-identifies patient identifying fields in the source extract e.g. NHS number or date of birth. De-identification is performed via an NHS Digital supplied de-identification toolkit. This process yields ‘transit pseudonyms’ which render the de-identified fields suitable for transmission across the organisational boundary between source system and DPS. The transit pseudonym provides an encrypted wrapper which ensures the de-identified field is not persisted or processed as clear text within the data store. Further processing steps within DPS generate usable pseudonyms from the supplied transit pseudonyms without requiring intermediate re-identification. Any subsequent re-identification of the de-identified fields is a separate process under appropriate governance.
3. In the case of the appointments extract the only de-identified field is the appointment to patient linkage by NHS number.
4. At some point in supplier processing, the appointment or patient extract is generated by transformation from the supplier’s internal data model to the specified output GP Data extract XML (this may occur at a different point to that shown in the supplier’s own implementation).
5. Multiple patient extracts are aggregated together to form the overall extract from the practice.
6. Depending on the transport used and the optimum file size for processing in DPS extract output files, extract output files may be split according to a configurable maximum file size, such that the daily output from a given provider may be split over multiple extract transmission files. In the case of the patient extract, a notional maximum extract size may be 1 Gb and in most daily extract operations this size would not be exceeded, however events like practice migrations and other bulk activities have the potential to vastly increase the number of changed patients within the delta feed and thus generate larger extract volumes in a given period which may require splitting.
7. Extracts are validated against the GP Data Validation Products before transmission (XML Schema Validation and Schematron rules)
8. All extract output is compressed prior to transmission (gzip).

### Transmission

DPS supports two inbound file transfer mechanisms and suppliers will adopt the most appropriate solution for the architecture of their solution.

1. MESH Transport (API or client)
2. AWS Transport

The AWS Transport option involves supplier systems writing extract files to and reading acknowledgements from a DPS landing bucket within DPS.

### DPS Landing and Ingestion

1. The receiving Landing Platform component of DPS presents a standard receiving interface for extract transmissions and accepts or rejects the extract transmission from the source GPCS.
2. The Landing Platform verifies technically validates received extracts (schema and rules based). Based on these checks received extracts are either accepted or rejected and an acknowledgement message signalling acceptance or rejection is returned to the sending system.
3. Once extracts have been accepted, the contained extracts are fully de-identified and extract data is further processed into segregated domains for utilisation in the context of specific use cases. This onward processing will may involve additional derivations and application of common and use case specific DQ rules. The detail of DPS processing is outside of the scope of this document.

### Acknowledgement Processing

1. Acceptance of an extract by the receiving landing platform, in the form of a successful extract acknowledgement asserts that responsibility for onward processing of the data contained in the extract has passed to the receiving system and is no longer a responsibility of the sending system. At that point the sending system can mark the transmission of the extracts making up the transmission file as complete.
2. Rejection of an extract represents a high priority technical failure of the system which will not generally be resolvable by automatically retrying or re-sending an extract. It is expected that the supplier solution will log the event and initiate an appropriate DevOps/Service Management process when the event is detected by the monitoring solution.
3. Processing of an acceptance acknowledgement may also be used to update the ‘last sent date’ associated with a patient’s extract send history. This timestamp is used to prioritise triggering of reconciliation extracts by supporting a least recently sent approach.

## Key Assumptions

|  |  |
| --- | --- |
| ID | Description |
| 1 | Patient Level Extracts may include patients not identified by NHS number |
| 2 | No additional patient matching or tracing is performed at source or within DPS. Any dissemination of re-identified location identifying data via DPS must consider the current PDS sensitivity indicator status of the patient. |
| 3 | There is no specific malware checking at source or destination, but strict structural validation of extract payload provides sufficient safeguarding of the receiving datastore. References to malware checking in this document refer to this type of validation rather than processing by a ‘malware scanning’ product. |
| 4 | No special measures are required to suppress the generation of patient level extracts when non-qualifying changes are made to patient records e.g. adding a purely narrative note to a patient record or an attachment will still result trigger an extract. This is because these activities will typically also be accompanied by the recording of other record content |
| 5 | Appropriate infrastructure capacity will be available within source systems to cope with the non-functional requirements arising from the ‘full extraction’ on change approach. |
| 6 | Appropriate network capacity will be available to transport extracts between source systems and DPS |
| 7 | Sizing and volumetrics for GP Data will be incrementally developed with suppliers during the implementation phase. |

## Operating Principles

### Snapshot Only Extraction

Patient Level Extracts operate on a full extraction on change principle (snapshot only).

This approach simplifies end to end data processing in comparison with an approach combining both full extract and incremental change-based data flows.

* There is a single DPS processing pipeline for all patient level extracts whether reconciliation or delta.
* Because each patient level extract is an accurate point in time copy of the source record in full, dealing with scenarios like patient movement within the data store is simplified within DPS and source systems.
* Scenarios like the need to perform ‘catch-up’ synchronisation in response to the removal of ‘type 1’ objections are avoided.
* There is a single extraction approach within source systems.

However, this simplification is at the expense of increased data extraction and transmission volumes from source systems which may require increased infrastructure capacity in comparison with an incremental change only based data flows.

Implementation of the GP data solution by suppliers will require careful consideration of the Non-Functional challenges arising from ‘snapshot only’. Non-Functional challenges extend from supplier capacity to shared concerns such as network capacity to NHSD concerns over MESH capacity.

Estimation of non-functional requirements should account for peak change volumes in addition to steady state volumes driven by patient facing workforce capacity e.g. bulk updates to patient records which may significantly skew change volumes where they occur.

### Extract Elimination

The sending system eliminates potential duplication between patient level extracts triggered by reconciliation extract activity and delta extraction i.e. where a record has changed and is included in the delta but would also have been included in the reconciliation extract.

### Data Equivalence with Source Systems

GP data extraction is primarily a data replication solution whose goal is to provide an accurate point in time (subject to latency) view of the data held about patients under a given registration in source systems. Reporting and analytics performed using the data store must yield the same or allowably similar results as equivalent reporting performed on the corresponding patient registrations on source systems.

While this registration-oriented reporting meets the needs of known use cases it is possible that future use cases could require the construction of a true longitudinal patient record and this is not precluded by the registration-based extraction approach.

### Patient Movement Handling

Patient movement between practices generates changes to patient registration which is a trigger for delta extraction like any other record change. This means that up to date patient information will be available to the data store when patients move between practices and the patient movement event will generate both an extract from the old practice as a result of deduction and an extract from the new practice as a result of registration. The effect of extraction on deduction is to update the status of patients at their old practice as inactive which provides a simple means of excluding deducted patients and their records at the previous practice from analysis (where appropriate)

### Inclusion by Registration Type and Status

The inclusion rules for each extract feed, differ.

