John Tanner

The DataHub – What is it?

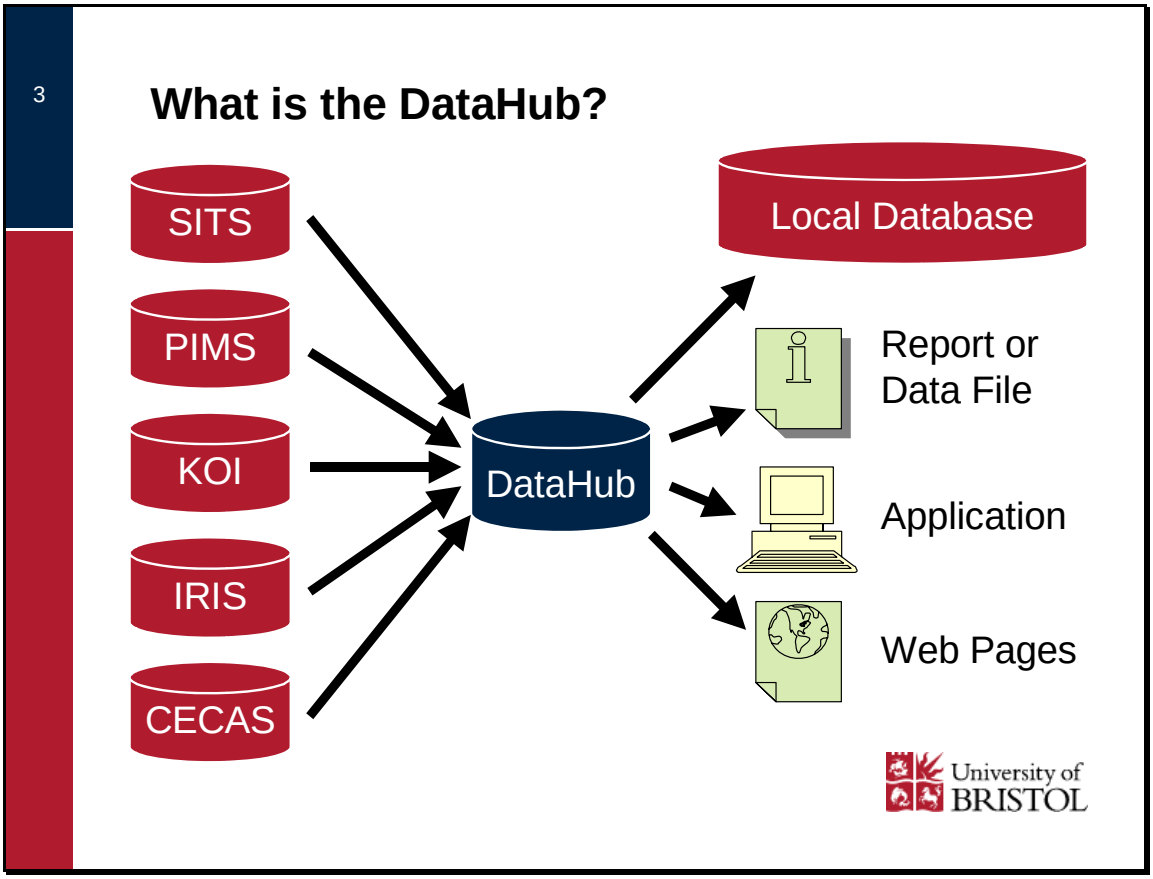
12 July 2005

University of BRISTOL

Overview

- What is the DataHub?
- Why use the DataHub?
- How to access the DataHub?
- DataHub usage
- DataHub security
- Future developments
- Questions and Answers

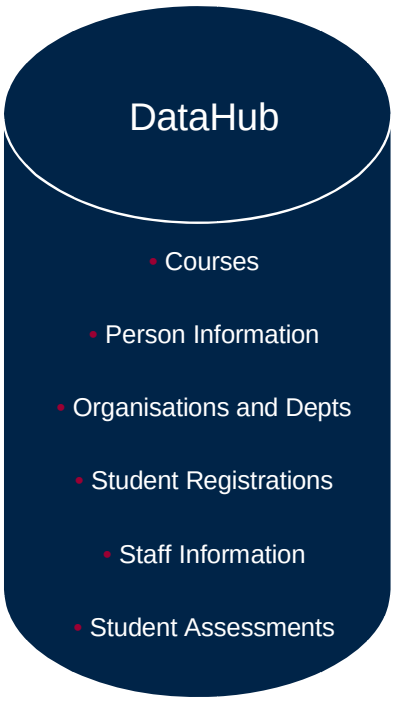
University of BRISTOL



What is the DataHub?

