Template for SUM: e-Framework editorial work

Usage Statistics

The library wishes to carry out an annual review of all their electronic subscriptions. They wish to get reports offering 'per use' or 'per download'
costs for each resource, and to compare their usage with average use across
similar institutions.

Such reviews tend to be problematic due to difficulties in collecting the data, and in access to data from other institutions. There are also issues in ensuring that like for like comparisons are made. Libraries also require more fine grained data than is currently available.

A central service for reporting statistics is proposed, where libraries will send their own local data, covering times of access and course information, retrieved from other university systems. E-resource providers will be harvested for their usage information (much as they are now, except that all such harvests will be undertaken by a central service, rather than many individual institutions). The central service is charged with validating the data and reconciling the two together.

It is additionally proposed that any solution be based on the COUNTER and SUSHI standards, ideally with a sector wide application profile to ensure that like for like comparisons are possible.

Problem description

Annually, libraries collect and analyse data related to the use of e-resources over the previous year. This data is termed "usage data", and typically it will show which resources are more heavily used than others. The library is coming under increasing pressure to justify spending on e-resources, therefore it is vital that the usage data is accurate.

The difficulty in collecting usage data is that the process is time consuming and complex, and is often not at a level of granularity that is useful to the library. Librarians will be interrogating many different systems, which will present data in different ways, and then a manual process involving spreadsheets is performed to collate and sanitise the data.

Relatively crude calculations are possible to determine the cost "per use" of an e-resource – the cost of the e-resource being divided by the number of uses – but this level of data does not allow for much further analysis. It is felt that as the number of e-resources used in the library increases, more fine-grained statistics will be required.

Another difficulty is that it is not often clear if like is being compared with like. Is an "access" the same as a "download"? Some modern e-resources (the example of a fashion database with 'trends' was given, it had no full text and or download), might not operate in a "per use" kind of way. How should data from these e-resources be collected and analysed?

There is also a need to compare usage statistics with those of other institutions and currently there is no way of doing this.

Goal

It is proposed that an above campus service be created which will allow libraries to share usage data. In simple terms, data would be aggregated centrally and shared across all institutions participating in the service. This would solve part of the problem – in that librarians could see the e-resources being widely used across the sector, and assuming that there is some central QA, there would be some guarantee that like is being compared with like. An application profile of the COUNTER standard for UK usage data should be considered.

This can be realistically achieved by the use of the COUNTER standard (http://www.projectcounter.org/) and the SUSHI protocol (http://www.niso.org/workrooms/sushi). Many (although not all) e-resource providers implement these protocols, and currently librarians obtain e-resource usage data from their providers. Librarians complain that this requires login to many different systems; a central system would be able to use SUSHI to query many different e-resource provider systems and then automatically aggregate and format the data.

However, the issue of providing more granular data is not so easily solved. E-resource providers do not track the courses that students are doing. Therefore it has been proposed that local data from the library is also fed into the central service. In particular, authentication data could be used to identify users of resources; and then via lookup in other systems (for example Student Management Systems) provide data regarding the courses, study levels, time of usage and so forth.

The difficulty of this latter point should not be underestimated – there are many authentication methods in place in institutions; and it could be the case that authentication data isn't the correct way of doing this. Local data collection for the level of granularity desired is an area for further work.

Use case (Business Process Modelling)

Actors:

The actors include all parties involved in e-resource usage.

Library Users:
- Use e-resources; thus being the raw source for usage data.

Library staff:
- Responsible for providing local usage data to the central service; for non COUNTER enabled publishers.
- Use the central service data within SCONUL returns and annual library reports.
- Promote e-resources to academics and students.

E-resource providers:
- Provide COUNTER information on usage (via the SUSHI protocol) to the central service.

Central Service Staff:
- Provide QA on the e-resource provider information to ensure that the various different data fields have been implemented in a like for like manner.
- Sanity check the aggregated data.

Business Objects:

Usage Data – This is the data that the library is interested in, probably described by some application profile of the COUNTER standard. The data contains information on the usage of a particular e-resource, connected to a particular library.

Note: Locally held e-resources will need to be described to the central service using the same application profile of COUNTER.

Note: Usage Data which is augmented by Local Access Data is termed "Consolidated Usage Data".

Local Access Data – this is the granular data that the library provides to the central service. This data will need to use the same institution identifier that this institution uses for e-resource access. The local access data is assembled from the following local sources:

- Holdings Data

- Course Data – This is data retained by the institution that matches a student to a course or study level. This data will be used in combination with authentication data and data to inform the central service, with a view to adding extra granularity to the usage data.

- Authentication Data – This is data that tracks who a library user is, and the time at which they have used an e-resource.

