



INTERWEAVE
CONNECTING CARE

Cookbook for Regional Interoperability
Detailed Design Paper #001

A Unified Distributed Data Model for FHIR

INTERNAL DISCUSSION DRAFT

Version 1.1 – 6th January 2019

Abstract Interoperability Cookbook Anchor Points

Section	Title
3.1.2	Regional and Model Component Specifications – Regional Infrastructure - FHIR Service Bus

Table of Contents

1	Introduction	4
1.1	Purpose of this Document	4
1.2	Intended Users of the This Document	5
1.3	A Very Brief Primer on FHIR.....	5
2	FHIR Data Models	6
2.1	A Contained Model	6
2.2	A Distributed Model	6
2.3	A Unified Distributed Model.....	8
2.3.1	An Illustrative Search Against a Unified FHIR Endpoint	9
2.3.2	Basic Search Execution Algorithm	10
2.3.3	Advanced Search Directives and Post Processing Requirements.....	11
2.3.4	Resource References in Search Terms.....	13
2.3.5	Creating and Modifying Resources Through a Unified Endpoint	13
2.3.6	Subscriptions.....	13
2.3.7	A Hybrid Model for Resource Identification	14
3	Master Data Management in a Unified Model.....	16

1 Introduction

1.1 Purpose of this Document

This document is one of a series of design papers which underpin the Abstract Cookbook for Regional Interoperability. These papers, in their totality, describe the technical components and the standards which form the YHCR system of systems. They are intended as a basis for developing or procuring software and so are expressed at a level of precision which is intended to avoid ambiguity but with a consequence that they are focussed to technical readers.

Design papers are anchored to topics which are discussed in the Abstract Cookbook for Regional Interoperability. They are elaborations of the concepts which were first introduced by the abstract and new content is further detail rather than variations of previously established core principles.

This document (design paper 001 - "A Unified Distributed Data Model for FHIR") explores the challenges of, and offers resolutions for, presenting data which are fundamentally distributed, data which are captured and located in care settings across the region, as a referentially integral, internally consistent, database which can be accessed and queried by data consumers from a single regional end point.

The system of systems architecture embraces federation of data but does so under two provisions:

1. Structurally and Semantically Normalised at Source

Data exposed by a locality is presented in format that strongly adheres to FHIR STU 3 standards, is designed in accordance with regionally curated FHIR profiles and, ultimately, is coded to regional standards.

2. Accessible as a Single Unified Data Model from a Regional Endpoint

Regional infrastructure mediates between data consumers and data providers and doing so simplifies access to data, applies standard governance constraints, ensures consistent treatment and representation of data.

The first of these provisions should be enforced by the data profiling work of a YHCR Data Architecture Design Authority. This body, at the time of writing, has yet to be formed.

This paper, in part, defines the second topic. It covers from the perspective of a regional FHIR Bus:

- Identification of FHIR resources.
- Cross-referencing of resources.
- Interpretation of search terms and execution of distributed searches.
- Master data management.

Subjects which are relevant to the design of a regional FHIR Bus which are covered in other design papers include:

- Technical implementation of FHIR endpoints (support for FHIR APIs, FHIR over non-HTTP protocols, asynchronous messaging).
- Application of consent, regional role-based access, and enforcement of legitimate relationships.
- Auditing.

- Security.
- Non-functional requirements (scalability, high availability).

1.2 Intended Users of the This Document

This design paper will inform the development of the FHIR aggregator functionality of the regional FHIR Bus (Abstract Cookbook section 3.1.4). The content is also relevant for data consumers looking to interact with the system of systems via the FHIR Bus (Abstract Cookbook section 5.4).

1.3 A Very Brief Primer on FHIR

Comprehension of this document will be facilitated by a working knowledge of the HL7 FHIR standard <http://hl7.org/fhir>.

However, in the interests of self-containment the following brief primer covers the core concepts relied on in the following chapters.

FHIR defines a data model that provides good coverage of an electronic patient record.

Data entities are known as resources. A resource fully describes a well encapsulated business object and covers master data (people, objects, locations, and entities) and transactional concepts (diagnoses, observations, care plans, test orders etc).

Every resource has a URL from which the resource can be obtained. The URL may be absolute (<https://endpoint.nhs.uk/Patient/1234>) or relative to endpoint which served the resource: (/Patient/1234). The URL embeds a unique resource id.

A resource may have one or more business identifiers – a patient has an NHS number.

Resources refer to other resources using either their URL or business identifier. For instance, an Observation may have a subject being a Patient. The observation resource references the patient resource.