Read only
Oracle 9.2 Database
Data extracted from centrally administered databases
Available to all members of staff within the university including:
Central applications
Local (departmental) applications
Individual members of staff
Typical usage includes
Updating local databases
Providing data for websites and intranets
Automating IS procedures
As an application data source
Examples of applications using the DataHub include:
Contact Directory
Porpoise
Various departmental applications

What is the DataHub? (2)



Subject Areas

University of BRISTOL

Course Tables - Course information about undergraduate, postgraduate and continuing education courses. Includes unit info and course ownership.
e.g. course, route, diet, modularisation_unit, and mod_unit_availability.
Person Information Tables – Personal information (full names, contact info, dates of attendance/employment at university) for staff and students.
e.g. person_id_admin, person, address, person_organisation and cd_person_phone_no
Organisations and Depts. Tables – Name and contact data for all units (academic and non-academic) in the university organisational hierarchy.
e.g. organisation_unit, organisation_address and organisation_headship.
Student Registration Tables – Courses and units taken by undergraduate, postgraduate and continuing education students.
e.g. student_registration, person, attendance_mode, organisation_unit, course and student_registration_unit.
Staff Information Tables – University employment history and qualifications held by current/near future staff. (Often Confidential)
e.g. person, staff_appointment, staff_appointment_history and category_group_grade.
Student Assessment Tables - Data relating to student assessments. Currently only exam timetable information.
e.g. assessment, exam_paper, student_unit_mark, organisation_unit, mod_unit, mod_unit_availability and academic_year.

Why Use the DataHub

Pros:

- Central data – Common to everybody
- Easily accessible
- Direct read only access
- Masks source system changes

University of BRISTOL

Central data – Common to everybody
The DataHub provides a common data source for applications including central services and departmental systems.
The format of the data has been standardised on input, regardless of the format in the source systems.

Easily accessible
All staff members can request an account which can be access from all PC's on the universities network.

Direct read only access
Since all access is read only, there is no need to worry about making accidental changes.
All updatable information must be held in your own applications.
Direct access to the DataHub database, removes the need to load data into individual systems using file downloads (e.g. CSV files).
Often direct database access is not permitted on other source systems, other than through front end screens.

Masks source system changes
Any changes to any of the underlying systems will be hidden.
It removes the need to re-write applications using the DataHub.

Why Use the DataHub (2)

Pros:

- Less complex than source systems
- Data availability guaranteed
- Security model included

University of BRISTOL

Less complex model than source systems
The DataHub's data model is far less complex than those of the underlying source systems, making the extracting of data much easier.
Source data has been combined into one model, removing the need to access multiple systems concurrently.
Fully documented schema model and table and column descriptions.
The simpler model often improves the performance when querying data.

Data availability guaranteed
Safety mechanisms have been established to ensure that when the daily refresh of the DataHub is taking place, if a problem occurs in one of the underlying systems whereby data isn't available, the problem won't be passed onto the DataHub.

Security model included
Individual and departmental users data access is controlled, to remove the chances of inadvertently publishing sensitive data.
Further details are discussed later in the section 'DataHub Security'.

Why Use the DataHub (3)

Cons:

- Not dynamic (data refreshed nightly)
- Only limited historic data
- Volatile
- Limited summary data
- Not as detailed as source system data
- Changes may take a long time

University of BRISTOL

Not dynamic (data refreshed nightly)
As the data only refreshes once every day, it is necessary to have to wait for any changes to the source systems to become available.
The daily refresh can be extremely useful for reconciling data.

Only limited historic data
DataHub only displays details of current members of the University.

Volatile
The records are completely re-loaded each day.

Limited summary data
Data is mainly only stored in a detailed format
Very few summary tables (More planned for future use)

Not as detailed as source system data
Only commonly required data is stored within the DataHub.
Detailed information, will still come from the relevant source system.

Changes may take a long time
Before any major changes are made they must be reviewed by the project board/DataHub users forum.
Only one developer who carries out all of the changes.
The reason for being so rigid in applying changes to the DataHub is to ensure they collectively benefit all users.