Processes:

Template for SUM: e-Framework editorial work

Create Local Access Data

This process is concerned with providing local access data to the central service from the local library. This process will help enable the required level of granularity for the usage of a particular e-resource; as well as detail locally held e-resources.

- Library staff must obtain the authentication data to identify users and times of access.
- The authentication data needs to be used to locate the appropriate course data for the user(s) e-resources (local and remote).
- Any usage data on locally held e-resources must be converted into a COUNTER conformant structure. In particular, the same institutional identifier must be used for both local and remote e-resources.
- Local access data is constructed from the data described above.
- Usage data and local access data are sent to the central service.

Get Data from Publishers

- Central service staff harvest (via SUSHI) all of the usage data for all institutions.
- The usage data is quality assured, to check that the e-resource providers have successfully implemented the application profile of COUNTER required by the central service.

Aggregate Data

- The Central Service staff aggregate the data together, combining the usage data received from institutions and from e-resource providers, with the local access data received from institutions.
- The resulting data is validated, and transformed into an agreed format for view / download by library staff.

View Data

- Library Staff authenticate with the central service.
- Library Staff download the usage data that they require (presumably their own for internal reports, and some comparison data if required).

Functionality

Create Local Data (Add)
The staff need to undertake the work required to produce the Local Access Data, and ultimately share it with the central system.

Local Service:
- Local services need to reconcile the Holdings Data, Authentication Data and Course Data.

- Local services need to assemble Local Access Data from this reconciliation.

Central Service:
An add request needs to be made to the central service. The request contains:
- The Local Access Data business object.

Behaviour:
1. The request is validated.
2. The Local Access Data is added to the appropriate database.

Harvest Data from Publishers (Harvest)
For all institutions, the central service needs to harvest the appropriate records from the e-resource providers.
Behaviour:
1. A set of requests is sent out to e-resource providers via SUSHI to obtain the Usage Data for all e-resources for all institutions.
2. The Usage Data for the period is validated to ensure that it is compliant to the application profile of COUNTER. The data is rejected if it does not pass validation.
3. The Usage Data is added to the appropriate database.

Aggregate Data (Assemble)
For all institutions, the central service needs to reconcile the Usage Data against the Local Access Data.
Behaviour:
1. The Usage Data for a particular institution is retrieved.
2. The Local Access Data for a particular institution is retrieved.
3. A new consolidated Usage Data object is assembled from the retrieved data. This object is likely an extended COUNTER report object, including the extra information detailed course/study level and access times (for example).

View Data (Read)
A read request needs to be made to the central service. The request contains:
- The authentication details of the institution.

Behaviour:
1. The request is validated to ensure that it comes from an actual institution.
2. The aggregated Usage Data is returned to the requester.

Service arrangement
Add

Template for SUM: e-Framework editorial work

Description: This function needs to assemble Local Usage Data and upload it to the central service.
Orchestration:
- Call Assemble
- Call Add

Service Name: Assemble
Target Data Source: Authentication Database, Course Database, Holdings Database
Target Business Object: Local Access Data
Actions:
- Retrieve Data from the Authentication, Course and Holdings Databases.
- Process the retrieved Data, to link time of access (from Authentication) and desired course information to the e-resource identifiers from the Holdings data.
- Assemble Local Access Data as above for all e-resources accessed.

Service Name: Add
Target Data Source: Local Access Database
Target Business Object: Local Access Data
Actions:
- If the Local Access Data already exists in the database, then this request should be rejected.
- Put the Local Access Data into the database.

Harvest
Description: this function needs to retrieve that Usage Data from all e-resource providers. This includes institutions with locally held resources. Once the Usage Data is retrieved, it is validated and stored for later use.
Orchestration:
- Call Harvest
- Call Validate
- Call Add

Service Name: Harvest
Target Business Object: Usage Data
Actions:
- Connect to each e-resource provider and retrieve the usage data or all e-resources accessed by all users at all institutions.

Service Name: Validate
Target Business Object: Usage Data
Actions:

- The Usage Data is validated to ensure that the application profile of COUNTER has been strictly adhered to.

Service Name: Add
Target Data Source: Usage Database
Target Business Object: Usage Data
Actions:
- If the Usage already exists in the database, then this request should be rejected.
- Put the Usage Data into the database

Assemble
Description: this function needs to create consolidated Usage Data, which is augmented by the Local Access Data.
Orchestration:
- Call Assemble

Service Name: Assemble
Target Data Source: Usage Database, Local Access Database
Target Business Object: Usage Data
Actions:
- Retrieve Data from the Usage Database and Local Access Database
- Process the retrieved Data, to link Local Access Data to the Usage Data, via institution identifier and e-resource identifiers.
- Assemble and store consolidated Usage Data.