A resource fully describes the concept which it models. This leads to redundancy in properties maintained for referenced resources. For instance, an Encounter is for a Patient (and therefore references the patient. An Observation is made for a Patient as part of an Encounter (and therefore references both the patient and the encounter). Intuitively, both the Encounter and Observation should be for the same Patient but FHIR doesn't enforce this and doesn't offer any interpretation of possible logical inconsistencies.

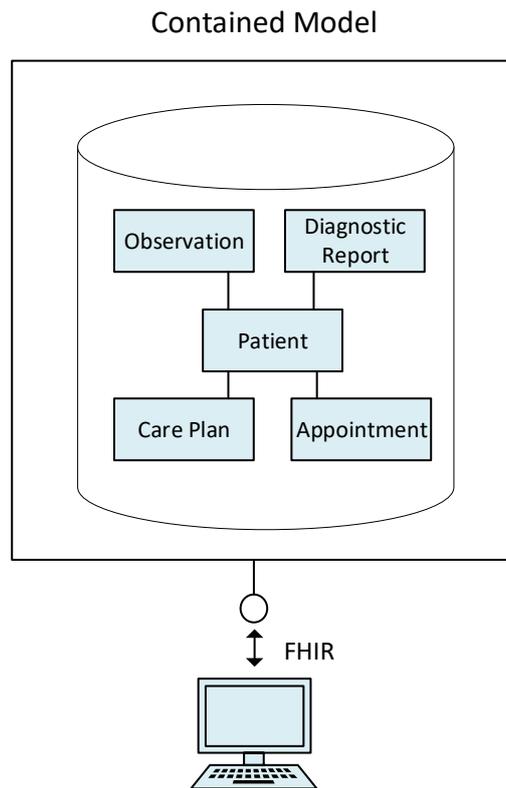
More than one resource may pertain to the same business object. If this duplication is known to a data owner, then the resources can be associated with each other using a Linkage resource.

2 FHIR Data Models

One of the attractions of FHIR is that it lends itself equally for modelling data for a single application as for representing a full patient record distributed across care settings.

2.1 A Contained Model

A patient record within a single care setting or managed by a single application might be represented as a contained FHIR model. All things known about a patient; test results, diagnoses, care plans, appointments etc. are represented as FHIR resources which are surfaced from a single endpoint.



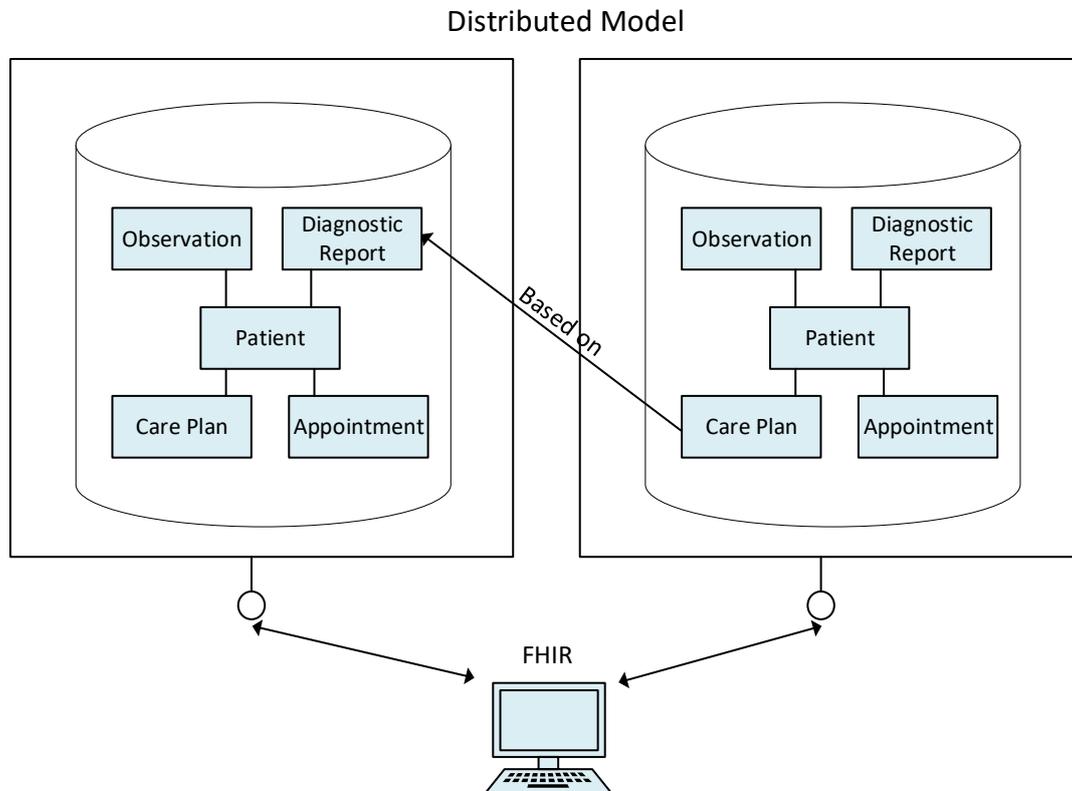
Resources are uniquely identified using a numbering scheme which is within the control of the application or care setting. The data model is complete, i.e. all resource references are to resources which exist within the application or care setting and are referenced using the contained model's numbering scheme.

