Supporting Data Management Infrastructure for the Humanities (Sudamih):

Database as a Service (DaaS) : A Tool For Researchers

Friday 21 January 2011

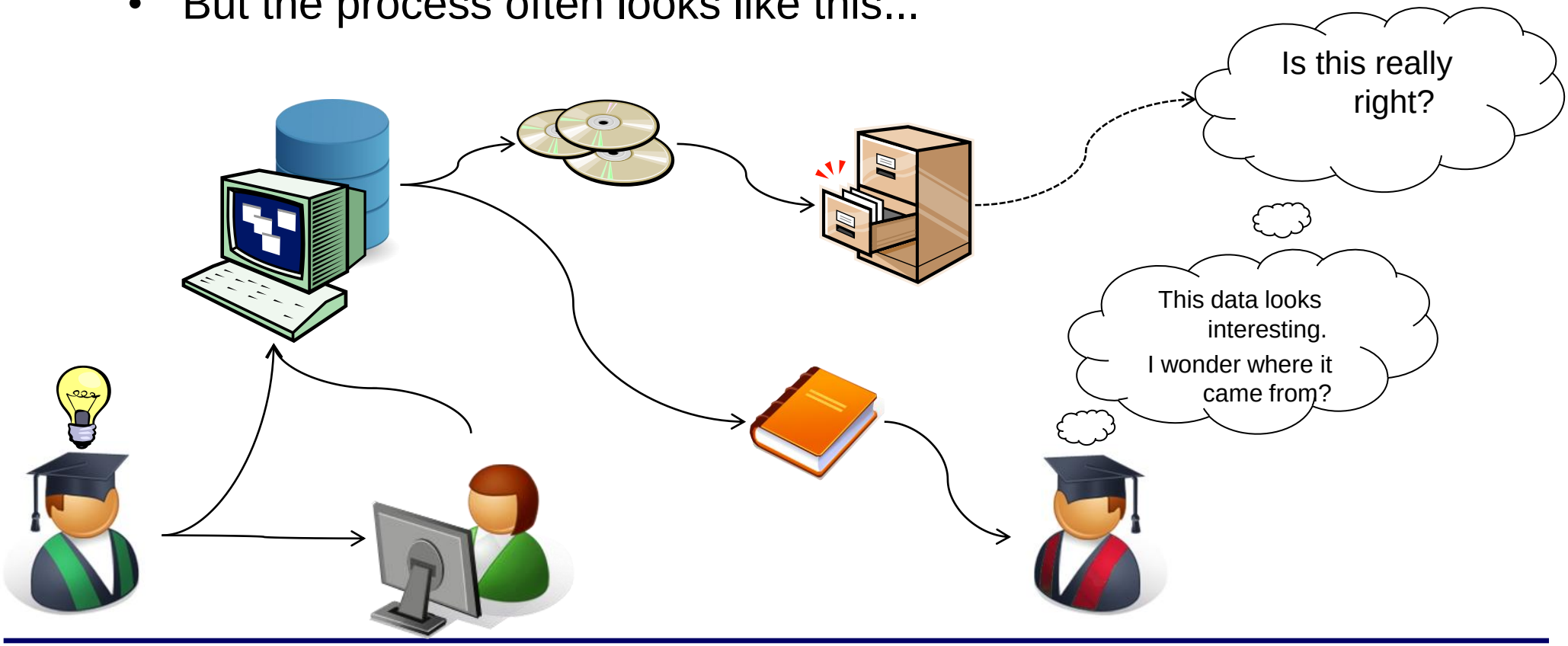
James A J Wilson
James.wilson@oucs.ox.ac.uk

What is Sudamih?

- Part of a programme to develop infrastructure for research data management at the University of Oxford
- Three significant areas of activity
  - Understanding existing data management practices in the humanities
  - Training for researchers
  - Development of a ‘Database as a Service’
- Project runs until end of March
- Long-term intention is to extend outputs to other disciplines

