Specification of Data Architecture

Contents

1 Preface

This document is an output of work conducted by the IDMAPS Project at Newcastle University.[1] It specifies the Data Architecture which has been created by the project.

It has been made available under a *Creative Commons Attribution-Share Alike 3.0 License* to the wider Higher Education community in the hope that our experiences will prove useful to other institutions undertaking similar activities.[2]

Any references to third-party companies, products or services in this document are purely for informational purposes, and do not constitute any kind of endorsement by the IDMAPS Project or Newcastle University.

2 Introduction

2.1 About this document

This document provides a specification of the data architecture provided by the Institutional Data Feed Service. The term data architecture in this context refers to the processes, tools, and policies which together allow the provision of data feeds to Application Providers.

---

[1] *Institutional Data Management for Personalisation and Syndication* (IDMAPS) is a JISC-funded Institutional Innovation project which aims to improve the quality and reliability of institutional data flows. For more information, please visit the project website at http://research.ncl.ac.uk/idmaps.
[2] http://creativecommons.org/licenses/by-sa/3.0/.

This specification is structured around a description of the data feed request process, illustrated by diagrams for clarity. It also specifies how technical tools are used to enable aspects of this process.

3   The data request process

3.1   Overview of the Institutional Data Feed Service

The Institutional Data Feed Service (hereafter IDFS) is a service set up by the IDMAPS Project at Newcastle University, and provided by Information Systems and Services.[3] It exists to supply bulk feeds of accurate, first-hand institutional data to Application Developers within the University.[4] This removes the need for developers to rely upon second- or third-hand data feeds (or copies of data) which may be sourced in a non-resilient manner, be out of date, or be otherwise inaccurate.

Through the IDFS, Newcastle University has also standardised on the process by which bulk data feeds are requested and supplied to Applications around the institution, and it allows the University to maintain a detailed and accurate record of which data is supplied to whom. This assists in ensuring compliance with data protection and freedom of information legislation, as well as helping to reduce the number of instances where multiple copies of the same core data exist.

Ultimately, end users will benefit from more accurate and resilient data feeds, which underpin the applications they use on a daily basis.

3.2   Audience

The intended audience for the IDFS is Application Developers around Newcastle University. These developers typically provide a wide range of applications to end users, such as students or staff, which assist in the day-to-day running of the University. In many cases, the applications they provide rely on particular pieces of institutional data, such as names and course registration details.

Applications can include programs which meet the specific needs of a particular school, or have wider relevance to the University as a whole. Syllabus Plus (the institution-wide timetabling system) and the Reading Lists system are examples of the latter.

Depending on the requirements of the application in question, developers will in many cases have identified the need for particular institutional data. In other cases, they might be unsure of their precise requirements, and their final request might depend on which data is available already.

The data architecture has been created in a manner which allows flexibility in dealing with the diverse needs of different Application Providers. As part of the IDFS, a "data dictionary" is available for developers to find out what core data is available for them to use. In addition, the data request process has been designed so that it encourages frequent communication between the IDFS and the Application Developer, in order to ensure that the resulting data feed meets the requirements of the individual or group requesting it.

---

[3] Information Systems and Services website, http://www.ncl.ac.uk/iss/.
[4] IDFS is designed to help manage regular bulk data transfers, not one-off data transfers with limited impact.
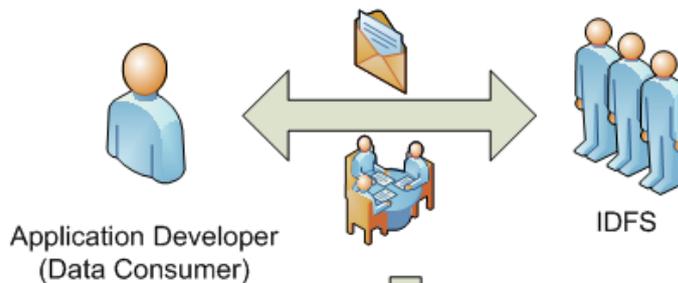
4   Requesting data through the IDFS

Requesting institutional data through the IDFS is a simple, four-step process.

IDFS Process

1. Contact IDFS

It is recommended that Application Developers contact IDFS for an informal discussion of their data requirements.

