Cookbook for Regional Interoperability Detailed Design Paper #017

Data Impairments – Reporting and Interpretaion

PRELIMINARY DRAFT

Version 1.0 – 7th December 2019

Abstract Interoperability Cookbook Anchor Points

Section	Title
6	FHIR Resource Profiles

Table of Contents

Version Control

Version	Release Date	Released By	Reason for Release
1.0	07/12/2019	R Hickingbotham	Preliminary draft

Reviewers

Initials	Name	Role	Organisation

1   Introduction

1.1   Purpose of this Document

This document is one of a series of design papers which underpin the Abstract of a Cookbook for Regional Interoperability (the Abstract Cookbook). These papers, in their totality, describe the technical components and the standards which form the YHCR System of Systems (SoS). They are intended as a basis for developing or procuring software and so are expressed at a level of precision which aims to avoid ambiguity but consequentially, their intended audience is technical readers.

Design papers are anchored to topics which are discussed in the Abstract Cookbook. 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 017 - "Data Impairments") concerns itself with imperfections in data quality: how data quality issues might arise, mechanisms for data providers reporting imperfections, and possible responses by data consumers to reported quality issues.

The paper uses a systematic definition of data quality: factors which are endemic to the way that the YHCR has been constructed which might lead to an interpretation of data by consumers which is different from the intention of providers who have been commissioned to supply it.

The paper sidesteps the broader understanding of data quality which is based on consistent representation of concepts in data by different individuals in different circumstances. Identifying, these data quality issues and taking measures to improve them is a topic for ongoing consideration.

The aim of the paper is to define how a participant in the YHCR which is aware of, or detects, a data quality issue informs other participants in a manner that allows them to act appropriately.

1.2   Normalisation, Data Quality, and Maturity

1.2.1   Improving Data Standards Through the YHCR Maturity Model

The YHCR operates a maturity model which aligns data providers within 6 discrete maturity bands. The providers within each band are assumed to comply with, and have been verified as part of an onboarding process as complying with, a common set of standards. The standards are multi-dimensional in that they cover several different aspects of data provision such as the:

- the data items which are available from a provider;
- coverage across patients and service lines;
- content of data items and use of coding systems;
- sophistication of interfaces through which data is provided.

The intention of the maturity model is to allow participants to join the YHCR with a relatively low barrier to entry but to also encourage evolution of capabilities so, at the highest level of maturity data providers are contributing to a normalised longitudinal care record which is fully structured, uses national data vocabularies, is bereft of duplicated data, and provides good coverage of the population of Yorkshire & Humber.

The YHCR's intentions regarding normalisation are detailed in design paper 26 - "Data Normalisation".

1.2.2 Maturity Levels and the Definition of Data Quality

Alignment with the maturity model explicates the interpretation of data quality. Data which is provided in accordance with the definition of a maturity level is "high quality" despite the possibility that data of the same type may be provided by a different data provider at a higher level of maturity and in closer conformance to the YHCR's ambition for normalised data. A data provider's maturity is published by the YHCR and a data consumer is aware of any constraints in the data that it receives from each data source and it can facilitate interpretation by its users of data supplied at different maturity levels. A data consumer is also able to align itself with a maturity level and only receive data from providers which are at or exceed that level of maturity (see design paper – 021 "Onboarding for Data Consumers".

1.2.3 Non-Compliance and Data Quality

"Poor quality" data is data which does not comply with the definition of the maturity level with which the provider is aligned. A data provider may knowingly provide "poor quality" data and this is acceptable in certain circumstances. For instance, there may be an imperative to release some data at a high maturity level but full compliance with the maturity level is not possible for all data types.

In this situation the provider is responsible for informing data consumers of quality issues by inserting statements into query result sets. This paper considers the format and content of these statements.

1.2.4 Restricted Coverage and Data Quality

A data provider's participation in the YHCR implies that it is able to supply the data types defined by the maturity level at which the provider operates for every patient that the provider registers with the regional PIX/MPI server (design paper 004).

Practically, there will be circumstances where there are gaps in the data which is available for some patients. Healthcare systems tend to be allied with service lines and specialties leading to inconsistency in data which is captured within an organisation and differences in accessibility for providing data to the YHCR.

Data providers must inform data consumers of gaps in data returned in query results using a similar method as that employed for non-compliance issues.

1.2.5 Out of Date Data

One dimension of maturity is the ability to surface data to the YHCR on a timely basis. The time lag between data being created and made available to the YHCR decreases with progression through maturity levels.

For various reasons a data provider may not be able to comply with the data currency stimulations of the maturity model and if so is obliged to inform data consumers.