How to Access the DataHub

- Obtaining a user account
- Connecting to the DataHub
- Querying the database

University of BRISTOL

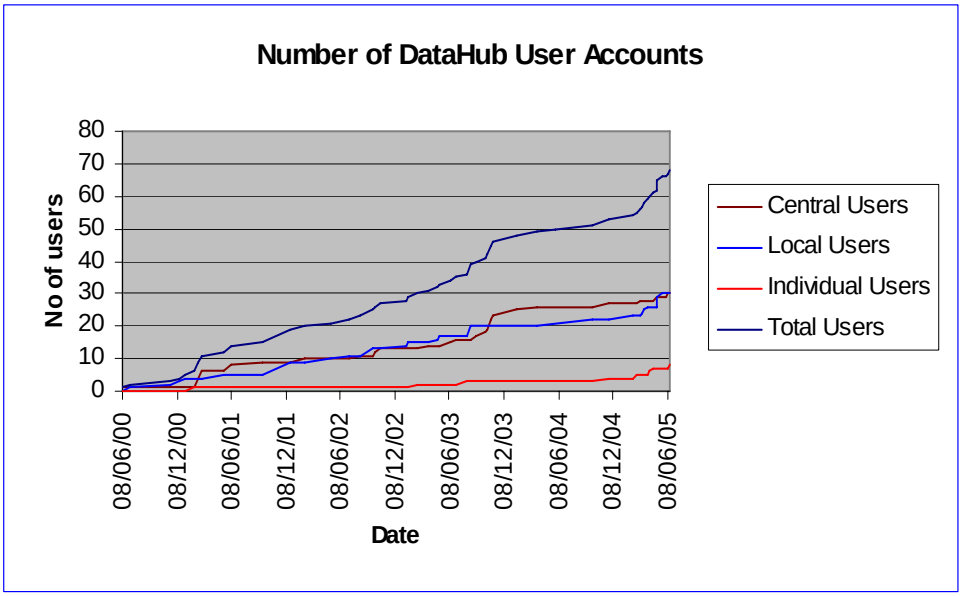
Obtaining a user account
Details of how to request an account can be found on the page:
<http://www.bris.ac.uk/is/services/corporateinformation/datahub/ new_account.html>
You will be asked to provide a few basic details about the account you need, so that the relevant type of account can be created.
The account login details will then be provided together with instructions on how to connect and view all of the schema documentation.

Connecting to the DataHub
The DataHub can be accessed in a number of ways including:
Direct connection to the Database, 2. Via an ODBC connection,
3. Using Zope/JDBC, 4. With Perl DBD via Oracle DBI driver

Querying the database
Some SQL skills are required in order to be able to extract information from the DataHub.
Schema details are available on the DataHub web-site, which should enable the rewriting of queries.
No formal DataHub training is available, however a database course is available to staff using Microsoft Access.
For assistance in extracting information from the DataHub contact the DataHub team via the e-mail *help-datahub@bristol.ac.uk*.

9

DataHub Usage

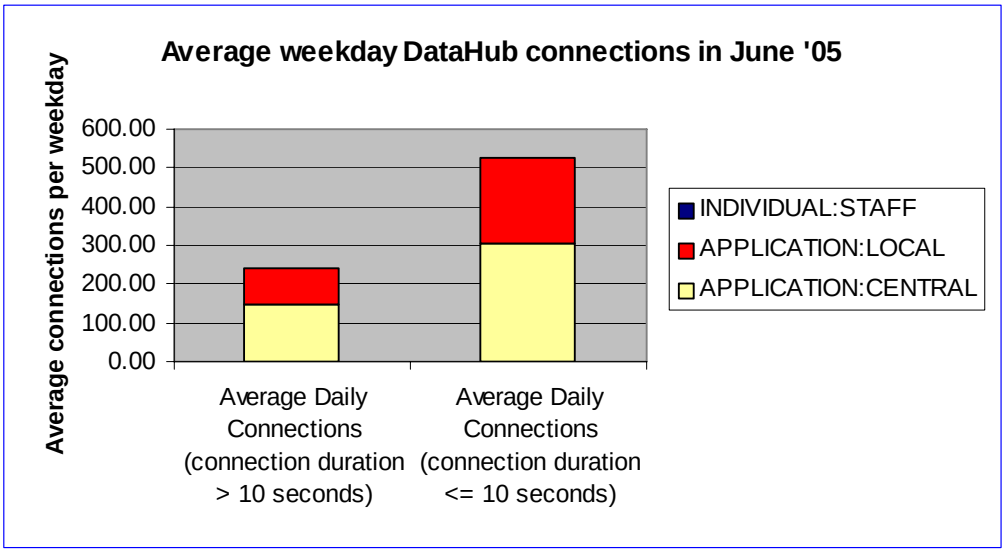


University of BRISTOL

Number of DataHub User Accounts
The graph shows a steady increase in the number of DataHub users since its launch in 2000. With the release of version 2 in July 2005, this number is expected to increase further.