#### Reconciliation Feed

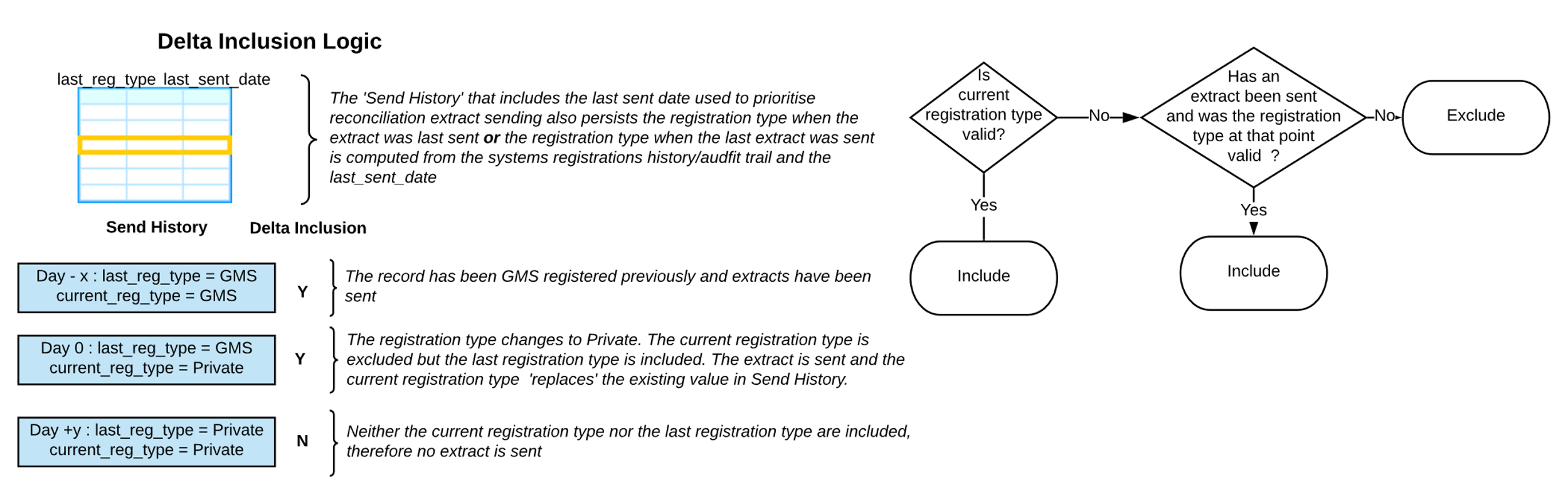
For the reconciliation feed, patient inclusion is by a specified set of registration types and status e.g. GMS and Temporary registration types with a status of active. This for example would exclude Privately registered patients by registration type and inactive or deceased patients by status.

#### Delta Feed

* For the delta feed, registration status does not apply. The intention is that in normal business process the records of inactive patients or deceased patients will not be routinely changing and if changes are being made then they are relevant to the patient record held in DPS.
* Registration type does matter and the delta feed inclusion criteria needs to handle not only the simple case of a registration type (new registration or change) coming into scope but also the more complex case of an existing qualifying registration type moving out of scope such as a temporary patient moving to a private registration. Where the registration type moves out of scope only one further extract should be sent while the registration type remains out of scope. In the temporary to private example, this means that the delta extract containing the change of registration type is sent once but further changes to the patient record afterwards are not sent which is the correct behaviour for a registration type which is out of scope for GP Data.

The following is one potential approach to implementing the required behaviour, other options may be available.

* Registration Type checking needs to consider not just the current registration type, but the registration type when the last extract was sent. This historic registration type may either be computed from registration history/audit trail or from information persisted specifically for GP Data like the extract ‘last sent date’.
* There are two components to the historic registration type check, firstly that a GP Data extract has been sent and secondly that the registration type current at the point the extract was last sent is one of the qualifying registration types. The implication of this is that during the notional ‘100 day’ running in period, there will be status changes missed that would otherwise be included, once the full synchronisation of patient data with DPS has occurred. This is expected behaviour as GP Data is not operational until the ‘100 days’ has elapsed.



### Type 1 Objection Handling

The solution respects type 1 patient objections at source.

**Patient Level Data**

The presence of a type 1 objection prevents patient level data flow.

Where a type 1 objection is removed the patient level data flows again.

**Appointment Data**

The presence of a type 1 objection prevents inclusion of the linking patient identifier (de-identified NHS number) with the appointment extract. If a type 1 objection is subsequently removed there is no need to restore patient linkage to appointments which have already been extracted.

Systems which cannot determine the type 1 status of the patient associated with the appointment should still populate the appointment extract with the patient linkage (type 1 objections by definition can only be observed by systems which record the type 1 objection against the patient).

### Appointment Extraction

Each execution of an appointments extract extracts all appointments changes since the last extraction at the appointment provider.

The first appointment extraction at a provider will be a snapshot of the current state of appointments within the system against which subsequent extracts provide incremental changes. The current state may include a historic component going back to a historic date cut-off.

The current state includes all currently pending appointment slots with no future cut-off.

This initial snapshot appointment extraction provides initial synchronisation of the data store with the appointment state held by the source system, without such a synchronisation it would not be possible to achieve synchronisation of all provider systems in a consistent period.

It should be possible for a supplier to ‘reset’ the state of the appointment extract as part of the scheduling process, such that it operates in snapshot mode on the next execution of the extract.

Creation of appointment slots from schedules/sessions is a change which will be reflected in the subsequent appointment extract. This means that following the initial snapshot and application of subsequent incremental extracts, the data store will have a full view of future appointment slots and their current state.

### Explicit Practice Participation

Although data collection will operate under a DPN, data controllers will still need to explicitly consent to extraction from their organisation. It is expected that this authorisation however it is captured will provide the technical switch that enables GP data to flow from an organisation. In the absence of this authorisation or the absence of such authorisation, the GP Data extract will not flow.

### Empty Extract Handling

It is expected that there could be instances where empty extracts are legitimately generated for a given provider and especially if the reconciliation feed is disabled e.g. at weekends where change may be minimal. Empty extracts transmission files should be suppressed at source either pro-actively (no records, therefore no extracts generated) or as a result of failed outbound validation.

### Data Minimisation

Longitudinal data minimisation is applied to patient level extracts. This takes the form of limiting the depth of historical record content selectively and via configurable parameters for specified content types.

Examples are restricting medication history or symptoms to 10 yrs.

In the case of medications, a hard date-based cut-off cannot be achieved because the scope of a medication authorisation may span the cut-off date with issues made after the cut-off date referring to the authorisation which lies beyond the cut-off date.

The delivery phase and forthcoming discussions with professional bodies will elaborate the categories of record content for which longitudinal minimisation is required and the definition/determination of each category.

It is expected that it will be possible to configure the cut offs for each content type and where terminological definitions of content type are applied, configure such definitions.

It is not expected that it would be possible to introduce a new type of restriction for a new content type by configuration alone.

Time restrictions apply to the clinically relevant date of a data item, not when it is recorded on the system.

Where the clinically relevant date is recorded as unknown, then longitudinal time-based minimisation cannot be applied.

Data minimisation applies to patient level extracts only and not to appointments.

### Configurability

It is expected that supplier implementations of the GP data extraction Subsystem will be configurable and hard coding of behaviours will be avoided. This does not mean that supplier solutions will be capable of ‘doing anything’ but instead that the key parameters that govern system behaviour will be modifiable without further software development.

Examples include but are not limited to.

* Scheduling and enabling of appointment and patient level extraction and transmission
* The extract transmission process including maximum transmission file size, MESH behaviours and addressing
* Whether type 1 suppression is enabled or not.
* The cut-offs associated with the data minimisation process
* Being able to disable the bulk synchronisation flow after the initial population of DPS is complete and subsequently re-enable it.
* The set of qualifying registration types and status for GP data extraction.

It will be possible to apply configuration changes across the supplier estate in an appropriate period.

