INTERWEAVE
CONNECTING CARE

Cookbook for Regional Interoperability
Detailed Design Paper #006

Reliable Messaging Infrastructure

PRELIMINARY DRAFT

Version 1.0 – 16th March 2019

Abstract Interoperability Cookbook Anchor Points

Section	Title
4.3	Subscriptions

Table of Contents

Version Control

Version	Release Date	Released By	Reason for Release
1.0	16/03/2019	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. 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. 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 006 - "Reliable Messaging Infrastructure" describes a regionally standard approach to sending messages between organisations in the YHCR, and beyond, in a manner where delivery to the intended recipient organisation is guaranteed.

There is a relationship between messaging and event publishing. Healthcare messages are often emitted as a consequence of an event happening in the context of a patient: a patient dies, a patient is attended in ED, a patient is discharged from a facility. NHS Digital operates the National Event Management Service (NEMS) which enables organisations to publish details of a predefined set of patient events and for subscribing organisations to receive notification of events as they happen to patients with whom they have a relationship. The service is limited in scope at the time of writing but the YHCR needs to be cognisant of potential future engagement with NEMS. This paper examines the relationship between event publishing and proposes a mechanism for regional infrastructure to facilitate contribution to NEMS as a biproduct of its role in message delivery.

1.2 What is Reliable Messaging?

Reliable messaging relates to the ability to make certain guarantees about message delivery when messages are being sent over unreliable infrastructure. From a messaging perspective infrastructure is inherently unreliable when messages and acknowledgements are sent between systems asynchronously: the message may travel through a number of intermediary devices on route from one endpoint to another; a failure in any one may cause the message to be lost.

There are a number of possible guarantees which may be desirable for senders and receivers of messages:

- a message must be delivered to its recipient at least once;
- a message must be delivered to its recipient at most once;
- a message must be delivered to its recipient exactly once;
- messages must be delivered to their recipient in the same sequence as they were dispatched.

From the perspective of the YHCR the only guarantee which will be considered is that messages must be delivered exactly once: this being the most commonly desired outcome.

It is recognised that maintaining sequential integrity is also a common requirement. However, it is asserted that the need for and interpretation of what sequential integrity means is determined at a business level and is best enforced by the recipient of messages with its understanding of the context in which messages are being received.

1.3 Reliable Messaging in the YHCR

The YHCR will use messaging for two main purposes:

- to deliver the results of subscriptions;
- to deliver inter-organisational transactions.

The last category may include referrals to services, transfers of care, correspondence, and reports. Examples encountered in the System of Systems pilot are referrals of patients receiving cancer care between oncology centres and transfer of care from an ambulance to an emergency department.

For both categories there is well defined sender of and a unique recipient.

This document does not concern itself with the content of messages. For subscriptions, content is defined by design paper 007 – “Subscriptions Infrastructure”. The content of transactions will be designed on a case-by-case basis either nationally or by the YHCR Data Architecture Design Authority.

It is possible that messaging in the FHIR could take place in a point-to-point manner whereby the originator of a message sends the message directly to its intended recipient (messages may pass through integration engines or other middleware, but these are operated by the messaging parties). However, this document focuses on messaging which uses regional infrastructure as an intermediary. The benefits of this approach are:

- firewall rules are simplified as all organisations send to and receive from a single endpoint;
- regional infrastructure can act on message content and perform secondary functions such as seeding NEMS.

1.4 The FHIR Messaging Standard

The HL7 FHIR standard includes consideration of messaging. It defines the structure of both a request message and a response message, and it defines a *MessageHeader* resource which contains metadata about the message.

The FHIR standard considers a number of messaging scenarios including messages which are broadcast to a number of recipients and a request message which results in many response messages. This design paper concentrates on a single paradigm: a request message is sent to a single recipient and receives a single response message. In FHIR terminology this is a message of “consequence”.