Application Developer
(Data Consumer)

IDFS

2. Complete the Data Integration Template

Every data request must be accompanied by a Data Integration Template, completed jointly by the Application Developer and IDFS.

This clarifies and documents the requirements of the data request in a standard way.

Data Integration Template

3. Implement data feed

As the various stages of the Data Integration Template are completed, the data feed is implemented, tested, and finally deployed.

Data Feed

<?xml version="1.0"

encoding="UTF-8"?>

<list>

...

4. Outputs

A fit-for-purpose data feed

The Data Integration Template, forming an up-to-date record of the feed.

Data Integration Template

Up to date record

Data Feed

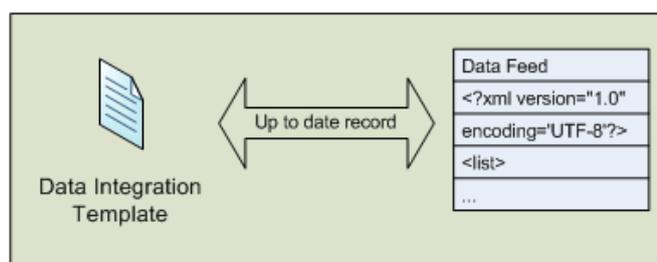<?xml version="1.0"

encoding="UTF-8"?>

<list>

...

Diagram 1. IDFS Data Request Process

4.1   Contact IDFS

Before going any further, Application Developers who may require feeds of core institutional data for their applications are advised to contact the IDFS team for an informal discussion of their requirements. An e-mail list for this purpose has been set up, and is available to all Newcastle University staff.

4.2   Complete the Data Integration Template

For each data request, a Data Integration Template must be completed by the Application Developer in conjunction with the IDFS. The Data Integration Template and accompanying Guide to Completing the Data Integration Template are available from the IDMAPS Website.[5]

The use of the Data Integration Template ensures that the Application Developer's requirements for the data feed are clearly documented. It provides structure to the data request process, ensuring that steps are completed in a logical sequence and that all requirements (including technical and compliance) are met.

The Data Integration Template is a multi-stage document: some sections are completed by the Application Provider, other sections by IDFS, as the data feed request is processed. Close co-operation is encouraged by the design of the Template, and is vital to ensuring that the process of discovering and documenting the details of proposed feed is successful.

More information about the Data Integration Template is provided in the *Guide to Completing the Data Integration Template*, Section 2.2 ("About the Data Integration Template").

4.3   Implement data feed

As the relevant sections of the Data Integration Template are completed, the data feed is implemented, tested, reviewed/monitored, and deployed.

This is carried out by IDFS, in conjunction with the Application Developer who requested the feed, so as to ensure that the feed is fit for purpose and well documented.

For most requests, existing data feeds from core systems will simply need re-purposing to provide the required feed to Application Developers. IDFS has selected the open source tool Talend, which allows data transformations to be modified with ease. In most cases, this work will be invisible to the Application Developer, who will simply be provided with the data they wanted via a data feed.

In certain circumstances, a data feed request may require additional development work by the teams who maintain the core institutional data store (SAP). In these cases, it has been agreed with the SAP teams that, having identified the requirement for additional development, the IDFS will queue the request through existing SAP work management/prioritisation channels.

This ensures that the IDFS is still the primary contact for all data requests, and allows the IDFS to maintain a comprehensive record of requests for data. Furthermore, it should reduce the number of unnecessary requests to SAP, so that only those requests with a genuine need for new SAP

---

[5] At present (September 2009), IDFS resources are located on the IDMAPS Project website. Upon conclusion of the project, they will be re-located to the relevant section of the Information Systems and Services website.

development (as opposed to requests could be met through re-purposing other flows) are passed to the relevant SAP teams.

This process is outlined in the diagram below.