### Extract Structure and Format

The extract format utilised in the GP data extract standard is based on an enhanced version of the GPES Interoperability Standard (GPES-I) [1]

1. The GPES-I standard provides a simple, concise record-oriented XML structure for bulk data exchange between GPCS and the destination datastore.
2. The XML structure facilitates content validation of extracts against the specified extract specification using standard XML validation tools.
3. The GPES-I format is modified to support an enhanced primary care data model which is closely aligned with the FHIR profiles defined for GP Connect and related domains.
4. FHIR vocabularies and datatypes are adopted where possible but extended where appropriate to support the needs of secondary uses including incorporation of existing vocabularies from NHSDD and GPES.
5. The mapping between the GP data extract format and FHIR is specified, allowing transformation between the two data representations.
6. The adoption of FHIR as the underlying resource model combined with a proprietary extract format provides for alignment around a strategic primary care data model and closes the traditional gap between direct care and secondary use data models.

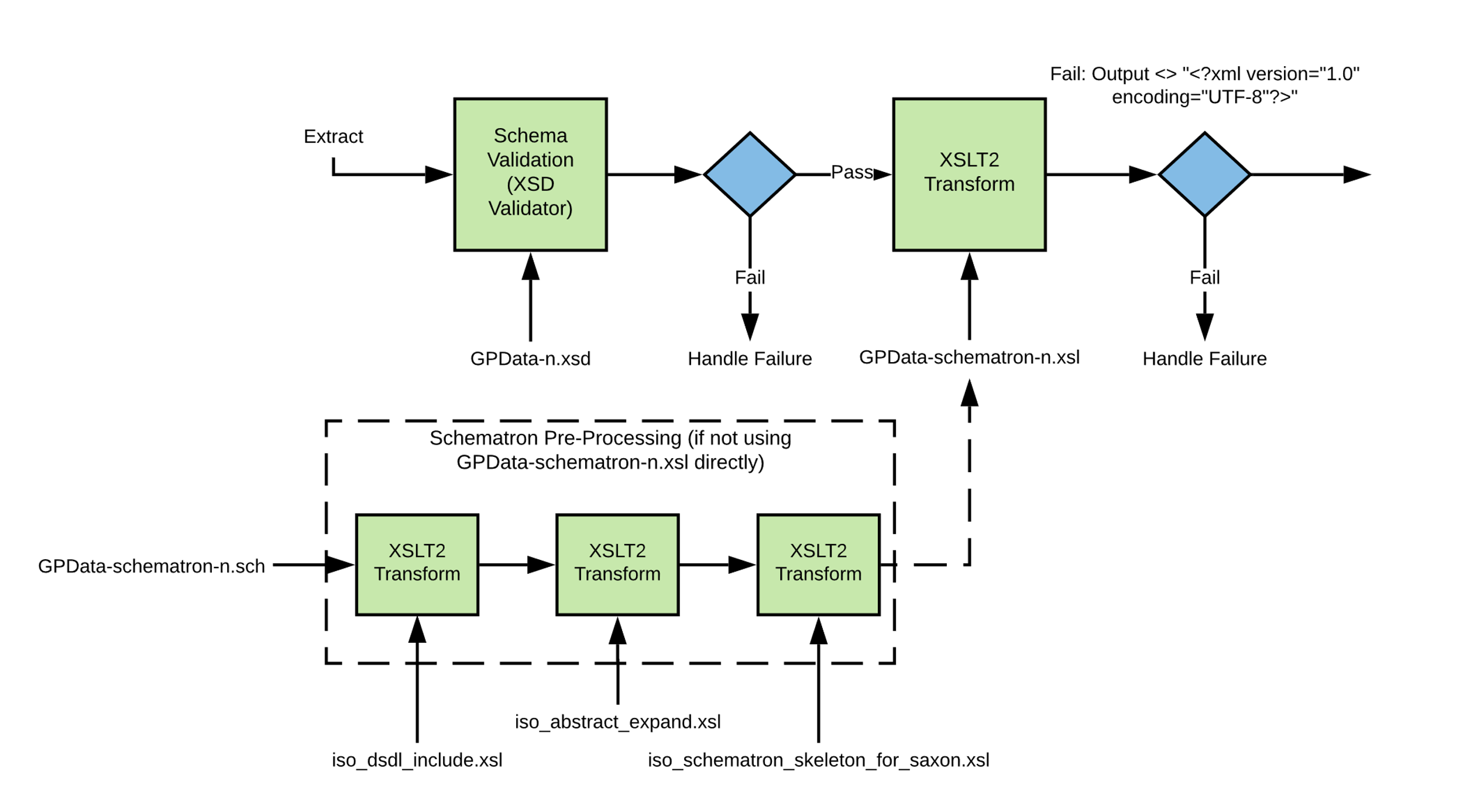
### Outbound Validation Pipeline

Outbound extract validation is a 2-stage process

1. XML Validation of extracts against the GP data Schema (GPData-n[[3]](#footnote-4).xsd)
2. Rules based validation implemented by transformation of the extract against an XSLT file generated from a set of GP data schematron rules (GPData-schematron-n.xsl generated from GPData-schematron-n.sch).

If the transformation stage generates anything other than an empty XML file (processing instruction only then it is considered a validation failure).

Although the XSLT file will be supplied with the GP data schematron products, suppliers may also generate the XSLT file from the GP data schematron file using the appropriate transforms.



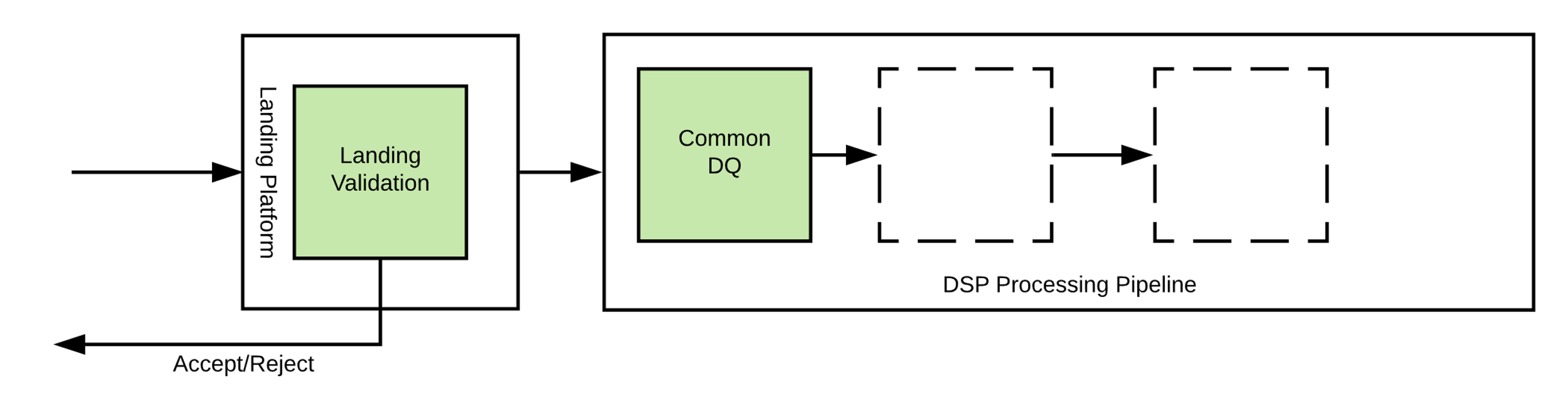
### Validation and Data Quality Approach

The GP data extract is principally a data replication solution in which an automatically generated representation of the primary care record is transmitted to the destination data store to support multiple use cases which would otherwise be satisfied by multiple extractions from the source systems.