10

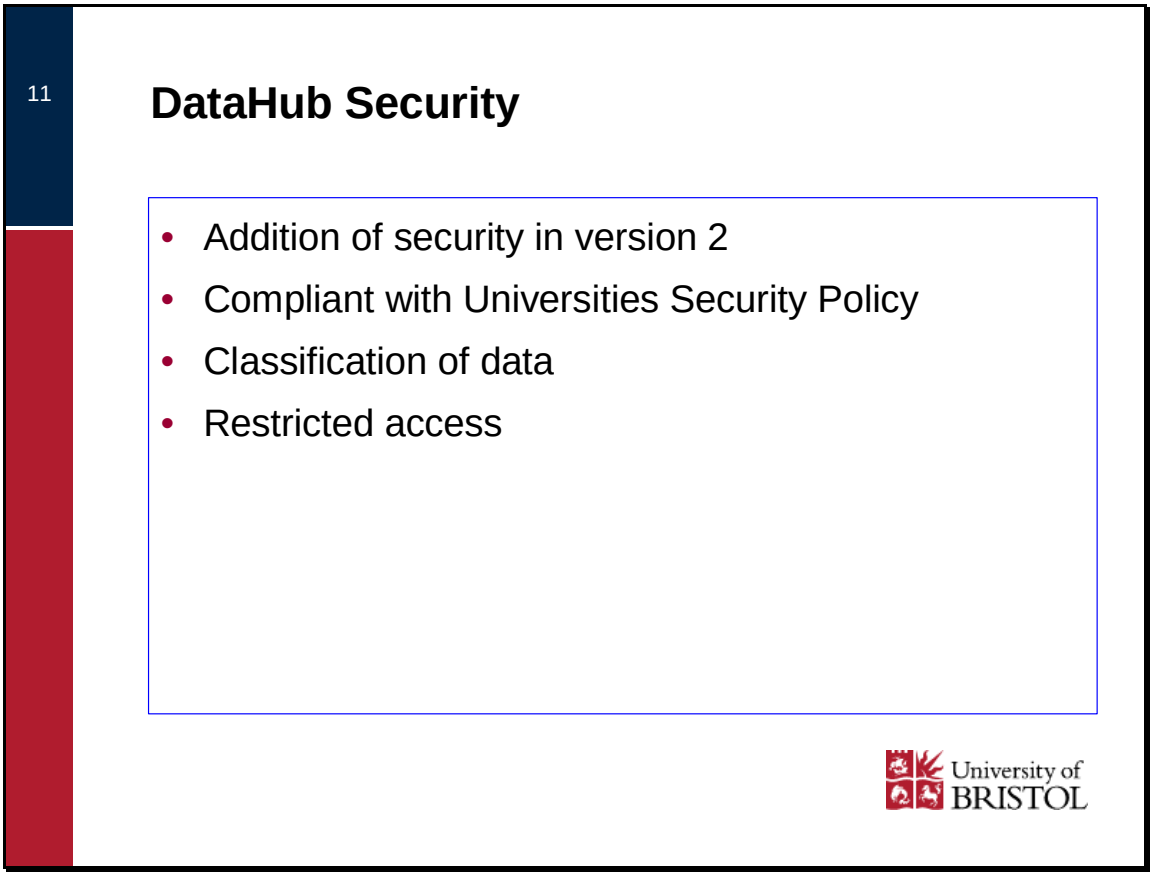
DataHub Usage (2)



University of BRISTOL

Average weekday DataHub connections in June 2005
This chart is divided into two columns to show the different nature of the connections to the DataHub:

Connections <= 10 seconds These connections relate to applications, which are written in such a way that they create a new connection to the DataHub each time some information is required.
Connections > 10 seconds These connections relate to applications, which keep a permanent connection to the database open. During each of these single connections a large amount of information could be obtained from the DataHub.