1.5 Relationship of this Document with Other Standards

The following standards form the basis for this document:

- FHIR Release 3 (STU) – Messaging Using FHIR Resources;
- InterOpen CareConnect FHIR profiles;

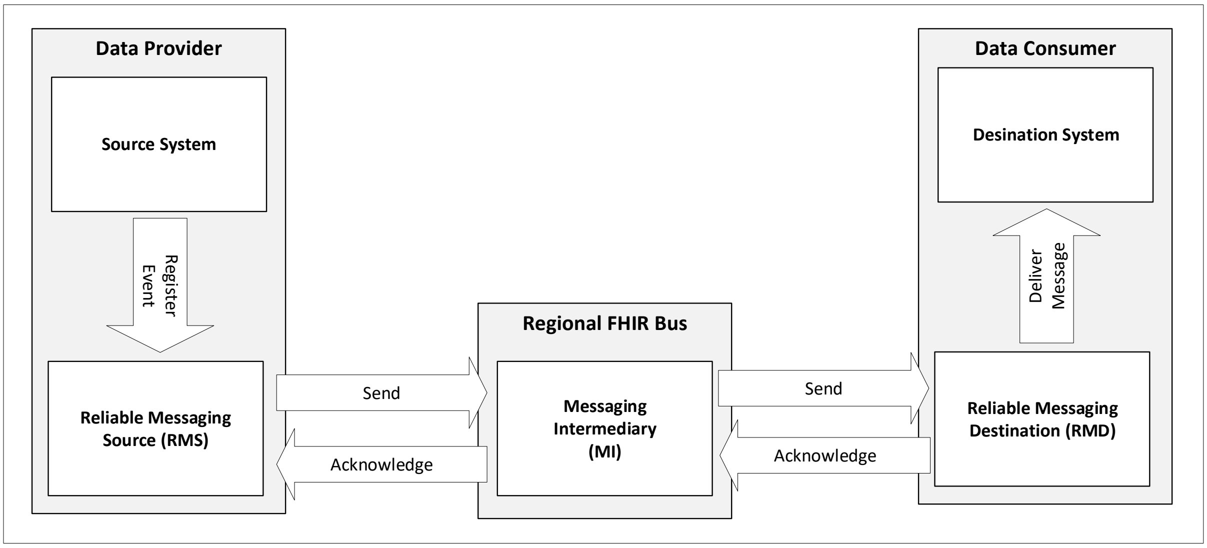
- National Event Management System (NEMS);
- ITK3 Messaging Distribution;
- Message Exchange for Health and Social Care(MESH)
- HTTP.

1.6 Intended Users of the This Document

This document is a reference guide for data providers intending to send messaged within the YHCR, data consumers intending to receive messages, and developers of regional infrastructure. The document should also be read by organisations participating in National Event Management Systems pilots or those intending to publish data to this service.

2 A Regional Messaging Architecture

The messaging infrastructure described in this paper is based on the following architecture:



The architecture is based on an asynchronous approach to data exchange. Data providers and data consumers employ integration technology to provide the reliable messaging components RMS and RMD. When an event is registered with RMS by a source system, then this component assumes responsibility for delivering a message which represents the event to its destination. It provides the source system with the guarantee that the message will be delivered exactly once. The reliability of the event registration mechanism between the source system and RMS is a matter for the data provider and is outside the scope of this document.

RMS sends the message to its associate RMD via a component on the regional FHIR Bus – the Messaging Intermediary (MI). RMD receives the message, acknowledges it and delivers the content of the message to the destination system, again the reliability of the delivery mechanism within the data consumer is outside the scope of this document. The acknowledgement is delivered via the regional MI.

Reliability, and the guarantee that a message is delivered exactly once, is ensured by adopting the following principles:

1. Every message has a globally unique identifier.
2. An acknowledgement is generated for every message received by RMD.
3. If RMS has not received an acknowledgement for a sent message within a certain time, then it resends the original message.
4. If RMD receives a duplicate message (based on its globally unique identifier) then it acknowledges it but does not deliver the message to the destination system.

The regional Messaging Intermediary does not participate (except where acting as a mediator in transport mechanisms outside of the YHCR) in ensuring reliability and acts only as a conduit for messages.

The FHIR standard describes the behaviour of messaging components when acting a reliable messaging environment. The text in the following sections in *italics* paraphrases the FHIR STU3 standard.

2.1.1 Reliable Messaging Source

A sender which implements reliable messaging shall, in the circumstance that it receives no response to a message within a configured timeout period, resend the same message (with the same MessageHeader.id and with the same Bundle.id).

Here the bundle id acts as the globally unique id required by reliable message principle #1.