Read
Description: this function needs to validate that a request to read Consolidated Usage Data has come from a real institution. If so, and if the data exists on the system, it is returned to the requester. If the request is valid but the checklist data does not exist, then the request is rejected.
Orchestration:
- Call Authenticate
- Call Read

Service Name: Authenticate
Actions:
- The central data service needs to satisfy itself that the request received has come from a bona fide institution. The model is silent on how this is achieved. It is suggested that institutions are given simple usernames and passwords in a user accounts database.
- If the request fails authentication, the request should be rejected.

Service Name: Read
Target Data Source: Consolidated Usage Database
Target Business Object: Consolidated Usage Data
Actions:
- If this Consolidated Usage Data does not exist in the Consolidated Usage Database, then this request should be rejected; otherwise
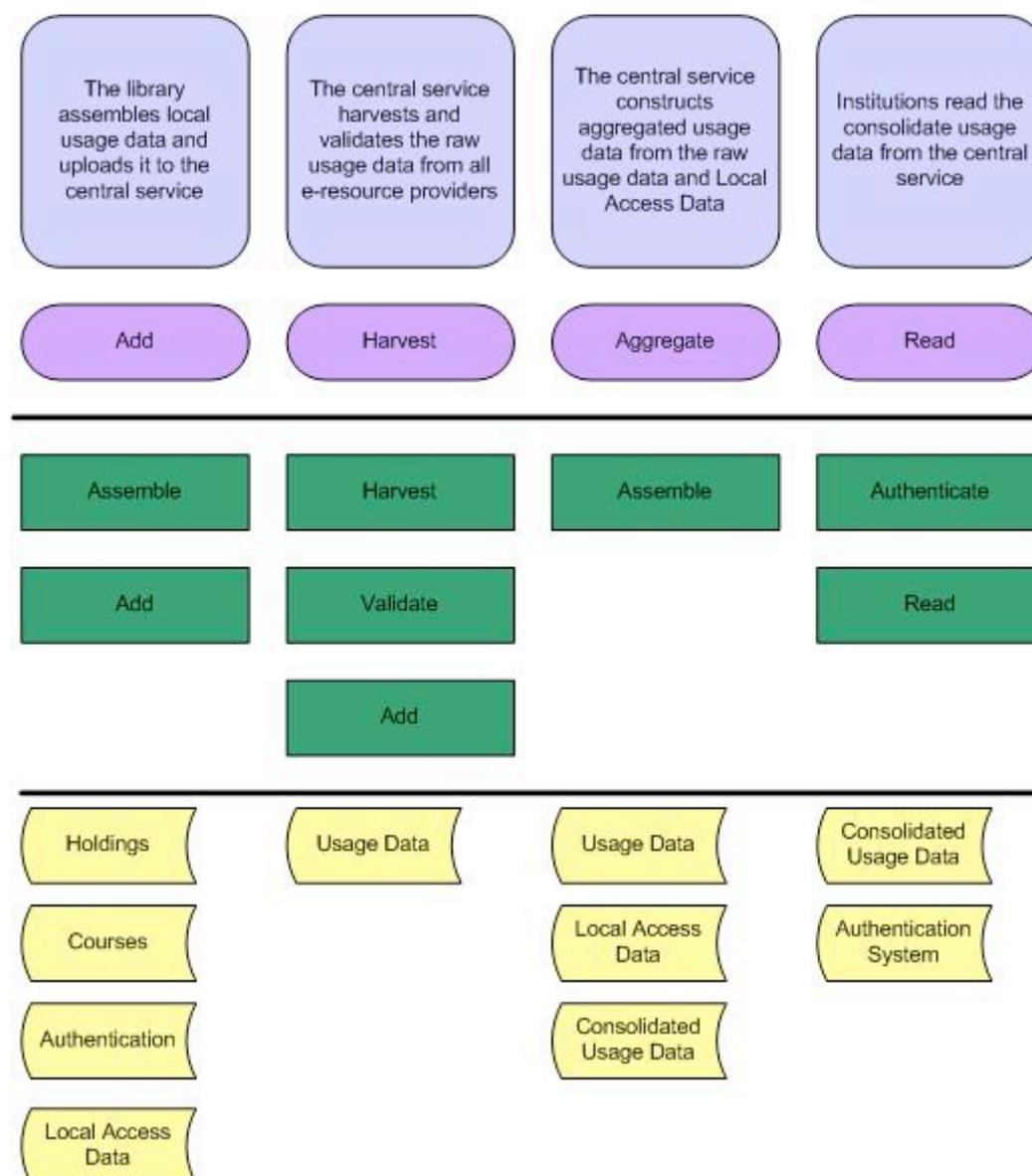- Return the stored Consolidated Usage Data.

SUM diagram

Figure 1: Usage Statistics SUM