2.2 A Distributed Model

When more than one data source contributes to the patient record then the FHIR model is distributed. A consumer of data must interact with more than one endpoint if it is to assemble a complete record for a patient.

Data sources in the distributed model may be contained or there may be interdependencies. A resource owned by one data source may reference a resource owned by another source. For example, a diagnostic report may be written as a consequence of a care plan. The care plan may

have been developed a primary care physician and the corresponding FHIR *CarePlan* resource is owned by the GP system whereas the diagnosis was performed by a psychiatrist and the *DiagnosticReport* is owned by a mental health care setting.



Clearly cross-boundary resource referencing can be avoided if resources are copied and held locally. One reason for copying resources is to place control of the evolution of the content of a resource in the hands of the organisation which has become dependent on it. The diagnostic report was prompted by a care plan as it existed at a point in time and the diagnosis may, to some extent, be dependent on the content of the care plan as it existed when the diagnosis was performed. Taking a local copy of the *CarePlan* resource and making the local copy the provenance for the *DiagnosticReport* is one mechanism for ensuring that the audit trail of dependency is preserved. Of course, the FHIR referencing standard allows one resource to reference a specific version of another resource.

As an example the following URL references the 5th version of a care plan:

https://GPSystem.nhs.uk/fhir/CarePlan/53479922/_history/5

Resource versioning, if implemented and trusted, goes some way to reducing the need to duplicate data. The YHCR will be implementing, at certain levels of the maturity model, resource versioning.

A distributed model introduces complexities for a data consumer:

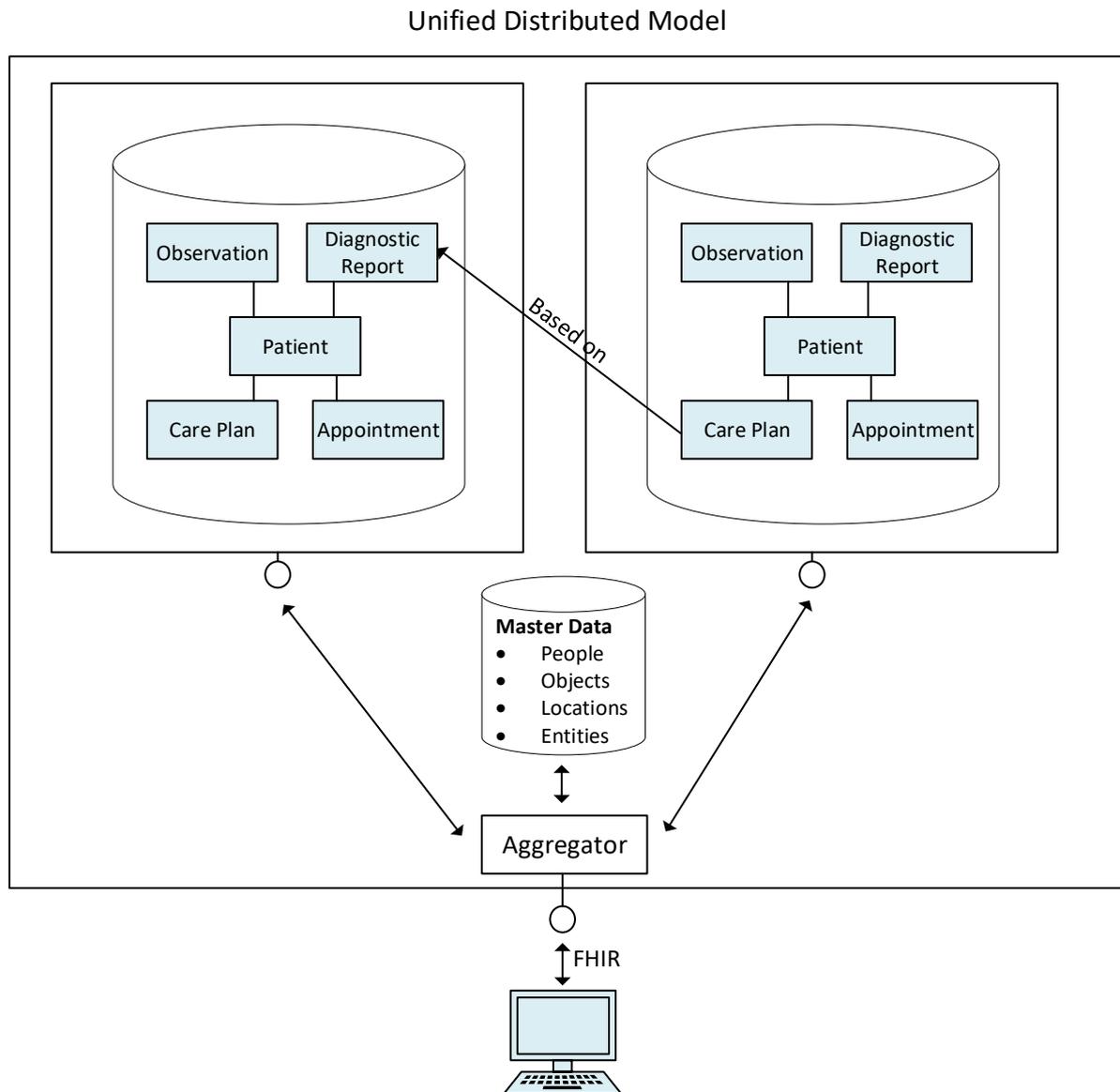
- The same business object may be represented by a number of resources, each owned by different data sources. Resources may disagree on the properties of the business object.
- Transactions may be duplicated across data sources.