The resend period will be advised by the YHCR and will be published in the YHCR Operations Guide. However, the selection of an appropriate value is ultimately a local decision. The value is published in the data providers FHIR *CapabilityStatement* messaging.event.category for the event associated with the message.

Note that an RMS may receive multiple acknowledgements for the same message. It must desist in resending messages on receipt of the first acknowledgement and discard any subsequent acknowledgements without action.

2.1.2 Reliable Messaging Destination

A receiver which implements reliable messaging shall check the incoming Bundle.id and MessageHeader.id against a cache of previously received messages. If both the Bundle.id and MessageHeader.id have not been received then this is the normal case, and the message should be processed and acknowledged. If both the envelope and message match cached ids then the original response has been lost and the original acknowledgement resent.

Note that in messaging environment which only trades messages of “consequence” then the *Bundle.id* and *MessageHeader.id* are synonymously unique. If messages are received where one the *Bundle.id* or the *MessageHeader.id* have previously used in a cached message but not both this is an error situation. The message must be discarded and a negative acknowledgement issued.

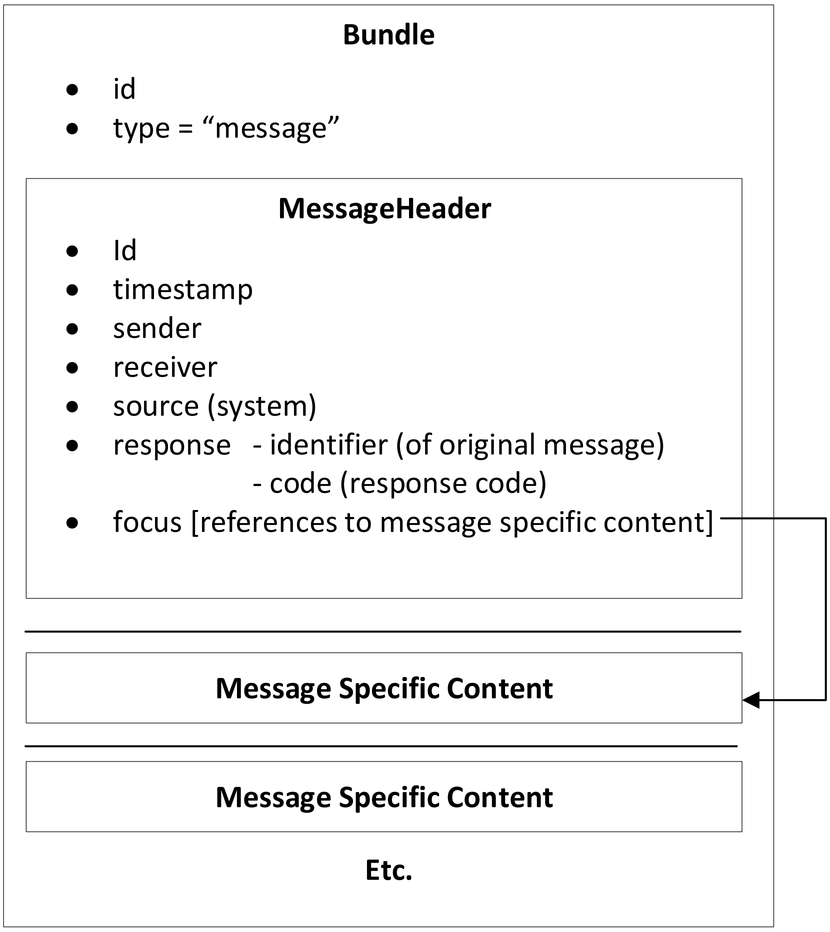
2.2 Structure of a FHIR Message

A FHIR message is a bundle of type “Message”. The first entry of the bundle is a MessageHeader resource and subsequent entries contain message specific content.

The same structure is used for both request and response messages. The response message (or acknowledgement) references the original request message in the MessageHeader and supplies a response code.

Message content is included in subsequent sentries in the bundle and resources are referenced through the focus property of the *MessageHeader*

The message structure is illustrated below.



2.3 The YHCR FHIR *MessageHeader* Profile

The YHCR *MessageHeader* is a constrained version of the STU3 standard resource. The profile adjusts the STU3 *MessageHeader* resource definition as described in the following table.