* Assurance (and regression testing) of participating GPCS will ensure that generated extracts are an accurate representation of source records.
* There are no additional manual coding activities or business process activities associated with the submission of GP data extracts.
* There no specific business rules or validations that data submitted via GP data extracts should follow beyond being an accurate representation of the underlying patient record.
* All GP systems are systems of record, accredited to GPSoC and forthcoming GP IT Futures standards and meet appropriate standards for clinical safety, data quality, utilisation of terminology and correctness.
* All GP data extracts will be validated at source by the sending GPCS (schema validation and rules based validations) before transmission.
* Source practices and responsible GPs are not directly participating in a data quality business process associated with the extraction and transmission of GP data.
* System suppliers providing GP data solutions are participating in a technical business process and participate in resolving technical errors with the solution rather than business or DQ errors.
* As an ‘extract once utilise many’ extract, DQ checks and acceptance and rejection are applicable in the context of individual use cases but not in the context of acceptance of an extract into the datastore. To do otherwise would open the possibility of rejecting an extract on the basis of acceptability to a single use case when 99 others would accept the data.

As a result, there are no specific Data Quality (DQ) checks to be performed by the Landing Platform on receipt of GP data extract transmission files beyond schema and rules based structural validation and limited content checks. Examples are checking that org codes or clinical codes conform to expectations of content type and field length. Rejections of extract transmission files at the point of landing are technical rejections relevant to source system suppliers.

Once an extract transmission file has been accepted by the landing platform, subsequent processing within the DPS processing pipeline will apply the appropriate DQ validations for GP data use cases and make relevant warnings or exclusions available at the point of utilisation in the context of a given use case.



### Extract Acceptance Granularity

A consequence of the Data Quality approach is that there are no actionable technical or quality errors that can specifically affect the processing of patient level extracts within the overall extract transmission file. Acceptance or rejection of the extract transmission file by the Landing Platform operates at the level of the transmission file.

### GPCS Architecture

The GP data solution is not prescriptive about the architecture of source GPCS and supports practice based and enterprise hosted deployments. For enterprise scale deployments work may be partitioned between replicated ‘GP data Subsystems’ which have responsibility for handling a partitioned subset of the overall GP data workload. Work may be partitioned by a variety of mechanisms (e.g. by practice, patient id modulo n) provided individual patient extracts are transmitted serially and in the correct sequence through a given subsystem and channel.

### Channel Abstraction

The GP data Interface standard is neutral with regard to the architecture of participating GPCS and is applicable to practice-based solutions and enterprise scale GPCS solutions.

Scalability to enterprise scale is achieved by partitioning workload between multiple replicated and independent GP data extract sub-systems. Each sub-system establishes a sender defined channel with the Landing Platform.

Within each channel the sender guarantees that extract files will be sent in an ordered sequence and patient level extracts within the context of a given registration will always be sent through a single channel i.e. the ordered set of changes to a patient record must be sent in sequence and not through two channels concurrently.

There will also be only one instance of a patient record from a given registration at a provider system per extract cycle.

The precise mechanism by which workload is partitioned between extract subsystems and channels in a given system is implementation and architecture dependent and not defined by this interface specification. Partitioning may be based for example on practice/registration boundaries or partitioning of patient identifiers (modulus n approach).

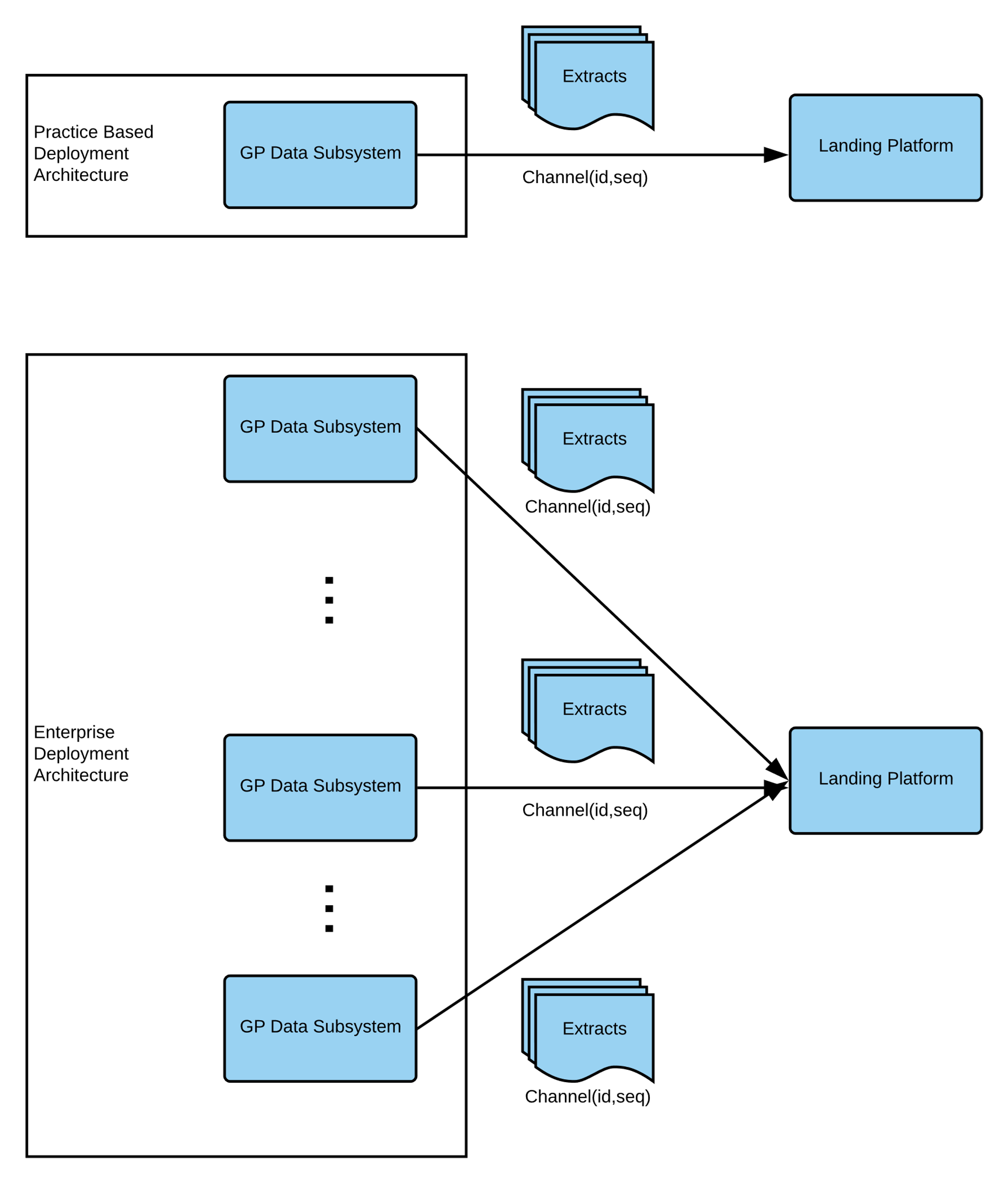
The Landing Platform validates that transmissions across a given channel occur in sequence as part of the verification stage of processing.

Out if sequence transmissions are not rejected outright but NHS Digital will monitor received transmissions and raised alerts for Service Management investigation in response to extracts being received out if sequence or with missing sequence numbers.

A channel is defined by a sending system defined unique identifier (Uuid) and sequence number and specified via extract file header. The channel acts as a proxy for the provider organisation and a separate channel identifier will be used for appointment and patient flows.