1.2.6 Transient Issues Affecting Data Completeness

Data access may be restricted because of system or network unavailability. Issues preventing a data consumer accessing the SoS are explicit and the data consumer can take appropriate action. Issues which are downstream from the SoS endpoint are hidden from the consumer and manifest themselves as data quality problems which should be reported using the standard mechanism.

Integration components in the YHCR or at data providers which detect a connection problem to a downstream component or data source must insert statements into query result sets which inform data consumers of potential missing data.

1.3 Reporting of Impairments to Data Quality and Interaction Patterns

The SoS implements 4 interaction patterns through which data consumers can obtain data from data providers. These are:

1. Synchronous query.
2. Asynchronous query.
3. Subscriptions.
4. Messaging.

All the patterns involve data being exchanged using the FHIR standard but are serviced by data providers in different ways. Data quality impairments have different interpretations for each these patterns. Statements of impaired data quality are only required for query type patterns.

A synchronous query (which includes direct resource access) is serviced on demand; a provider's assessment of data quality pertains to the point at which the request is made, and it is able to report its assessment of data quality in its response to the query.

An asynchronous query is serviced at some time after the request is made. A data provider can address certain data quality problems before serving results. The format in which results are served allows the data provider to embed assessments of data quality in its response.

Subscriptions or messaging do not need a mechanism for reporting impaired data quality:

- if a subscription is accepted by a provider, then the consumer can expect that it will be serviced accurately. The provider has a mechanism (see design paper 007 – "Subscriptions Infrastructure") to reject subscriptions which cannot be services because of known shortcomings in maturity or gaps in data coverage. In common with asynchronous queries, a subscription is processed as a background task and a provider can address transient problems in accessing data before issuing a notification.

- the messaging pattern is only used to disseminate data which conforms to a universally agreed structure and content. Organisations should only choose to implement a message type if they can
    a) recognise the event
    b) supply the data content representing the event.

1.4 FHIR and Statements of Impaired Data Quality

FHIR offers a resource, the *OperationOutcome*, which provides detailed information about the outcome of an attempted system operation. They are provided as a direct system response or component of one and provide information about the outcome of the operation.

This paper proposes to use *OperationOutcomes* as a YHCR standard for data providers informing data consumers of detectable data quality issues. It is applied to the synchronous and asynchronous interaction patters as a resource that can be inserted into search results alongside actual results content.

1.4.1 Data Access Controls and *OperationOutcomes*

Design paper 008 – "Data Access and Consent Management" describes the use of *OperationOutcome* resources to inform data consumers that data has been withheld or should be treated a restricted because of a data access control including a consent restriction. Whilst not strictly an impediment to data quality, the possible reaction by a data consumer is similar and the use of *OperationOutcomes* for this purpose is within the scope of this paper.

1.5 Relationship of this Document with Other Standards

This paper is a statement of intent rather than a design and does not rely on any particular standard although many standards will be used in the implementation of the intent:

- FHIR;

1.6 Intended Users of the This Document

Developers and users of the regional model FHIR proxy and other data providers supplying their own FHIR proxy services.

Developers of data consuming software.

Providers of data to the YHCR.

2 The YHCR *OperationOutcome* Profile

The YHCR *OperationOutcome* is a constrained version of the STU3 standard resource as follows:

Element	Resource Cardinality	Profile Cardinality	Implementation Notes
issue	1..*	1..*	One or more issues pertaining to the operation.
issue.severity	1..1	1..1	The issue's severity using the FHIR STU3 http://hl7.org/fhir/ValueSet/issue-severity value set.
issue.code	1..1	1..1	Uses the FHIR STU3 value set: http://hl7.org/fhir/ValueSet/issue-type
issue.detail	0..1	1..1	Details of issue are mandatory and must use the http://yhcr.nhs.uk/fhir/valueset -issue-detail coding system. The text value of the CodeableConcept should be a narrative that is interpretable by an end user of a system which describes the issue.
issue.diagnostics	0..1	0..1	Additional diagnostic information about the issue.
issue.location	0..*	0..*	Path of element(s) related to issue (note the location is deprecated in the FHIR standard and expression(s) should be used instead).
issue.expression	0..*	0..*	FHIRPath of element(s) related to issue. The YHCR profile extensions to the FHIR string type of this element to allow it be used to describe data content which is missing or restricted in search bundles. These extensions must be used when using OperationOutcomes to describe a data quality issue or a data access restriction. See section 2.2 for details.