- Searches for resources must be repeated against multiple data sources and the results aggregated.
- Processing required for advanced manipulation of search results (ordering, pagination, etc.) moves from the data provider to the data consumer.

It also adds a complication for data providers: chained search parameters of arbitrary length which traverse resource references across the boundaries of data sources are computationally difficult to implement for real-time searching. For this reason, support for chained search parameters in the YHCR will be targeted to a predefined set of search terms.

2.3 A Unified Distributed Model

The unified model aims to abstract the complexity of interacting with a distributed model from data consumers. Unification wraps the distributed data model within a single FHIR endpoint and presents it as though it had the internal structure of a contained model.



The unified model has the following characteristics:

- All resources which are accessible from the unified endpoint are identified using a numbering system which is owned by the unification layer and can be accessed from a URL targeted at the unified endpoint (a unified namespace).
- The data model is complete. All resources which are referenced by resources served from the unified endpoint can themselves be accessed from the unified endpoint.
- Master data items exist as a resource in the unified namespace. Unified resources are linked to local resources which represent the same business object using the FHIR Linkage concept.
- Searches are expressed by data consumers in an identical manner to searches made against a contained model. The unification layer defers fulfilment for searches to local data sources.
- The full search syntax is supported by the unified endpoint. Implementors will need to post-process search results from local data sources to support certain search directives (details below).
- Resource references can be used in parameters in search terms. The type of resource reference used in a search controls the way in which a search is performed (see below).