A channel may be reset by the sending system dynamically introducing a new channel identifier.

Channel sequences should commence at 1.



***Because there is an additional constraint of one transmission file per provider per feed, then the channel is a proxy for the provider/organisation supplying the extract data but is used in preference to the provider organisation code, because of the complexity of handling a transmission sequence tied to an organisation code, which can move between suppliers.***

### Transfer of Responsibility

Once an extract file has been accepted by the Landing Platform responsibility for onward processing and delivery of extract data passes from the sending GPCS to the Landing Platform. This transfer of responsibility allows the sending GPCS to mark the transmission on in-flight extracts as complete, freeing resources and allowing the transmitting system to move on the next batch of work.

### Asynchronous Transmission

Solutions should not wait for the business acknowledgement from the preceding transmission on the same channel to be received back before transmitting the next extract file.

### Exception Handling, Reliability and Disaster Recovery

#### Extract rejection and retries

Acceptance of submitted extract files is based on structural validation rather than DQ or other measures, consequently there should be no automatic re-send of extract files in response to rejection/failure acknowledgements as such rejections are likely to arise from system configuration or software issues that will not be resolved by re-sending. Instead a high severity issue should be logged and escalated through the supplier monitoring solution for onward management via operational and service processes.

#### Transmission Reliability and Retry Behaviour

In the case of solutions utilising the MESH transport mechanism and the MESH client, transmission retry behaviour is built into the solution therefore no additional retry behaviour is specified.

In the case of solutions using MESH API, the expectation is that retry behaviour equivalent to the MESH client would be available.

In the case of solutions, using the AWS transport, transfers occur within a highly reliable co-located cloud infrastructure with atomic write fratures. Consequently, no retry capability is mandated.

#### Disaster Recovery

In the case of catastrophic supplier solution platform or DPS Landing Platform unavailability, then data recovery may be performed by.

* Re-send of un-transmitted extracts by re-initiating the supplier solution to DPS transmission process. It is expected that DPS will be capable of either eliminating duplicate extracts received (identified by Extract Id) or when processed in the correct order duplicate extracts can be processed with no impact on data accuracy.

and/or

* Modifying the run-time parameters of supplier extraction processes such that extracts which were not successfully generated initially are regenerated including changes from a wider time period than the usual extraction process. This may mean for example generating patient extracts that include more than one day of reconciliation and delta extracts.

#### Supplier Monitoring and Escalation

It is expected that all solutions operate with appropriate logging and monitoring capabilities that can detect significant issues with the operation of the service and notify supplier DevOps and Service Teams for resolution and where appropriate these processes will involve NHS Digital Service Management.

### Authentication and Authorisation

#### MESH Transport

GPCS and Landing Platform attempting to send GP data messages via MESH must authenticate using the correct MESH credentials and utilise the correct client certificate. There is no additional authentication or authorisation beyond these MESH features.

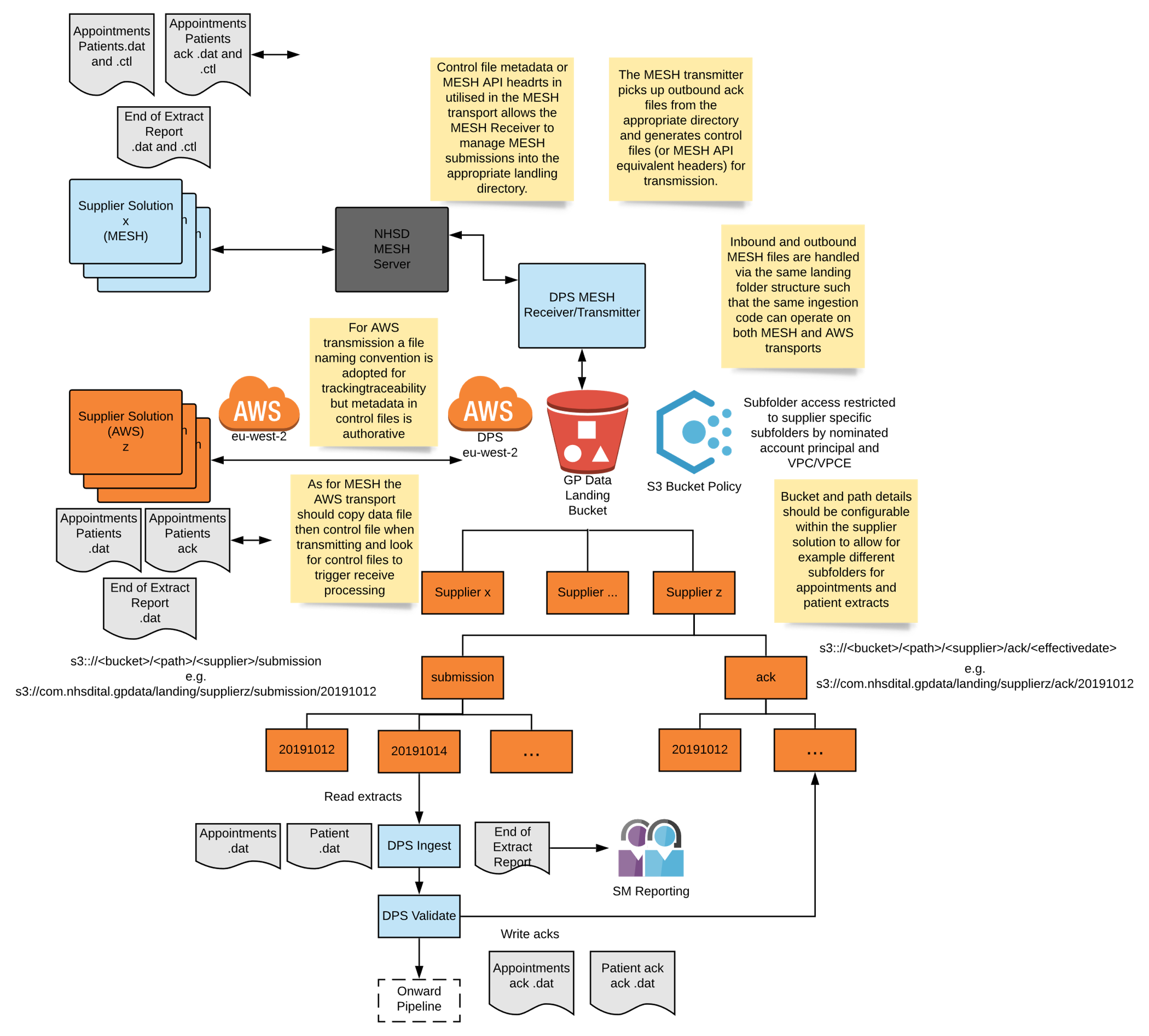
#### AWS Transport

NHSD managed AWS policies control access to the supplier specific bucket subfolders used to write extract transmission files and read acknowledgements. Access to a given supplier subfolder is limited to a supplier’s nominated principle and S3 VPC Endpoint. The VPC Endpoint constraint ensures that GP Data extracts and acknowledgements are transmitted entirely within AWS with no egress across the internet.

### De-Identification

De-Identification is performed at source by sending systems using the NHS Digital supplied De-Identification toolkit. The **De-Identification Toolkit Implementation Guidance** provides information on the deployment, operation and invocation of the De-Identification toolkit to perform the set of De-Identifications required for GP data.