Element	Resource Cardinality	Profile Cardinality	Implementation Notes
event	1..1	1..1	Uses the YHCR coding system: https://yhcr.nhs.uk/STU3/ValueSet/EventType-1
destination.endpoint	0..1	0..0	The endpoint is derived from receiver Organisation
receiver	0..1	1..1	A receiving must be provided. This is a Reference to an Organisation resource. STU3 also permits a reference to a Practitioner which is not supported by this profile.
sender	0..1	1..1	A sender must be provided. This is a reference to an Organisation resource. STU3 also permits a reference to a Practitioner which is not supported by this profile.
reason	0..1	0..0	The reason for the message is sufficiently encapsulated by the event code. This refinement is consistent with profiling for NEMS.

Note that the profile is intended for *Organisation* to *Organisation* messaging. The STU3 resource definition also supports *Practitioner* to *Practitioner* messaging. If, given the business context of message, delivery to an individual is required then this is the responsibility of the recipient organisation and is to be undertaken based on the content of the message.

Constraining message senders to be *Organisations* does not prevent the author of the message or person responsible for the emitting event being identified in the corresponding properties of the *MessageHeader*

The receiver of a message must be a data consumer which is registered with the YHCR (reference design paper 021 “Onboarding Data Consumers”). The *MessageHeader* receiver property must reference an *Organisation* profiled by Care Connect which is identified using an ODS code. The ODS code is used by the regional FHIR to determine the destination endpoint URL to which the message will be delivered.

The sender of a message must be a data provider which is registered with the YHCR (reference design paper 020 “Onboarding Data Providers”). The *MessageHeader* sender property must reference an *Organisation* profiled by Care Connect which is identified using an ODS code. The PKI certificate used in the connection to the regional FHIR Bus must be the one been issued by the YHCR to that provider.

2.4 Event Type Value Set

The YHCR Event Type value set is defined at: https://yhcr.nhs.uk/STU3/ValueSet/EventType-1. The value set is a super set of the NEMS event type value set at the time of writing comprises:

Code	Description
FM001	National Population Failsafe Alert
FM002	National Population Failsafe Alert Nullify
PDS001	PDS Change of GP
PDS002	PDS Change of Address
PDS003	PDS Birth Notification
PDS004	PDS Person Death
YH001	Transfer of care from ambulance to ED
YH002	Cancer centre secondary referral

2.5 REST Messaging Interface

All interactions between RMSs, the regional FHIR Bus Messaging Intermediary and RMDs are RESTful.

The REST interface is over HTTPS secured with regionally issued certificates using mutual TLS authentication.

The messaging endpoint address is published in the respective *CapabilityStatements* for the FHIR bus of the sender, receiver, and YHCR. Participants may cache *CapabilityStatement*s of their peers to improve the efficiency of endpoint discovery. The maximum cache period is specified in the YHCR operations guide.

2.6 Functional Specification for Regional Messaging Infrastructure

The following specification draws together the concepts discussed above into a specification for the regional component in the Messaging Intermediary.

1. The Messaging Intermediary operates a RESTful POST interface over HTTPS, the endpoint for which is published in the regional FHIR Bus *CapabilityStatement*.
2. The Content-Type HTTP header in the POST request must be application/json.
3. The data body of the POST request must be a *Bundle*.
4. The request is validated as follows:
    a. The content is a valid Bundle of type message.
    b. The first entry of the bundle is a *MessageHeader.*
    c. The *MessageHeader* conforms to the YHCR profile.
5. Should validation fail the service returns a 400 response.
6. The *Organisation* resource referenced by *MessageHeader* sender property is retrieved.
7. If resource retrieval fails, then then the service returns a 503 response if the failure is transient or else a 400 response.
8. If the sender organisation is not registered as a data provider with YHCR or the TLS certificate does not match the registration, then the service returns a 401 response.
9. The *Organisation* resource referenced by *MessageHeader* receiver property is retrieved.
10. If resource retrieval fails, then then the service returns a 503 response if the failure is transient or else a 400 response.
11. If the receiver organisation is not registered as a data consumer with then the service returns a 502 response
12. The message is accepted, queued for dispatch, and a 200 response is returned.
13. When attempting to dispatch a message to its destination the service:
    a. Retrieves the capability statement either from cache or from the receiver FHIR Bus.
    b. Issues a POST request on the receivers messaging endpoint containing the message in the HTTP body.
    c. If an HTTP error code is received other than 408 then the error is logged and the message is discarded.
    d. If a 408 error is received then the message is returned to the tail of message queue.
14. If an error occurs performing the function 13 then it is logged, and the message is returned to the tail of the message queue.