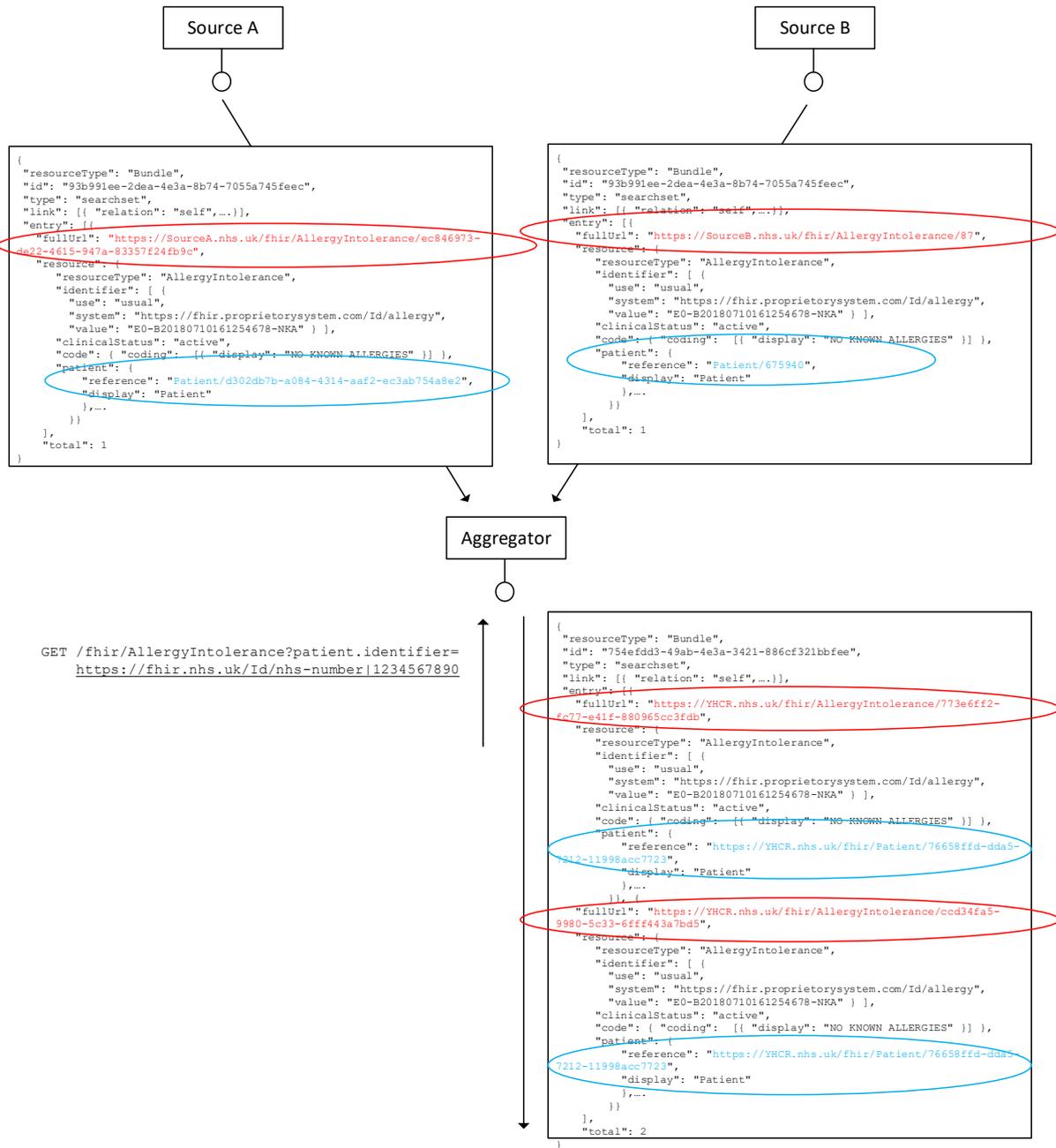
The YHCR architecture includes a regional FHIR Bus which acts as a unifier for the regionally distributed FHIR model. The concepts expressed here apply directly to the design of this component and are implied by the Abstract Cookbook's use of the term FHIR aggregator to describe one of the functions performed by the regional FHIR Bus. FHIR aggregator is used going forward as a synonym for a unifying endpoint.

The concepts are also relevant to individual data providers. Most major healthcare providers operate a number of healthcare system. In consolidation these supply the local patient record. Each of these systems could be seen operating over a contained data model and the challenge for the care provider is to unify this distributed data model for external consumers.

2.3.1 An Illustrative Search Against a Unified FHIR Endpoint

In the following, a search is issued against an aggregator for all allergy intolerances known for a patient. The search is issues to 2 data sources and a single result is retrieved from each. Resource ids for each allergy intolerance are relocated to the unified namespace and embedded references to patient are rebased to reference master data items.

Principles of Search Aggregation



2.3.2 Basic Search Execution Algorithm

In the main, responsibility for executing a search is deferred to local data sources. The FHIR aggregator passes the search, unmodified, to each data source and consolidates the results. The following describes the logical consolidation process:

1. Create a consolidated array from the union of all search results.
2. Replace each resource id in the array with a unified equivalent as follows:
 - a. If the resource is NOT a master data item then use a simple map (local identifier => regional identifier), creating a new identifier if necessary.
 - b. Else if a linkage exists from the local resource to a unified resource then use the unified resource id.

- c. Else if the resource is deterministically equivalent to unified resource then create a linkage to the unified resource.
 - d. Else create a new unified resource from the local resource and trigger an asynchronous process to embellish the resource with a national dataset.
3. Remove any duplicated resources.
4. For each resource reference replace the local reference with a unified reference as follows:
- a. If the resource is NOT a master data item then use a simple map (local identifier => regional identifier), creating a new identifier if necessary.
 - b. Else if a linkage exists from the local resource to a unified resource then use the unified resource id (note that this will be the case if the resource was included in the result set using the `_include` directive say).
 - c. Else create a new unified resource id and trigger an asynchronous process to retrieve the resource, link to any existing unified resource that is deterministically equivalent and embellish the resource from a national data set.