2.1 Use of the *OperationOutcome* resource

OperationOutcome resources can be used to return information about any FHIR operation as a response to an HTTP request. They can be used for purposes other than reporting data impairments. Of particular interest for the latter case is the use of *OperationOutcomes* when responding to search results.

The resource content of entries in a FHIR bundle can represent both search results and *OperationOutcomes*. *OperationOutcomes* embedded in a bundle in this way can inform a consumer of data which is missing because of unavailability, non-compliance with a maturity level, or data withheld because of a data access policy (design paper 008 – "Data Access and Consent Management") but it is not useful if the intention is to qualify the interpretation of a specific resource which appears as a another entry in the search result bundle. Examples of where this situation may occur includes:

- the data source is aware that some or all of search results may be stale having been served from an out-of-date offline copy of a live database;

- the resources are subject to data access to controls and are being released under condition that the user is warned that the resource content is restricted.

Under these circumstances the *OperationOutcome* should cross-reference the resources in the bundle to which it relates. Whilst this is theoretically possible using an issue expression to locate the

resources, this would place a complex processing responsibility on consumers and could result in unnecessarily large bundles being exchanged.

The YHCR will use the response attribute of a *Bundle* to associate an *OperationOutcome* with one or more resources. The response attribute is intended for use in transaction and batch processing but is structurally legitimate to use for other purposes. The response attribute is a property of a *Bundle* entry and references an *OperationOutcome* included elsewhere in the *Bundle*.

If the interpretation of a resource returned in response to a search requires qualification then the data source will insert an *OperationOutcome* into the results and reference it using the response.outcome property of the entry.

Note that all OperationOutcome resources, in common with all resources in the YHCTR, must identify the source using a meta tag in the format:

```
"meta": {
    "tag": [
        {
            "system": "https://yhcr.nhs.uk/Source",
            "code": "HELM",
            "display": "Helm"
        }
    ]
```

2.2 Using Expressions to Specify the Scope of Data Impairments

The following extensions enables an expression to be used to construct a search term.

Name	Flags	Card.	Type	Description & Constraints
❋ Selector		0..1	Extension	URL = http://fhir.yhcr.nhs.uk/schema/StructureDefinition/OperationOutcome-selector Modifies an expression to allow it to be used to describe a resource scope
comparitor		1..1	code	A comparitor to apply to the expression eq,ne,gt,lt,ge,le,sa,eb,ap
value		1..1	string	The value to compare in the exporession

The extension applies to issue.expression which is a string property and is used to provide a FHIRPath to element properties to which the *OperationOutcome* relates. This extension adds a comparator and value which allows search to be constructed from the expression.

The search string relates the *OperationOutcome* to resources which were requested by the consumer but could not be provided. The Consumer can interpret the search string and explain constraints in the data being accessed to an end user.

The following examples show how the Selector extension can be applied to an expression in order to relate an *OperationOutcome* to observations arising from laboratory tests.

JSON

```
"expression" : "Observation.category"
"expression_" : {
    "extension" : [{
```

```
    "url" :
"http://fhir.yhcr.nhs.uk/schema/StructureDefinition/OperationOutcome-
selector",
    "extension" : [
        {
            "url" : "comparator",
            "valueCode" : "eq"
        },
        {
            "url"  : "value",
            "valueString" : "http://hl7.org/fhir/ValueSet/observation-
category|laboratory"
        }]
    }]
  }
```

XML

```
<expression value="Observation.category"
    <extension
    url="http://fhir.yhcr.nhs.uk/schema/StructureDefinition/OperationOutcom
e-selector"
        <extension url="comparator" >
            <valueCode value="eq"/>
        </extension>
        <extension url="value" >
            <valueString value="http://hl7.org/fhir/ValueSet/observation-
category|laboratory" />
        </extension>
    </extension>
</expression>
```

2.3   Code Systems

Issue.detail

http://yhcr.nhs.uk/fhir/valueset -issue-detail extends http://hl7.org/fhir/ValueSet/issue-type

The following codes are defined:

Code	Display	Definition
For use with code = informational		
MSG_RESTRICTED_RESOURCE[1]	Restricted Resource	The resource is controlled by a data access policy that allows it to be released on the condition that the user is informed of its restricted status.
MSG_POSSIBLY_STALE[1]	Possibly Stale	The resource has been served from a data source that is out of date in accordance with the aligned maturity level.
For use with code = suppressed		
MSG_RESTRICTED_RESOURCE[2]	Restricted Resource	Referenced resources are controlled by a data access policy that prohibits the resources being released.
For use with code = incomplete		
MSG_UNAVAILABLE[2][3]	Data Source Unavailable	A data source could not be reached, or the connection timed out.
MSG_PARTIAL_COVERGE[2][5]	Partial Coverage	The data source is only able to serve a partial result

		set for a resource type as defined by the aligned maturity level.
MSG_NO_COVERAGE[2][4][5]	No Coverage	The data source cannot serve the resource type requested despite it being included in the definition of the aligned maturity level.