3 Messaging with ITK3 and MESH

MESH is a messaging technology which has evolved from the Connecting for Health Data Transfer Service. Those wishing to send data using MESH are allocated a mailbox. Data can be sent to or received from a mailbox using a RESTful interface.

ITK3 is a standard developed by NHS Digital which uses FHIR messaging over MESH to enable 4 basic messaging patterns:

1. Fire & Forget (a message is sent to a recipient, but no acknowledgement is required)
2. Technical Only (a message sent to a recipient is acknowledged)
3. Technical and Business (a message is sent and both an acknowledgment, and a business reply are returned to the sender)
4. Full Acknowledgement (a message sent to a recipient is acknowledged and a reply is sent by the recipient which in turn is acknowledged by the sender.

Patterns 2) and 4) are consistent with the approach to reliable messaging set out in this document. Patterns 1) and 3) can co-exist with the reliable messaging infrastructure by must be supplemented by YHCR provided functionality to be consistent.

ITK3 is being used (in pilot status) for transfer of care documents from acute units to general practitioners, from mental health units to general practitioners, and in handover from acute to social care (Admissions, Discharges, Withdrawals).

The YHCR obviates the need for ITK3 when both the source and target organisations are direct participants in the YHCR but has a valid role when an organisation provides data to the YHCR through a proxy service and has no direct connection point to the YHCR. This is likely to be the case for most general practises. These will, in general, be receivers of messages rather than senders of messages.

The regional FHIR Bus Messaging Intermediary will offer a routing mechanism to transmit messages over MESH to selected recipients who are identified by their ODS code. The routing mechanism will abstract the sender of a message from the transport mechanism used to deliver it.

3.1 Compatibility between YHCR and ITK3

ITK3 provides its own profile for a *MessageHeader*. The profile defines an extension property which specifies the messaging pattern to be employed. It is asserted here (without full knowledge of how ITK3 will evolve) that the details of the pattern can be derived by the Messaging Intermediary using the event type in the sender's messages. If this is the case, then the sender will be abstracted from the detail of how the message is delivered. If not then the YHCR *MessageHeader* profile will need to evolve.

3.2 Reliable Messaging and MESH

MESH and ITK3 are not designed for reliable messaging and the Messaging Intermediary may need to intervene to provide reliance to message senders of guaranteed delivery. Interventions will include:

1. Suppression of messages resent from YHCR RMS's in compliance with MESH service levels and the expectations of receiving applications.

ITK3 and MESH do not require applications to comply with reliable messaging standards. As transport layer MESH has not been designed for high-frequency replay of messages and even when an application, by convention, is able to accommodate replayed messages then the rate at which replays are transported over MESH may need to be controlled. The Messaging Intermediary will throttle replayed messages or supress replays completely on an application by application basis.

2. Regeneration of acknowledgements in response to an RMS's retried message.

   If the messaging pattern requires the receiving application to acknowledge receipt of a message and the original acknowledgement has been dispatched by the Messaging Intermediary but lost in transit then the RMS will resend messages in accordance with reliable messaging principles. The Messaging Intermediary when receiving a replayed message which has already been acknowledged must reply with the original acknowledgement.

3. Automatic generation of acknowledgements for fire & forget messages.

   When the receiving application is not required to generate a technical acknowledgement then the Messaging Intermediary must generate the response required for reliable messaging. The MESH Online Enquiry Service (MOLES) gives the Messaging Intermediary sight of the status of message delivery. An acknowledgement can be automatically generated when a fire & forget message is picked up by the receiving application.