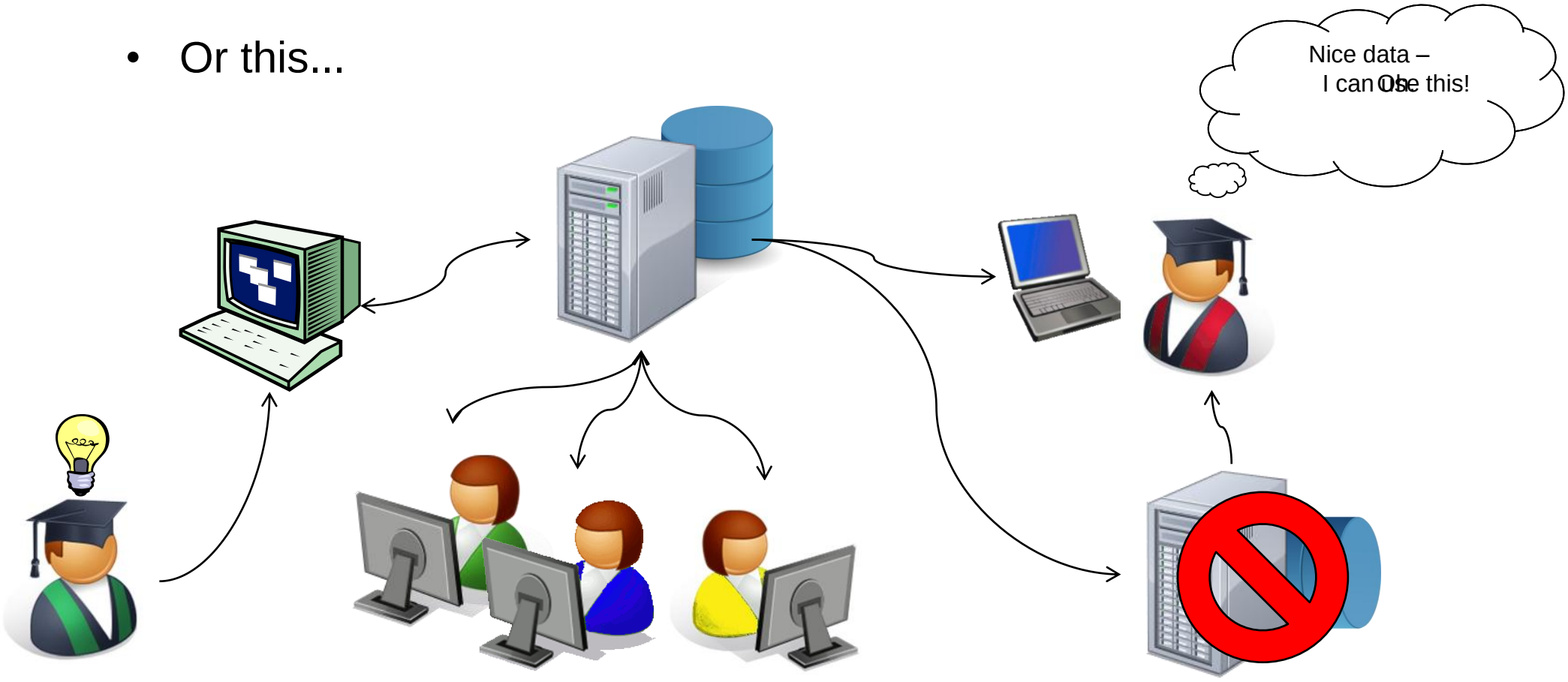
Databases in the humanities – current practices

- Database practices vary greatly in the humanities
- But the process often looks like this...



Databases in the humanities – current practices

- Database practices vary greatly in the humanities
- Or this...



Problems identified

- Lack of technological awareness
- Poor backing-up practices
- Collaboration difficult
- Difficult to re-discover and re-use data
- Risk of technical obsolescence
- Servers and Websites costs money
- Funding only lasts as long as the project, but servers and websites require maintenance
- Technical expertise required
- What happens to data when its creators move?