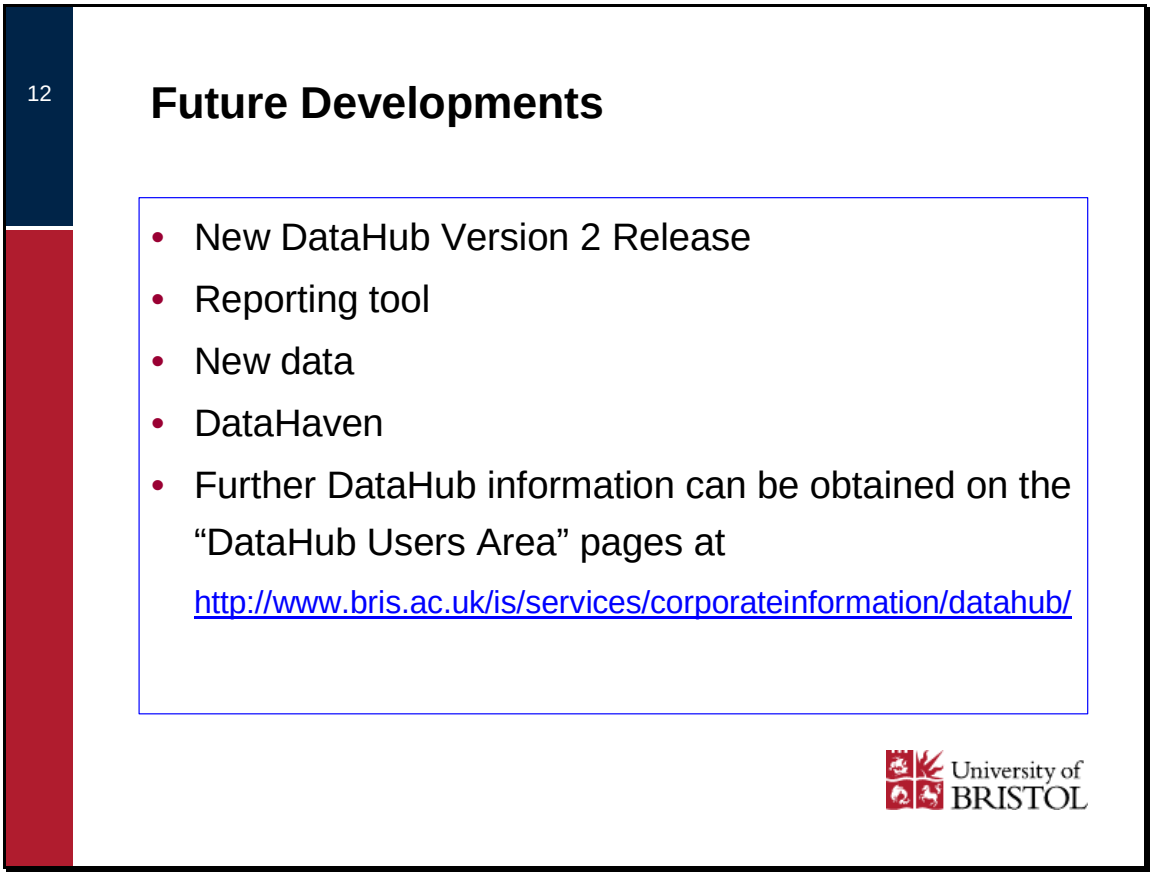


Addition of security in version 2
A new security model has been implemented in the DataHub version 2

Compliant with Universities security policy
The new version of the DataHub contains security that complies with the Universities Information Access and Security Policy.
Further details of which can be found on the page:
<https://www.bris.ac.uk/WorkingGroups/CITG/IASPolicy.pdf>

Classification of data
Each piece of data within the DataHub has a classification level attached controlling who can access it.
These are either Public, Open, Confidential or Strictly Confidential.

Restricted access
Access to data is restricted based upon which of the following user types is assigned:
Application Central
For corporate IS applications e.g. the Portal
Application Local
For departmental applications
Individual Staff



New DataHub Version 2 Release
A new version of the DataHub was released yesterday. It's new features include:
A new security Model
Updated framework to enable growth in the size of the DataHub
Many new fields
Improved documentation

Reporting Tool
Oracle Discoverer
Work starting on student data first then staff data
Roll-out in the Autumn

New Data
There are plans to add more data to the DataHub

DataHaven
Proposed writable version of the DataHub
Work scheduled to commence shortly, starting on a single subject area of the DataHub.