IDFS-SAP Interaction

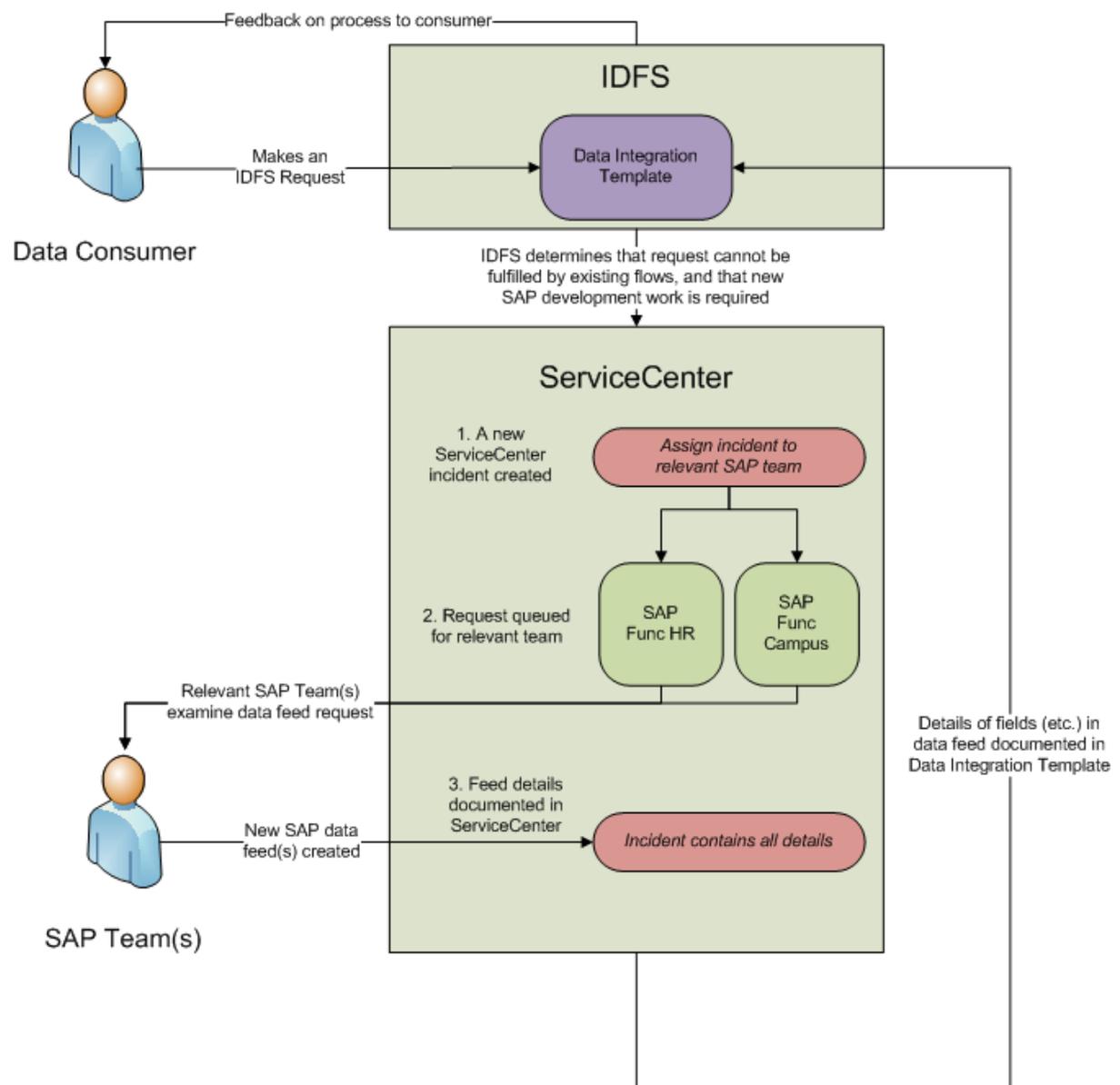## Where new SAP development work is required

Diagram 2. IDFS-SAP Interaction

4.4   Outputs

By the end of the process, the outputs should consist of:

- A reliable, fit-for purpose data feed.
- The Data Integration Template, forming a comprehensive and up-to-date record of the feed. This includes information on what data is involved, how it is transferred (and to/from which systems), for what purposes it is transferred, associated security and integrity arrangements, and contact details for the individual(s) responsible for that data and feed.

It is intended that the Data Integration Template will be a live document, maintained by IDFS and the Application Developer, which is kept up to date as any changes to the data provision are requested and implemented.

As the IDFS continues to develop and documents data feeds in conjunction with various Application Developers around the University, it will identify areas where such feeds can be consolidated and rationalised.