# Extract Transmission



## Principles

Extract transmission files transmitted top DPS and Acknowledgement files received from DPS are transmitted over MESH or AWS depending on the option selected for the supplier solution.

* No control files are utilised for the AWS interface.
* In the case of MESH transmissions, control file metadata (Subject field) and Workflow Id provide the metadata that allows both MESH and AWS submissions to be managed into the same landing folder structure from which a common ingestion process is triggered.
* In the case of S3 transmissions, file naming and appropriate placement in the landing folder structure provides sufficient information for onward file processing.
* Files transferred using MESH and AWS are managed into and out of DPS via landing bucket and subfolder structure. For MESH transport the landing structure is opaque to suppliers as the DPS MESH component manages files into and out of the landing structure. Similarly access control to the landing bucket for solutions using the MESH transport is managed via the DPS MESH component.
* In the case of the MESH transport it is expected that the DPS MESH component will be able to determine the sending supplier from the sending mailbox identifier.
* Files are segregated by supplier, workflow (extract or acknowledgement) and by date.
* Files are written to and received from dated sub-folders that correspond to the 24-hour period a given extract or related ack is for (the **effective date)** i.e. does not correspond to when an extract is processed or an ack is generated.
* There is currently no explicit segregation in the landing folder structure by workflow (patient or appointment) but solutions using the AWS transport option should cater for flexibility in this area by making bucket and subfolder paths configurable per extract type.
* Access to the AWS landing structure for solutions using the AWS transport is limited by policy such that suppliers can only write and read from the supplier specific bucket sub-folder and suppliers are identified within the policy by principal and S3 VPC endpoint (the use of a VPC Endpoint ensures that all traffic to the S3 landing bucket routes within AWS rather than externally.
* All data landing in DPS is encrypted at rest, hence data copied by AWS transport is encrypted transparently on landing. For this to operate correctly the supplier principal needs appropriate AWS KMS permissions to perform operation on encryption keys and on the NHS Digital side appropriate KMS permissions to access the encryption key will be granted to the supplier principal.

## Message Sizing

The maximum allowed message size (uncompressed) will be a configurable value and determined during the GP data delivery phase.

Where the uncompressed limit is reached the sending system should ensure that only whole patient level extracts are included in any given extract file i.e. a given patient level extract (snapshot or delta) should not be split between extract transmission files. Where a given patient level extract cannot be included in an Extract transmission file it will be included in a subsequent extract transmission file.

## Compression

All extract transmission files are compressed (gzip) by the sending system and uncompressed by the receiving landing platform for processing. Acknowledgement messages are not compressed.

## Networking

MESH transmission may over the public internet (MESH is an encrypted channel) in line with internet first principles or via HSCN.

In the case of the AWS transport all file transmission is over encrypted channels within AWS with no egress to the internet.

## MESH Addressing

There will be at least one MESH endpoint address (MESH mailbox) for each sending and receiving party.

This specification is not prescriptive about the precise MESH topology and configuration employed.

1. MESH mailboxes utilised for GP data may be shared with other message flows provided the application can distinguish GP data traffic from other messages by workflow id and route them appropriately and the ‘shared’ nature of the mailbox is understood from an architectural, operations and service management perspective.
2. Multi-mailbox configurations in which the same MESH client services multiple mailbox addresses.
3. Each instance of a GPCS data sending subsystem will be associated with a single destination Landing Platform MESH mailbox as the destination for extracts.
4. The Landing Platform may utilise more than one MESH mailbox and client
5. There may be a one to one association between GPCS GP data extraction subsystems and mailboxes and clients or a one to many association, provide the application the sending subsystem can adequately discriminate between Ack messages received.

## Workflow Ids

A distinct MESH workflow is assigned to each interaction (also applicable to file naming in AWS transport)

### Extract Work Flows

|  |  |
| --- | --- |
| Interaction | MESH Workflow Id |
| GP data Patient Level Extract (GPCS -> Landing Platform) | GPDRP\_PAT\_EXTRACT |
| GP data Patient Level Ack (Landing Platform -> GPCS) | GPDRP\_PAT\_EXTRACT\_ACK |
| GP data Appointment Extract (GPCS -> Landing Platform) | GPDRP\_APPT\_EXTRACT |
| GP data Appointment Ack (Landing Platform -> GPCS) | GPDRP\_APPT\_EXTRACT\_ACK |

### Extract Report Work Flows

|  |  |
| --- | --- |
| Interaction | MESH Workflow Id |
| GP data Patient Level Extract Report (GPCS -> Landing Platform) | GPDRP\_PAT\_REPORT |
| GP data Appointment Extract Report (GPCS -> Landing Platform) | GPDRP\_APPT\_REPORT |

### Control File[[4]](#footnote-5)

|  |  |  |  |
| --- | --- | --- | --- |
| **DATA ITEM** | **VALUES** | **DESCRIPTION** | **MESH Applicable** |
| Version | 1.0 | Version of control file | Y |
| AddressType | DTS | All transmission is via DTS type messaging | Y |
| MessageType | Data | Identifies the type of transfer   * Data will have a data file and a control file. | Y |
| WorkflowId | Allocated Workflow ID see above | Use the allocated workflow Ids for each interaction | Y |
| From\_DTS | MESH mailbox Name | MESH mailbox of sender | Y |
| To\_DTS | MESH mailbox Name | Destination MESH mailbox | Y |
| Subject | SMTP Subject | See Correlations | Y |
| LocalId | Application Specific Value | Local identifier of the data transfer. This is specific to the host application sending via the MESH client. This will allow for correlation with DTSId. | Y |
| Compress | N | Always N  GP Data Extract files are always compressed by the application GP Data Ack files are never compressed Extract Report files are never compressed | Y |
| Encrypted | N | MESH encryption will be utilised therefore set Encrypted = N | Y |
| AllowChunking | TBD | Whether chunking is enabled will be determined during the delivery phase. | TBD |

### File Naming

In the case of the MESH transport, file naming conventions are optional have local significance only and are only useful to assist supplier side operations.

In the case of the AWS transport, following the file naming convention is mandatory

* Extract Naming

**<Workflow Id>**\_**<Message Id>**\_**<Effective Date Time>**\_**<Channel Id>\_<Sequence>**.**dat**

e.g. GPDRP\_PAT\_EXTRACT\_ ec403dac-4f3d-4278-b3a5-a9ada512a828\_20190511000000\_ 866b7fef-ed9e-476e-9192-aa23e06183d6\_12.dat

Compressed files are given the .gz suffix

GPDRP\_PAT\_EXTRACT\_ ec403dac-4f3d-4278-b3a5-a9ada512a828\_20190511000000\_ 866b7fef-ed9e-476e-9192-aa23e06183d6\_12.dat.gz

* Acknowledgement Naming

**<Workflow Id>\_<Acknowledgement Id>\_<Response to Extract Message Id>.dat**

e.g.GPDRP\_PAT\_EXTRACT\_ACK\_ 9bf5b105-bbc3-4036-962f-9e15e6f37a7c\_ec403dac-4f3d-4278-b3a5-a9ada512a828.dat

* Extract report file naming

**<Workflow Id>\_<Effective Date Time>.dat**

GPDRP\_PAT\_REPORT\_20190511000000.dat

### Population of Subject Element

In the case of extract files, the **Effective Date Time** component of Subject allows extract files landing over MESH to be placed in the correct date subfolder of the landing bucket.