Notes:

- i) Step 4c can result in resource ids being created which are transient. Once the actual resource has been retrieved then it may be discovered that it is deterministically equivalent to another unified resource and the allocated resource id becomes redundant and is linked the existing resource id. Transient resource ids could be avoided by the implementor adding `_include` directives to searches which executed against local data sources. The technique would ensure that all referenced master data items are known to the aggregator before references are resolved.
- ii) Deterministically equivalent resources are master data items which pertain to the same business object. Deterministic equivalence implies that there is a function that can compare 2 resources and classify them as equivalent or otherwise.

$f(a,b)=1$ implies that a and b are equivalent

$f(a,b)=0$ implies that a and b are not equivalent.

The normal rules of equivalence follow:

$f(a,b)=1 \Rightarrow f(b,a)=1$

$f(a,b)=1 \ \& \ f(a,c)=1 \Rightarrow f(b,c)=1$ etc.

2.3.3 Advanced Search Directives and Post Processing Requirements

The above algorithm is sufficient for basic searches where the result set returned from the FHIR aggregator is effectively the sum of result sets returned from individual data sources. Further processing is required to support the following directives:

<code>_sort</code>	The result set is sorted by one or more resource properties.
<code>_count</code>	The result set is returned in pages each containing a maximum number of resources.
<code>_include</code>	Resources referenced by members of the result set are also included in the result set. The include statement lists those search parameters which are counted as references to be included.
<code>_reinclude</code>	Resources referencing resources contained in the result set are also included. The include statement lists those resource types and their references to be searched.

Sorting

The `_sort` directive specifies search terms by which the result set must be sorted. The aggregator interleaves sorted results returned from data sources and maintains the sort order across the aggregated result set.

Sorting within the aggregator requires that all properties on which the sort is dependant are returned to the aggregator by data sources. TF the search uses the directives `_elements` or `_summary` to restrict properties returned, then the aggregator will expand the elements list to include all terms required for the search and strip out unrequested elements when returning results to the data consumer.

Pagination

The `_count` directive is an instruction to return the result set in pages of a maximum size. The aggregator maintains a cursor of sorted results which is periodically refreshed from data sources by issuing a query embellished with a similar `_count` directive. If the cursor contains at least a page size of results and does not contain the last item returned by any one data source (a condition which would indicate a potential gap in a sorted results) then the next page can be served without refresh. Otherwise a refresh is required.

Includes

The includes directive allows a consumer to both search for results and retrieve resources which are referenced by those returned in search results in one interaction. Inclusion can be recursive `_include:recurse` requests resources referenced by those included in a result set as a consequence of an include directive. Including resources in this way reduces the number of interactions required between data consumers and data providers and results in efficient client-side processing.

A non-unified data source must include (to a reasonable level or recursion) requested referenced resources within its contained mode. It should not include resources references outside of its contained model. This is not an error condition and no error or warning should be returned.

An aggregator should post process search results from non-aggregating data sources to include resources which match the `_include` directive but are outside of the boundary of resources returned by local sources. The aggregator should recurse through additional resources retrieved and process recursive includes. The aggregator may limit the number of recursions without reporting an error. A data consumer must have alternative stratagem' s for obtaining resources requested with an `_include` directive but not provided in a result set.

The aggregator should not attempt to process the wildchar `*_includes` property.

Reverse Includes

Reverse includes allows resources to be obtained which reference those included in result set.

A non-unified data source must include (to a reasonable level or recursion) requested referencing resources within its contained mode. It should not include referencing resources outside of its contained model. This is not an error condition and no error or warning should be returned.

An aggregator should attempt to locate resources referencing those resources returned in searches issued to local data sources by issuing queries which explicitly target properties listed in the `_revincludes` directive. It should recurse through additional resources retrieved and process recursive `_revincludes`. The aggregator may limit the number of recursions without reporting an error. Because of the unreliable completeness of the results of a `_revincludes` directive issued to a unified data source a data consumer should generally use more direct search techniques to identify resources which are of interest.

The aggregator should not attempt to process the wildcard `*_revincludes` property.

2.3.4 Resource References in Search Terms

Resource references (absolute and relative) can be used in search terms. Relative references are treated as being relative to the aggregator endpoint.

Absolute references may be local references or unified references.

Unified references to non-master data items are substituted in queries against data sources with an 'or' term which allows the query term to match either an absolute local reference or an absolute unified reference.