Notes

(1) Outcomes of this type must be referenced by the bundle entry to which they relate using the response attribute of the entry.

(2) Outcomes of this type do not relate to specify bundle entries and must provide and expression extended as per 2.2 which allows a consumer to determine the scope of missing data.

(3) An MSG_UNAVAILABLE outcome will be returned by the FHIR Aggregator for each requested resource types if a data provider cannot be reached. The expression will identify the provider using the _tag search term against the https://yhcr.nhs.uk/Source coding system.

(4) An MSG_NO_COVERAGE outcome will be returned by a data provider for each resource type that it is unable to provide which would otherwise be available in accordance with the provider's maturity level. The expression identifies the resource type.

(5) A data provide should not return an outcome if it cannot provide data because its organisation does not offer the services to which the requested resources pertain. For example if the request is for observations captured at an encounter with an Ophthalmology service and this service is not provided then an empty bundle should be returned.

3   Guidance on Interpretation of *OperationOutcomes* for Data Consumers

The presence of an *OperationOutcome* in search results is an important indicator for data consumer that data can only be interpreted by an end user with full understanding of the meaning of the issue being reported. Without this information the user can be reasonably be expected to treat the data as being complete and any unreported gaps introduce a clinical safety risk.

A consumer can choose to reject any search results which contain OperationOutcome resources (with appropriate notice to a user). If they choose to accept the results, then the information in the outcomes must be interpreted.

Outcomes which are attached to individual search entries require specific treatment by a data consumer.

The data associated with outcomes of type (code=informational, detail=MSG_RESTRICTED_RESOURCE) must not be revealed to users unless the user acknowledges the restriction in the text of the outcome's detail. The acknowledgement must be logged by the consumer.

Data associated with outcomes of type (code=informational, detail=MSG_POSSIBLY_STALE) might be out-of-date (more so than would normally be expected for data obtained from a provider at the declared maturity level) and data consumer may, depending on clinical risk, offer an indication of this to a user.

Other outcomes describe data which is missing from search results. The textual content of the outcome detail should be sufficiently precise to inform a user of the scope of the missing data. This will be verified to be so as part of the Clinical Safety Assessment for data providers as part of the onboarding process. However, a consumer may choose to interpret the coded expressions on *OperationOutcomes* to perform more complex processing such as removing fields or tabs from a user interface. It is anticipated that the expression attached to a MSG_UNAVAILABLE outcome will frequently be used to identify the participant organisation and provide a visual indicator to a user that data from a particular source is not currently available.

4 Implementation in the YHCR Model FHIR Proxy

The YHCR Model FHIR Proxy is introduced by design paper 004 – "Conceptual Design for a FHIR Proxy Server". This paper outlines the functionality of a software product which is available for deployment by a data provider and is intended to reduce the barrier for entry in complying with YHCR standards.

The Model FHIR Proxy must facilitate support for the standards for reporting data impairments as specified in this document. Specifically, it must enable implementors to describe gaps in their alignment to a level in the maturity model in a way that the FHIR Proxy can automatically insert *OperationOutcome* resources in search result *Bundles*.

Three points in the FHIR Proxy processing pathway are affected:

1. When interacting with a data source through an SQL or FHIR Adapter the FHIR Proxy must return an *OperationOutcome* of type (code= incomplete, detail=MSG_UNAVAILABLE) if the data source cannot be reached. The invoking pathway returns the outcome to the service user wither directly as a response to the HTTPS request or by inserting it into a search result *Bundle*.

2. When applying consent an *OperationOutcome* may be inserted into search results if data is withheld or classed as restricted. This behaviour is described by design paper 008 – "Data Access and Consent Management" and is not considered further here.

3. When assembling a search result bundle *OperationOutcome* resources may be inserted for searches for resources which are not available because of incomplete coverage at the providers declared maturity level. This functionality is considered further below.

4.1 No Data Coverage

If a data provider does not support a resource type which is defined at the maturity level with which the provider is aligned then, through configuration, the FHIR Proxy must insert an *OperationOutcome* into the result Bundle for any search which might be expected to return resources of the unsupported type.

This should be achieved through simple configuration whereby unsupported resource type are identified and associated with a narrative to include in the *OperationOutcome* detail text.