**Extract File**

GPDATA-Extract:[Effective DateTime]:[Message Id]:[Channel Id]:[Channel Sequence Number]

e.g. GPDATA-Extract:20190511000000:ec403dac-4f3d-4278-b3a5-a9ada512a828\_20190511231021:866b7fef-ed9e-476e-9192-aa23e06183d6:12

**Acknowledgement File**

GPDATA-Ack:[Message Id]:[Response To]

e.g. GPDATA-Ack:9bf5b105-bbc3-4036-962f-9e15e6f37a7c:ec403dac-4f3d-4278-b3a5-a9ada512a828

**Report File**

GPDATA-Report: [Effective DateTime]

e.g. GPDATA-Report:9bf5b105-bbc3-4036-962f-9e15e6f37a7c:ec403dac-4f3d-4278-b3a5-a9ada512a828

### Correlations

The MESH control file subject field is overloaded to provide an end correlation/traceability between application and transport layers and communicate out of band information such as the extract checksum.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item | Extract Report | Filename[[5]](#footnote-6) | Control File Subject Element | Extract or Acknowledgement File Contents | Notes |
| Message Id | Y | Y | Y | gpdata-extract/@ID GPData-Ack/@ID | Provides correlation between MESH control file and extract/ack identifiers |
| Channel Id | Y | Y | Y | gpdata-extract/@channel-id | Provides correlation of channel id and sequence across control and extract transmission files |
| Channel Sequence | Y | Y | Y | gpdata-extract/@channel-sequence | Provides correlation of channel id and sequence across control and extract transmission files |
| Effective Date Time | Y | Y | Y | n/a | Provides the nightly cut off date time that the extract applies to. |
| Sending MESH Mailbox | N | N | N | gpdata-extract/@source-message-address\*  *This is a mandatory field in the extract but not applicable to AWS transport and should be populated with dummy values.* | Provides source MESH mailbox to send ack to from within extract transmission file. |
| Response-To | N | Y | Y | GPDATA-Ack/@Response-To | Allows an ack to be linked to an extract transmission file without opening the Ack message file. |

### MESH Client Configuration

Unless specified by further guidance client configuration values should be set according to the MESH specification [3,4]

# Message Specification

## GP Data Extract

The GP Data Extract specification is provided by the **GP Data Extract Technical Output Specification** and the **GP Data Extract Implementation Guidance**

## GP Data Ack Message

The GP Data Ack message is used to provide positive and negative acknowledgement to the sending GPCS.

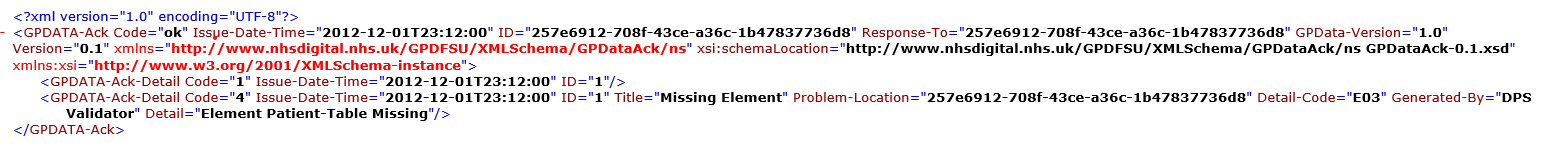
A screenshot of a cell phone

Description generated with very high confidence

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Message** | GP Data Ack | | | |
| **Description** | Provides the acknowledgement message between Landing Platform and sending GPCS | | | |
| **Element** | GPDATA-Ack | | | |
| **Attribute** | **Description** | **Type[[6]](#footnote-7)** | **M/O/R** | **Notes** |
| ID | Unique identifier for Acknowledgement message | Id | M |  |
| Version | Version of GPDATA-Ack message | String | M |  |
| GPData-Version | Compliant with GPData-Version | String | M |  |
| Issue-Date-Time | Timestamp of when the GP Data Ack was generated | Datetime | M |  |
| Response-To | The GP Data Extract being Acknowledged | Id | M |  |
| Code | The overall response code. | Code | M | <http://hl7.org/fhir/stu3/valueset-response-code.html> ok – the receiver has accepted the extract for onward transmission fatal-error – there is a non-recoverable problem with either the extract or the receiver. There is no point in retrying. |
| **Content** | **Description** | **Type** | **M/O/R** |  |
| Detail | Detailed description of error condition | GPDATA-Ack-Detail | M | 0..n Will not generally present for a success/acceptance acknowledgement but detail may be supplied for informational purposes |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Message** | GP Data Ack Detail | | | |
| **Description** | Error Detail and description | | | |
| **Element** | GPDATA-Ack-Detail | | | |
| **Attribute** | **Description** | **Type[[7]](#footnote-8)** | **M/O/R** | **Notes** |
| ID | Ascending numeric identifier of error detail within context of the specific overall Acknowledgement | Id | M | Start with an index of 1, where multiple errors are reported the primary error should be reported first |
| Issue-Date-Time | Timestamp of when the error detail was generated | Datetime | M |  |
| Code | Specific GPData assigned Error/Information Code | Code | M | |  |  |  |  | | --- | --- | --- | --- | | Code | Display | Detail | Overall ack code | | 1 | Sender Invalid | The receiving landing platform has been unable to validate the sender as a valid source of GP data extracts | fatal-error | | 2 | Malware detected | Malware was detected by the receiver | fatal-error | | 3 | Checksum Failed | The supplied checksum was invalid | fatal-error | | 4 | Invalid Extract | The extract has failed validation | fatal-error | | 5 | Other error | Other error type – details in the detail attributes | fatal-error | | 6 | Sequence Error | Out of sequence extract on channel | fatal-error | | 7 | Informational | Returned detail is informational | ok | |
| Detail-Code | Detailed local error code | String | O |  |
| Title | Title of detailed error | String | O |  |
| Detail | Detailed error description | String | O |  |
| Problem-Location | If a problem is detected with specific extract content describes the location of the issue | String | O |  |
| Generated-By | The source of the detailed error information e.g. a validator or other component | String | O |  |

### Example GPData Ack Message



## Extract Report File

The purpose of the extract report file is two-fold.

* Provides a timely signal to NHS Digital that the supplier has completed file generation and submission for the extract day.
* Provides content that satisfies the majority of the reporting requirements set out in the **GPDfSU GP Dataset Supplier Requirements.**

The need for timely delivery of the report file following completion of extraction activity prevents inclusion of information correlating the submission with Acknowledgements returned by NHS Digital.

### Timeliness

The daily extract reports for the appointments and patient feeds should be generated and sent when the extract generation and submission process is completed from a supplier perspective.

* This is not dependent on receipt of acknowledgement messages from NHS Digital for every extract submitted
* In the case of AWS submissions transmission is considered complete when the last file for each feed is written to the landing bucket.
* In the case of MESH submissions, submission is considered complete when the last file for each feed has been uploaded to the MESH gateway server (is not dependent on download by NHS Digital).