Query terms involving unified references to master data items are substituted with a chained search to a unique business identifier for the referenced resource.

For example, a search against the aggregator for observations for a patient which is constructed as:

```
GET /Observation?subject=https://yhcr.nhs.uk/fhir/Patient/12
```

Is issued to data sources as:

```
GET /Observation?patient.identifier=https://fhir.nhs.uk/Id/nhs-number|1234567890
```

(Assuming that Patient 12 at the aggregator endpoint has the NHS number 1234567890).

2.3.5 Creating and Modifying Resources Through a Unified Endpoint

Resources can be modified or patched via the unified endpoint. Modifications to unified master data resources are applied directly by the aggregator. Modifications to local resources are passed to the local data source, the resource first being modified to replace resource references in the unified namespace with local equivalents (if they exist in the data source's namespace).

Unless otherwise specified, new resources are created in a default data source (the regional FHIR aggregator this is the regional FHIR Store). A data consumer specifies an alternative target for a created resource using the HTTP header:

```
FHIR-Base: <url>
```

As for resource modifications, resource references in the unified namespace in newly created resources are replaced with local equivalents (if they exist as the data source's namespace).

2.3.6 Subscriptions

Subscriptions allow a data consumer to express an interest in a data point at a data provider. Subscriptions use search paths to define the data point. If a resource is created or modified that matches the data point, then the data provider sends the resource to the data consumer. From a functional perspective, subscriptions can be used:

- To have a clinician to be alerted about a change in a condition or a drug prescription.
- To define data that should be collected for processing by an algorithm looking for opportunities for early intervention.
- Etc.

Subscriptions created at the unified endpoint are passed onto all primary data sources. Resource references in search terms are modified as described in 2.3.4.

In order to simplify regional firewall configuration subscriptions are delivered to local subscribers via a regional endpoint. The channel endpoint in the subscription is replaced with an endpoint on the aggregator which in turn delivers subscription content to the local subscriber.

2.3.7 A Hybrid Model for Resource Identification

The unified resource identification system described above is conceptually simple but carries a cost of centralising resource acquisition and harmonisation.

To a data consumer, the unified model behaves precisely as a contained model – all resources, regardless of their owner, are identified using common numbering system and are served in a normalised format where all resource references resolve to the same unified endpoint.

The unified endpoint is responsible for all serving all resources, and for every resource served, for normalising references to other resources so that they are contained within a common identifier namespace.

For non-master-data items the unified endpoint is acting merely as a proxy for the data source, introducing an additional hop to acquire data and requiring the resource to be naturalised – local resource references replaced with unified references. All resource requests are concentrated on the aggregator and full; aggregator functionality is executed for every resource interaction including auditing, application of consent rules, resource versioning etc.