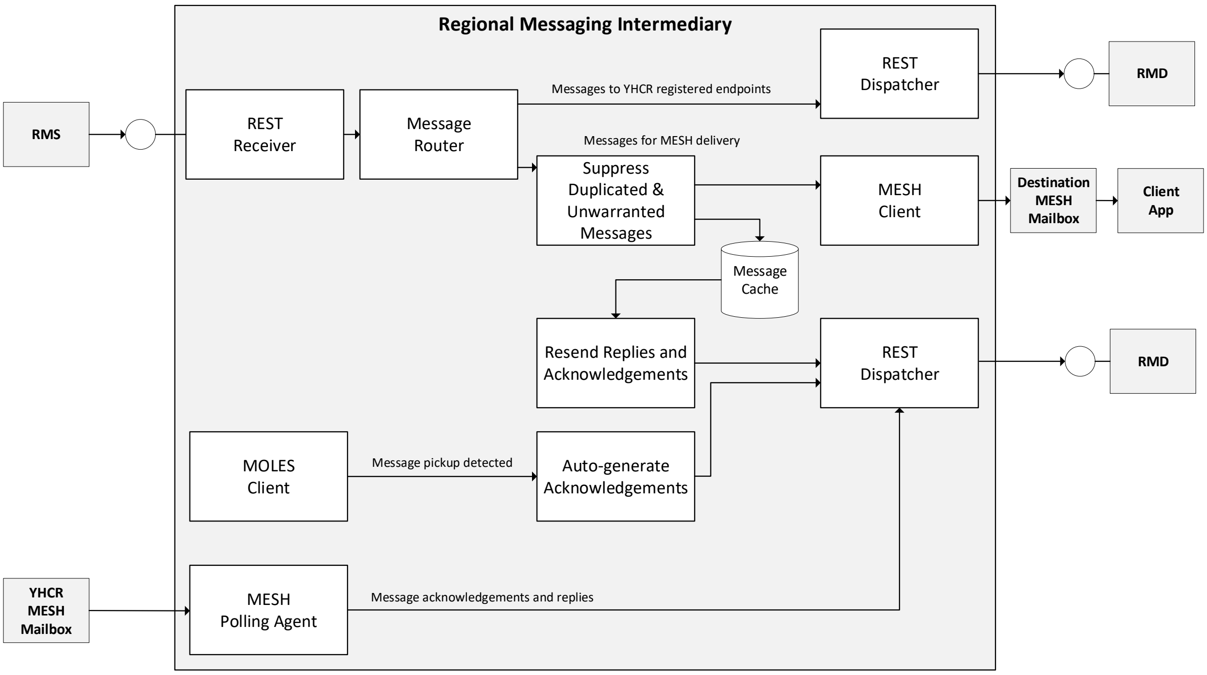
4. Resending of business replies.

   Applications which don't implement reliable messaging will not resend a business reply that has not been acknowledged. The Messaging Intermediary will resend replies to an RMD which have not been acknowledged within the expected time.

5. Suppression of acknowledgements for business replies.

   Message pattern 3 required a receiving application to provide a business reply but does not support acknowledgements for the business reply. An RMD will generate an acknowledgement and this circumstance the Messaging Intermediary will supress it.

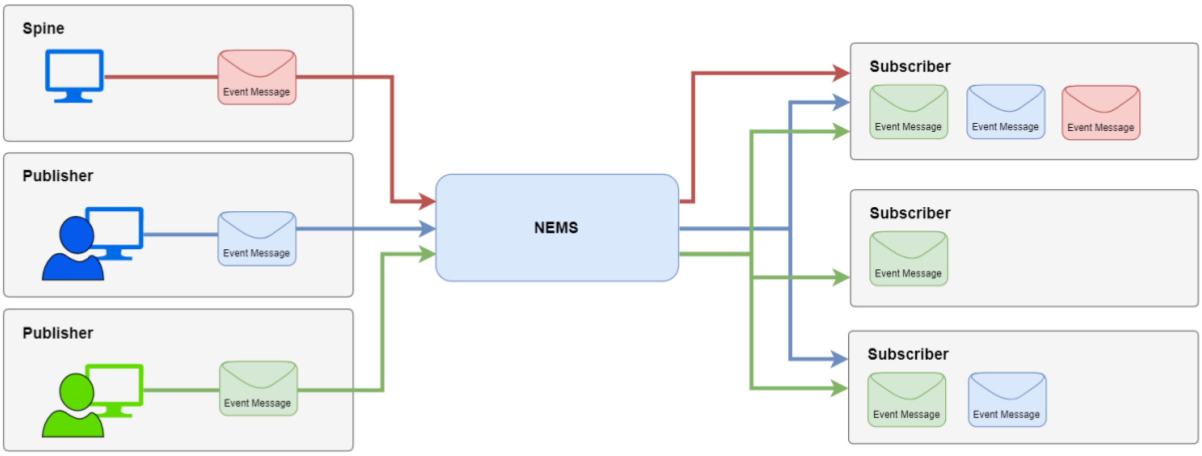
A logical design for the Messaging Intermediary routing capability is as follows:


Regional Messaging Intermediary
RMS
REST Receiver
Message Router
Messages to YHCR registered endpoints
REST Dispatcher
RMD
Messages for MESH delivery
Suppress Duplicated & Unwarranted Messages
MESH Client
Destination MESH Mailbox
Client App
Message Cache
Resend Replies and Acknowledgements
REST Dispatcher
RMD
MOLES Client
Message pickup detected
Auto-generate Acknowledgements
YHCR MESH Mailbox
MESH Polling Agent
Message acknowledgements and replies

4 Relationships with the National Event Management Service

NEMS is a service operated by NMS Digital which allows data about certain patient related events to be published to the service and for the subscribers to the service to be informed of the events as they occur.

A schematic which is used by NHS Digital to describe the service is as follows:



The service uses FHIR messaging to receive publication of events and to notify subscribers.

There is the possibility for interoperability between the YHCR Messaging Infrastructure and NEMS. Specifically, it would be possible for the YHCR Messaging Intermediary to publish data to NEMS when a message which corresponds to a publishable event passes through the regional service.

There are benefits to this approach:

- It obviates the requirement for local data provides to interact directly with NEMS.
- The technical solution for interfacing with NEMS is developed and accredited once and used by many.

There is also a possible relationship between the YHCR and the subscription element of NEMS – the YHCR enables data consumers to subscribe to data points in FHIR. Certain of these data points correspond to events which are published on NEMS. It may be possible for the YHCR to translate a local data subscription to a national event description. This concept is explored more fully in design paper 007 – “Subscription Infrastructure”.

NEMS is currently in its infancy and only supports events relating to failsafe alerting in pilot form and this paper proposes NEMS integration as a future possibility rather than seeking to direct immediate implementation.

The impact on the YHCR on supporting NEMS focuses on the Messaging Intermediary as follows:

- The first copy of a message received by MI which references a NEMS supported event is published by MI to NEMS;
- Subsequent copies (resent for reliable messaging purposes) pass through the MI without consequence for NEMS;
- AN RMS may use MI to publish an event without wishing to send the containing message to another organisation in the YHCR in which case the receiving organisation in the YHCR *MessageHeader* must be the YHCR itself.

4.1 Compatibility between the YHCR and NEMS publishing

NEMS publishes it own profile for the MessageHeader Resource which is based on FHIR STU3. The profile:

- introduces an extension property (eventMessageType);
- defines a coding system for the event property;
- constrains the destination of the message to be NEMS;
- constrains the sender to be a CareConnect *Organisation*;
- makes the timestamp optional;
- removes the enterer property;
- removes the author property;
- constrains the responsible party to be a CareConnect *Organisation*;
- removes the reason property;
- does not permit a response to be sent to NEMS;
- mandates the focus and constrains it to be a CareConnect *Encounter* or a *Communication* (also profiled by NEMS).

The NEMS *MessageHeader* can be derived from the YHCR *MessageHeader* in the following respects:

- The eventMessageType extension declares and event to be a new event, an update to an existing event, or a deletion of a previously registered event. The Messaging Intermediary would interpret messages as being new events and would derive this property;
- The event coding system required by the profile is a subset of the YHCR coding system and NEMS event codes can be derived from YHCR event codes;
- The responsible party for the message can be assumed to be the organisation sending a message to the Messaging Intermediary;
- The sender of the message to NEMS is the YHCR;
- Messages can be adjusted for downgraded cardinality by the Messaging Intermediary.

The focus of YHCR message is not constrained in a manner which is automatically compatible with the NEMS profile. An organisation sending a messaging through the YHCR for publication to NEMS must include a focus property which is a NEMS *Communication* or a CareConnect *Encounter*.

The content of a message classified as a NEMS Event must conform with the NEMS content specification.

4.2 Non-functional Requirements for Interaction with NEMS

To interact with NEMS the YHCR must be registered as a Spine endpoint and be in possession of an Accredited System Identifier (ASID).

The YHCR must comply with the HTTP requirements of NEMS such as the inclusion of a JWT claim and the presence of certain HTTP headers.