ICT Project Guidance

Data Hubs

Author:

Sky Sigal, Solution Architect

## Description

A description of Data Hubs constraints, key design decisions and approach to developing them.

## Synopsis

Data Hubs requires a baseline of capabilities to meet expectations.

## Contents

[Description 1](#_Toc145929572)

[Synopsis 1](#_Toc145929573)

[Contents 2](#_Toc145929574)

[Introduction 3](#_Toc145929575)

[Consumer System 3](#_Toc145929576)

[Consumer Identifiers 3](#_Toc145929577)

[Appendices 4](#_Toc145929578)

[Appendix A - Document Information 4](#_Toc145929579)

[Images 4](#_Toc145929580)

[Tables 4](#_Toc145929581)

[References 4](#_Toc145929582)

[Review Distribution 4](#_Toc145929583)

[Audience 4](#_Toc145929584)

[Structure 4](#_Toc145929585)

[Diagrams 4](#_Toc145929586)

[Terms 4](#_Toc145929587)

## Introduction

Background, Objective, Scope, Stakeholders [Users], Constraints, Assumptions, Risks, Deliverables.

Background

When an organisation can provide a common information service for consumption by others, it is referred to as an authoritative system of record. The sharing of information around the network is facilitated by sharing between service and all clients a common unique universal identifier(UUID) for each object[[1]](#footnote-2).

When instead data is managed by different localised systems of record, using their own identifiers, the interchange of information requires a central brokering data hub.

Issue

Without a central authoritative source of information to provide a common identifier to each piece of information, each system persists its copy of the data using locally generated identifiers.

The end result is that each system is using a different identifier than all other systems, and does not have the capability of keeping track of identifiers that may have been used by another system.

Resolution

A data hub is a central broker that persists data from one system to reshare with another system – but keeps a record of the identifiers used by each system.

  
Figure : TODO Image

### Concepts

#### Versioning

In a network where there are multiple systems of authority for individual records, there is the possibility that systems will disagree on the validity of changes made by another system.

While validation of incoming changes signals help, there still remains an ongoing issue that systems will disagree on the validity of changes made by another system.

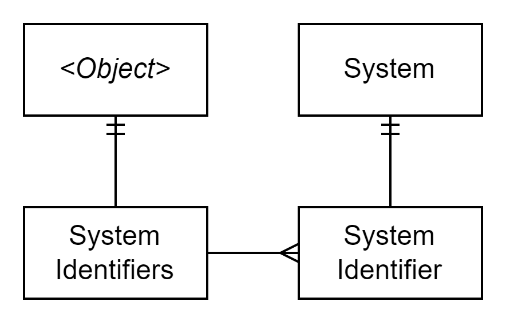
In such scenarios, it is important to have access to the multiple values ot weigh.

#### Consumer System Tracking

Each remote service call that connects to the data hub’s services is associated to a Consumer Service.

#### Consumer Identifiers

Each object that is transferred into the data hub is associated to one or more service identifiers. The service identifiers are the remote service’s identities it gave to the object in the client service, associated to a Consumer System.



TODO

Data Hubs are systems in themselves, but also share information with other systems.

In each system -- including the Data Hub itself, being a system in its own right -- items require an identifier unique to each system, which is unique to each system.

For example, in a source Student Management System (SMS) a record may have an ID of “12345”. This resource is shared with the data hub, which persists it with a different Universal Unique ID (UUID) based identifier.

When the record was shared with an.other SMS, it will be stored under another Id (“765”), and probably discard the Data Hub’s Guid based Id for the record, as it had no place to record it.

All 3 Identifiers are correct, within different contexts.

If a change to the record is made in this SMS, after it sends the information back to the data hub using the only id it knows (765), the Datahub will signal back to the first SMS, that the record has been updated, using the only Id that the first school recognises (12345).

It is the data hub’s responsibility to keep tabs on the identifiers used in each system.

Appendices

Appendix A - Document Information

### Images

[Figure 1: TODO Image 2](#_Toc144995112)

### Tables

[Table 1: TODO Table 3](#_Toc145048484)

[Table 2: TODO Table 2 3](#_Toc145048485)

### References

**There are no sources in the current document.**

### Review Distribution

The document was distributed for review as below:

|  |  |
| --- | --- |
| Identity | Notes |
|  |  |
|  |  |
|  |  |

### Audience

The document is technical in nature, but parts are expected to be read and/or validated by a non-technical audience.

### Structure

Where possible, the document structure is guided by either ISO-\* standards or best practice.

### Diagrams

Diagrams are developed for a wide audience. Unless specifically for a technical audience, where the use of industry standard diagram types (Archimate, UML, C4), is appropriate, diagrams are developed as simple “box & line” monochrome diagrams.

### Terms

Refer to the project’s Glossary.

##### Data Hub

: center of data exchange between disparate systems that adds security, data quality, value and provides it via a standard API. Differs from an operational system (as not constrained to operational information of a single system), a data warehouse (as not for integrated information), data lakes (as not for unhomogenised data).

1. Facilitated if the service client develops UUID based identifiers. [↑](#footnote-ref-2)