### File Content

The file contains a JSON object populated as follows.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Field | **Optionality** | **Type** | | | | | | | | | **Description** |
| serviced\_provider\_list | M | Array of Provider ODS Organisation Codes (0..n) | | | | | | | | | On any given day, provides the set of organisations services by the supplier who are potential GP Data providers |
| participation\_list | M | Array of Provider ODS Organisation Codes (0..n) | | | | | | | | | On any given day, provides the set of organisations services by the supplier who are participating in GP Data extraction |
| expected\_submissions | M | Array of Expected Submissions (0..n) | | | | | | | | | The expected submissions for a given day. This is the prospective extract activity based on reconciliation or delta activity or in the case of appointments snapshot or delta. The structure differs between appointment and patient feeds because of the need to distinguish between reconciliation and delta extraction in the case of the patient extract. |
| **Patient Extract Expected Submission** | | | | | | | | |
| **Field** | | **Type** | | | **Optionality** | | | **Description** |
| org\_code | | String | | | M | | | The organisation code of the provider generating the submission |
| delta | | Boolean | | | M | | | Whether the provider will generate a patient delta extract |
| recon | | Boolean | | | M | | | Whether the provider will generate a patient reconciliation extract |
| **Appointment Extract Expected Submission** | | | | | | | | |
| **Field** | **Type** | | | **Optionality** | | | **Description** | |
| org\_code | String | | | M | | | The organisation code of the provider generating the submission | |
| snapshot | Boolean | | | M | | | Whether the provider will generate an appointment snapshot extract | |
| delta | Boolean | | | M | | | Whether the provider will generate an appointment delta extract | |
| actual\_submissions | M | Array of Actual Submissions (0..n) | | | | | | | | | A record of each extract transmission file, generated and successfully submitted rom the supplier perspective and without correlation with Acknowledgement messages received in response. |
| **Patient Extract Actual Submission** | | | | | | | | |
| **Field** | | | **Type** | | **Optionality** | **Description** | | |
| org\_code | | | String | | M | The organisation code of the provider generating the submission | | |
| date\_generated | | | FHIR Datetime | | M | The date and time of when the extract file was generated | | |
| date\_transmitted | | | FHIR Datetime | | M | When the transmission of the extract to NHS Digital commenced. For MESH transport may be approximate in the case of MESH Client. | | |
| message\_id | | | Uuid | | M | The Extract Id | | |
| channel\_id | | | Uuid | | M | The Channel id | | |
| channel\_sequence | | | Positive Integer | | M | The Channel sequence number | | |
| effective\_date | | | FHIR Datetime | | M | The effective date for the extract i.e. the midnight cut off for extract inclusion | | |
| record\_count | | | Positive Integer | | M | The number of patient records in the extract file | | |
| type | | | String | | M | The type of patient extract transmission file. Differentiates between delta, reconciliation and combined extracts. “combined” is used where reconciliation and delta extracts are not segregated and are combined into the same extract transmission file. | | |
| delta recon combined | |
| **Appointment Extract Actual Submission** | | | | | | | | |
| **Field** | | | **Type** | | **Optionality** | **Description** | | |
| org\_code | | | String | | M | The organisation code of the provider generating the submission | | |
| date\_generated | | | FHIR Datetime | | M | The date and time of when the extract file was generated | | |
| date\_transmitted | | | FHIR Datetime | | M | When the transmission of the extract to NHS Digital commenced. For MESH transport may be approximate in the case of MESH Client. | | |
| message\_id | | | Uuid | | M | The Extract Id | | |
| channel\_id | | | Uuid | | M | The Channel id | | |
| channel\_sequence | | | Positive Integer | | M | The Channel sequence number | | |
| effective\_date | | | FHIR Datetime | | M | The effective date for the extract i.e. the midnight cut off for extract inclusion | | |
| record\_count | | | Positive Integer | | M | The number of appointment records in the extract file | | |
| type | | | String | | M | The type of appointment extract – snapshot or delta | | |  |
| snapshot delta | |
| status | M | FHIR ResponseType | | | | | | | | | Whether from the supplier perspective there was an error detected that prevented transmission of the complete set of extracts. This is indicative only and may only be populated for recognised error conditions on the supplier side |
| ok  fatal-error | | | | | | | | |

### Example

#### Patient Extract Report File

{

“status”: “ok”,

“serviced\_provider\_list”: [“A81001”, “X90802”,“D90002” …. “E67271” ],

“participation\_list”: [“A81001”, “D90002” …. “E67271” ],

“expected\_submissions”: [ {

“org\_code”: “A81001”,

“delta”: true,

“recon”: true

}

…

{

“org\_code”: “D90005”,

“delta”: true,

“recon”: true

}

],

“actual\_submissions”: [

{

“org\_code”: “A81001”,

“date\_generated”: “2019-10-01T02:15:00”,

“date\_transmitted”: “2019-10-01T02:20:00”,

“message\_id”: “674f628f-a78b-4a45-9ed0-07dcfefa8d3f”,

“channel\_id”: “b4831188-da94-467f-9a0e-01cbf8728eb5”,

“channel\_sequence”: 2667,

“effective\_date”: “2019-09-31T00:00:00”,

“record\_count”: 1200,

“type”: “delta”

},

{

“org\_code”: “A81001”,

“date\_generated”: “2019-10-01T02:15:01”,

“date\_transmitted”: “2019-10-01T02:20:01”,

“message\_id”: “674f628f-a78b-4a45-9ed0-07dcfefa8d2e”,

“channel\_id”: “b4831188-da94-467f-9a0e-01cbf8728eb5”,

“channel\_sequence”: 2668,

“effective\_date”: “2019-09-31T00:00:00”,

“record\_count”: 800,

“type”: “snapshot”

}

…..

]

}

#### Appointment Extract Report File

{

“status”: “ok”,

“serviced\_provider\_list”: [“A81001”, “X90802”,“D90002” …. “E67271” ],

“participation\_list”: [“A81001”, “D90002” …. “E67271” ],

“expected\_submissions”: [ {

“org\_code”: “A81001”,

“delta”: true,

“snapshot”: false

}

…

{

“org\_code”: “D90005”,

“delta”: true,

“snapshot”: true

}

],

“actual\_submissions”: [

{

“org\_code”: “A81001”,

“date\_generated”: “2019-10-01T02:15:00”,

“date\_transmitted”: “2019-10-01T02:20:00”,

“message\_id”: “674f628f-a78b-4a45-9ed0-07dcfefa8d3f”,

“channel\_id”: “b4831188-da94-467f-9a0e-01cbf8728eb5”,

“channel\_sequence”: 2667,

“effective\_date”: “2019-09-31T00:00:00”,

“record\_count”: 1200,

“type”: “delta”

},

{

“org\_code”: “A81001”,

“date\_generated”: “2019-10-01T02:15:01”,

“date\_transmitted”: “2019-10-01T02:20:01”,

“message\_id”: “674f628f-a78b-4a45-9ed0-07dcfefa8d2e”,

“channel\_id”: “b4831188-da94-467f-9a0e-01cbf8728eb5”,

“channel\_sequence”: 2668,

“effective\_date”: “2019-09-31T00:00:00”,

“record\_count”: 800,

“type”: “snapshot”

}

…..

]

}

1. The term de-identification is used in this document reflecting the technical naming of the feature within the de-identification product itself and surrounding documentation. However the GDPR term is pseudonymisation because pseudonyms are still considered identifiable data. [↑](#footnote-ref-2)
2. The actual point of origin for the initial appointment snapshot will be configurable [↑](#footnote-ref-3)
3. Where n is a version number or versioning text [↑](#footnote-ref-4)
4. Or MESH API Equivalents [↑](#footnote-ref-5)
5. File naming is most relevant to the AWS transport but may be implemented by a supplier using the MESH transport for local use [↑](#footnote-ref-6)
6. Datatypes conform to the FHIR Datatypes specification https://www.hl7.org/fhir/datatypes.html [↑](#footnote-ref-7)
7. Datatypes conform to the FHIR Datatypes specification <https://www.hl7.org/fhir/datatypes.html> [↑](#footnote-ref-8)