A search might be expected to return a resource type if it is:

- the subject of the search;
- referenced in an _include directive;
- referenced in a _revinclude directive.

4.2 Partial Data Coverage

Partial data coverage occurs when a provider is unable to provide subset of the data which might covered by a search a given maturity level. Some results may be returned but there is a recognised gap in the data. For instance a provider may be able to supply *Observations* for laboratory tests, and also collects vital-signs but, for whatever reason, is unable to offer them as a FHIR resource. If they

are aligned with a maturity level which requires support for both then there would be a gap in the data returned for a search returned for all *Observations.*

The FHIR Proxy must be configurable to allow such gaps to be specified and associated with data which enables an *OperationOutcome* to be generated.

A data gap can be identified using a set of FHIR search terms, each targeted at a resource type.

For instance, the following determines:

- Observation.category=http://hl7.org/fhir/ValueSet/observation-category|social-history
- Observation.subject.deceased=true

that social-history observations are not available for deceased patients.

The FHIR Proxy must establish whether defined data gaps may apply to a search and if they do then to insert and *OperationOutcome* into search results which explains the gap.

How to apply the definition of a data gap depends on the construction of the search.

4.2.1   Searches Targeted at the Gap Resource Type

If the gap relates to the subject of a search, then the search term and the gap definition must be resolved to determine whether there is an intersection. An intersection exists if, for each search parameter in common between the gap definition and the search term, there is a possible overlap.

A search of Observations using the following search term:

```
Observation?subject.deceased=false
```

intersects with the gap definition specified above because there is a search parameter in common and there is no possible common data point between the two.

Whereas a search against

```
Observation.category=http://hl7.org/fhir/ValueSet/observation-
category|laboratory
```

does not intersect.

An *OperationOutcome* of type (code=incomplete, detail=MSG_PARTIAL_COVERGE) must be inserted into the result set for any intersection.

4.2.2   Search Terms Dependent on Data Gaps

A search may be conducted against a resource type which includes terms targeting gaps in other resource types.

For instance, a search for *DiagnosticReports* might stipulate that report was collected as part of an encounter at a particular clinic. A data gap based on *Encounters* might constrain results.

Determination of the gap and whether to add ObservationOutcomes is logically similar to 4.2.1 except that the search term must be re-expressed to make the dependent resource the subject of the search.

A search term should be constructed from any chained search terms in the original search that relate to properties (or chained properties) of the resource type that is the subject of the gap definition.

If this, reduced search term, intersects with the gap definition then an an *OperationOutcome* of type (code=incomplete, detail=MSG_PARTIAL_COVERGE) must be inserted into the result set.

The outcome's expression should be based on the original resource being searched.

4.2.3   Resources included in search results

A resource which is included is a resource which is referenced by one of the resources returned in the result set.

Determination of whether to add ObservationOutcomes for included resource types is logically similar to 4.2.1 except that the search term must be re-expressed to make the resource being included the subject of the search.

A search term should be constructed from any chained search terms in the original search that relate to properties (or chained properties) of the resource type that is the subject of the gap definition.

If this, reduced search term, intersects with the gap definition then an an *OperationOutcome* of type (code=incomplete, detail=MSG_PARTIAL_COVERGE) must be inserted into the result set.

The outcome's expression should be based on the included resource tyoe resource being searched.

4.2.4   Resources reverse included in search results

An *OperationOutcome* of type (code=incomplete, detail=MSG_PARTIAL_COVERGE) must be inserted into the result set for every gap expression for every resource type which is reverse included in a query.

Appendix 1 – Maturity Matrix

Section	Narrative	Consultative	Draft	Normative
1 Introduction 1.1 Purpose of this Document	X			
1.2 Normalisation, Data Quality, and Maturity	X			
1.3 Reporting and Impairments to Data Quality and Interaction Patterns	X			
1.4 FHIR and Statements of Impaired Data Quality		X		
1.5 Relationship of this document to other standards	X			
1.6 Intended Users of the This Document	X			
2 The YHCR OperationOutcome Profile 2.1 Use of the OperationOutcome resource			X	
2.2 Using Expressions to Specify the Scope of Data Impairments			X	
2.3 Code Systems			X	
3 Guidance on Interpretation of OperationOutcomes for Data Consumers		X		
4 Implementation in the YHCR Model FHIR Proxy 4.1 No Data Coverage			X	
4.2 Partial Data Coverage 4.2.1 Searches Targeted at the Gap Resource Type			X	
4.2.2 Search Terms Dependent on Data Gaps			X	
4.2.3 Resources included in search results			X	
4.2.4 Resources reverse included in search results			X	