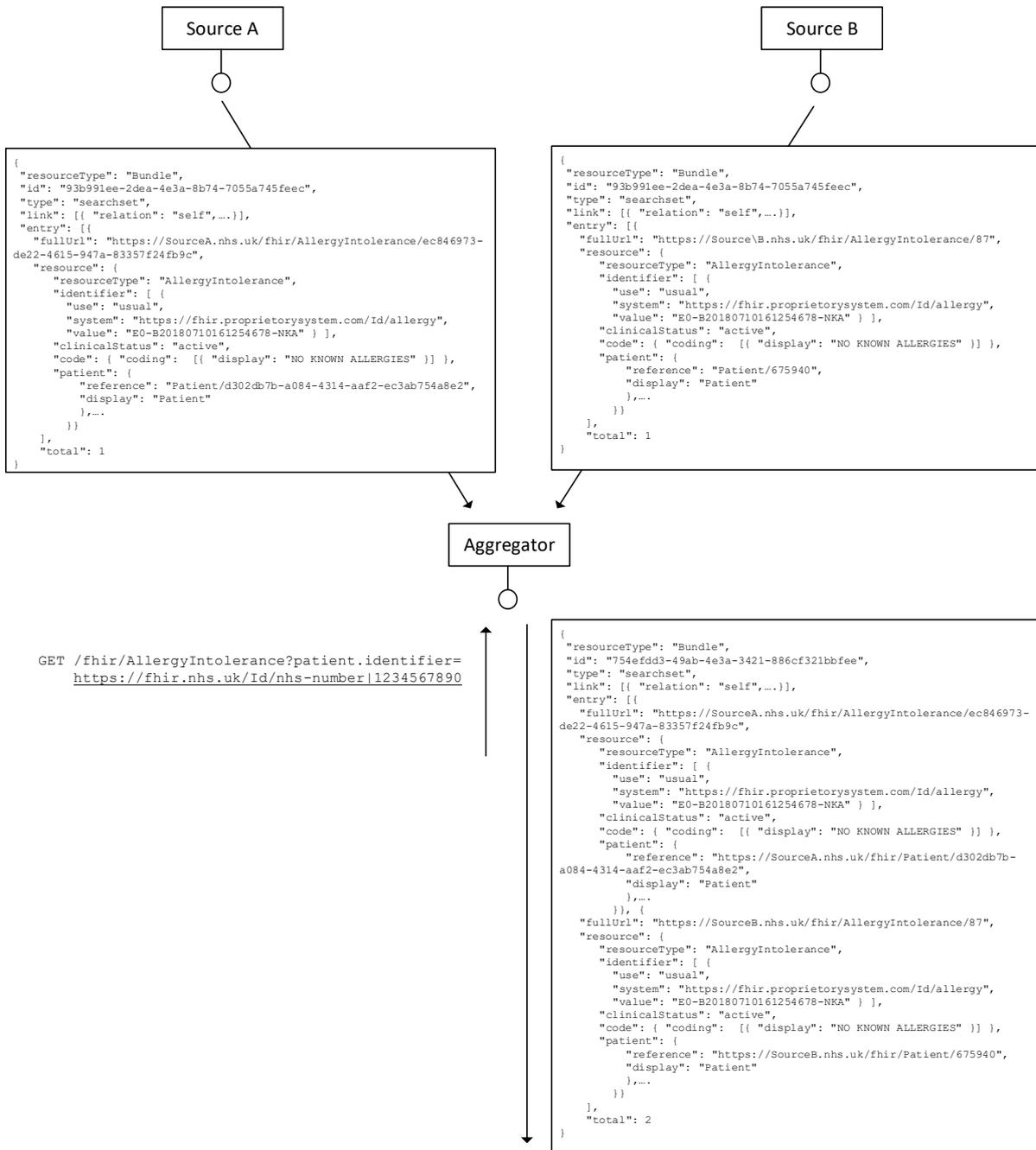
However, many consumers will be targeting individual data sources (the application may be structured to show data from different sources in different tabs, or it may be a local application interacting predominantly with local data). For these, a direct relationship between the data consumer and data provider may be more efficient. In these circumstances the unified endpoint is used to conduct an initial search, with subsequent resource interactions targeted at the data source. A aggregator must operate under two different modes: in unification mode (default behaviour), resource id's are translated into the unified namespace. In local model, local resource ids are preserved. The following HTTP header is used to select the mode:

```
FHR-Identifier-Translation: unified  
Ids should be translated to the unified namespace (default).
```

```
FHR-Identifier-Translation: none  
Ids should not be translated.
```

Using this later mode, the search in 2.3.1 now returns as follows:

Search Aggregation without Identifier Harmonization



Note that resource ids and references are preserved. The only translation is from relative to absolute identifiers.

3 Master Data Management in a Unified Model

Master data management relates to the maintenance of master data items in the unified namespace for physical concepts such as people, objects, locations, and entities. Master data items duplicate resources which are maintained locally and pertain to the same concepts. Local resources pertaining to the same concept may differ in their properties. For instance, a local resource representing a patient may have a different primary address from a resource representing the same patient in another location.

Master data items maintained by the regional FHIR aggregator comprise:

- Patient.
- Organisation.
- Practitioner.

Master data items are constructed on demand whenever a local equivalent resource is retrieved by the aggregator. Local resources and master data items are linked using FHIR Linkage resources. A data consumer can locate all local equivalents for a master data item by querying the regional Linkage resources. This technique could be used by a user interface, for instance, to obtain and display a patient's home address as it is recorded at a particular data source rather than the home address on the unified data item.

Master data management relies local and unified resources being matched using a deterministic mechanism. For the YHCR aggregator this is:

Master Data Item	Matching Method
Patient	NHS Number
Organisation	ODS Code
Practitioner	ERS Number

Master data items are constructed from one or more sources. Sources are process according to a tiering system. Each tier is interrogated in turn for properties missing properties of the resource. The YHCR FHIR aggregator uses JSON Path to specify properties collected from each tier. Initial configuration will be:

Master Data Item	Tier	Properties	Tier Source
Patient	1	\$.*	PDS
Organisation	1	\$.*	ODS
Practitioner	1	\$.*	ERS

Master data items are rebuilt periodically. A rebuild is triggered by access to a stale item. The definition of stale is parameterizable for the YHCR FHIR aggregator.