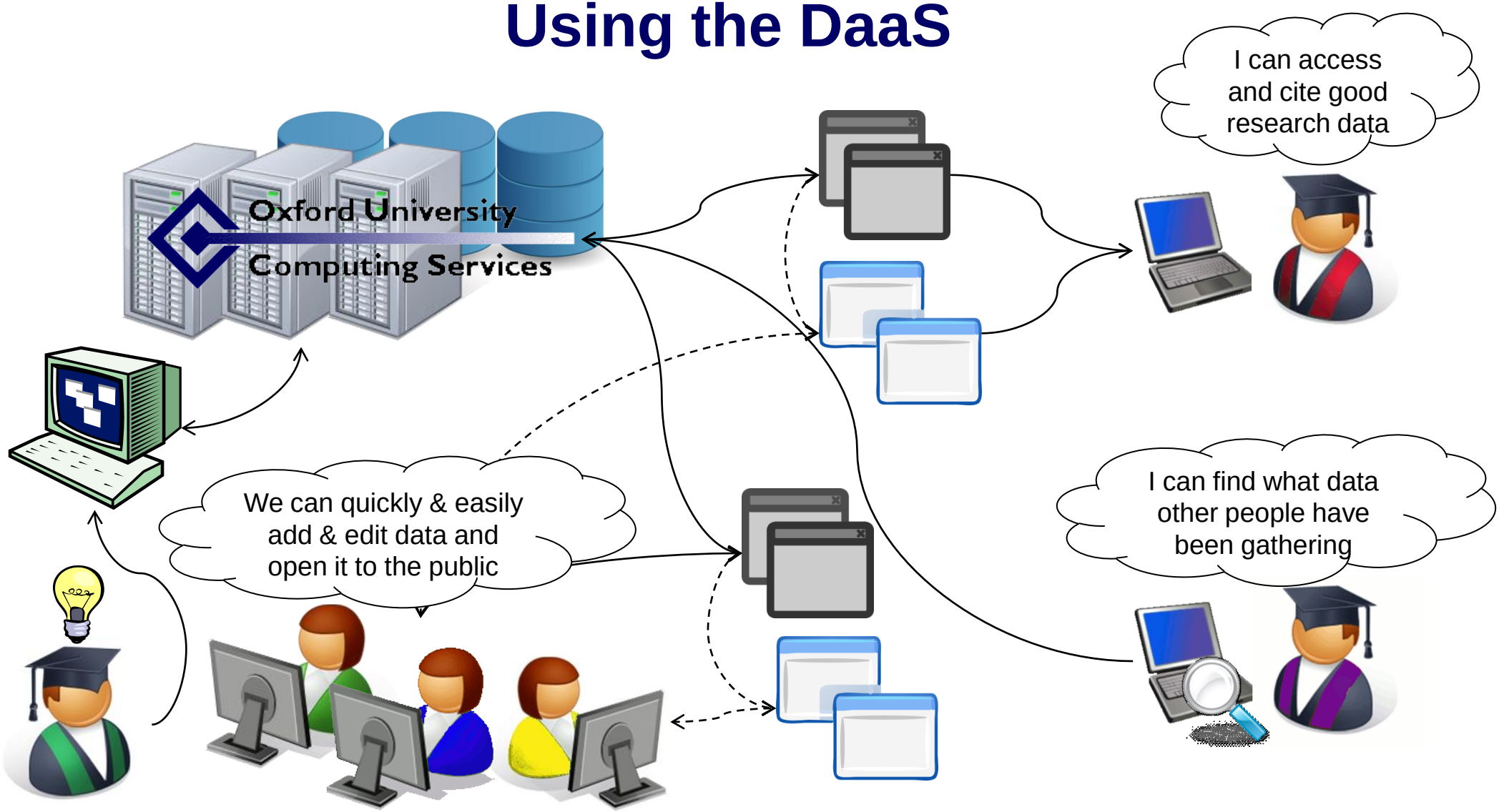
Solutions Proposed

- Centrally-provided database system with institutional support
- Databases hosted by institution, with automated secure back-up and version control
- Web-based interface to facilitate collaboration
- Simple metadata captured to describe each database, plus an interface for searching this metadata
- Technical updates rolled out centrally
- Costs of service built into funding proposals
- Very long-term hosting with fixed PURL/DOI addresses
- Possibly national cloud-based hosting

What is the DaaS?

- A web-based system that will enable researchers to quickly and intuitively
  - build a relational database from scratch, or
  - Import an existing database in common formats (such as Access)
- Offering support (initially) for text, image, and geospatial data
- Databases centrally hosted and maintained
- Access controls to determine who can view or edit each database
- Research groups may, if desired, develop their own Web front-end interfaces to databases hosted by DaaS

Using the DaaS



Initial features

- Import / export databases in common formats
- Visual interface for editing database structure
- Admin interface for registering and managing databases
- Access controls for adding/editing data
- Visual form-builder for creating online data editing forms
- Full Unicode compliance (allowing non-Roman character sets)
- Generic query-builder tool
- Various query response formats
  - Including geospatial data plotted on maps
- Facility to add metadata to describe databases

DaaS Technical Information

- DaaS hosted on JBoss Application Server 5.0.
  - Web Portal uses open source frameworks JBoss Seam in conjunction with Facelets; AJAX support-based RichFaces
- Hibernate framework used as bridging technology between DaaS Portal and backend database
- DaaS supports various data protocols and formats
  - Researchers can upload Microsoft Access Database (.mdb) as it is or data in CSV, XML or SQL format
- MySQL used for admin front-end
- PostgreSQL back-end
- Entirely open source

Future possibilities

- Full productization and scalability
- Registry service for descriptive metadata
- Extended functionality beyond the humanities
- Incorporation of data storage models other than relational databases
  - e.g. XML-based, document-based, and CouchDB
- Deployment as a cloud-based software service
  - Involves advance monitoring and management tools
  - Capable of running on other institutional virtual infrastructures
- End-user documentation and support

Questions?

http://sudamih.oucs.ox.ac.uk/