

ORDSProposal

Abstract

ORDS is a user and project centered, web-based database-as-a-service (DbaaS) system with data publishing.

Proposal

ORDS imports data from recognized sources such as csv files, spreadsheets, and relational database files into structured online databases to provide DbaaS functionality. The databases are stored in an RDBMS (PostgreSQL with ORM access). It has a full set of high-level tools for managing and editing these databases, including tools that give users the ability to create versioned snapshots of data and share specific datasets. New databases can also be created using these tools. Databases are housed in projects and associated with users with given roles. The data can be exported in common formats.

Background

The ORDS system was developed, deployed and utilized at the University of Oxford as an online tool for researchers working with relational databases. It offered secure hosting for the data with sharing and editing interfaces to allow researchers to work together with colleagues and collaborators to collect, organize, manage data then to cite research data for publication audit. The original instigation of the project was in response to requests for a DbaaS system from university researchers and lecturers.

Rationale

The team believes strongly that the tools developed within ORDS could relatively easily be generalized for the wider community; not only the research community but also anyone who has need for centrally managed data.

There are several underlying principles that should be of interest to the wider community:

- The database is generated from the data being inputted, so whether the data has been collected in a spreadsheet or a standalone database it can be imported into ORDS and fully integrates with the features and tools of ORDS
- The data is preserved in a managed space
- The data is citable in a controlled way, i.e. snapshots and subsets of the data can be defined as publishable
- All transactions on the data are recorded and are therefore auditable in an easy to manage way.
- User roles and containment for the data currently referred to as projects make group management, control and sharing of the data possible.

While there are larger DBaaS projects such as [OpenStack Trove](#), ORDS provides a relatively straightforward solution for the “long tail” of “Small Data” commonly found in organisations.

Current Status

ORDS is currently a tool being used by researchers at the University of Oxford. The same team is also maintaining the source in an SVN repository.

Information about the service as it stands is available at [+http://ords.ox.ac.uk/+](http://ords.ox.ac.uk/)

Meritocracy

There is a strong belief in the meritocratic model of software development within the team. Two of the initial committer list are already committers on Apache projects and have experience of the Apache way of working. The intention is to open-source this software and an important part of doing that is adopting a model which ensures that it is done in a realistic and sustainable way. It is the belief of the initial team that the Apache Incubator is the place to encourage the development of this project and to realize its potential. The team believe that only by encouraging the community of developers for the project can it exist as sustainable piece of software.

Community

There are two areas of potential growth for the community of users and developers. There is growth in terms of its existing use as a research database system and also growth into other industries. Both options will be looked into. Much of the initial funding to develop the software came from JISC (<https://www.jisc.ac.uk/>), who are keen to encourage other UK universities to evaluate and adopt the ORDS. Work has already been done promoting it's use in other universities through the JISC Data Spring initiative which has generated interest from institutions.

Expressions of interest have been received from the following institutions:

- University of Bristol
- University of Southampton
- University of Leicester
- University of St. Andrews
- University of London Computer Centre
- University of Lincoln

To broaden the reach of the system a possible avenue is to target companies that help people move to open source database systems, for instance companies who help with migration them from local Access databases to open source dbms.

Code

The code base is written in Java and is managed and compiled using Maven. It uses Tomcat as its application server. It uses PostgreSQL as the backend RDBMS.

There is an existing deployed system of the software which has its own maintenance schedule, so we recognise that we would have to maintain updates of the live system whilst the apache version is not released. This would mean that some parallel development would be required until full migration to the apache version is possible. There is support from the system administration team for deployed system for transitioning to an apache release.

We are familiar with the release process and recognise that there will be some changes required in the dependencies. We have already done some due diligence on dependencies to try and smooth this process.

Core Developers

The core developers are from the University of Oxford and partners of CETIS LLP. The Oxford developers are employed partly to support this software and its use in Oxford on an ongoing basis. However, in order to make ORDS sustainable we need more users and more customers and the team are fully committed to this and are actively encouraging the systems use and other collaborators in its development from that user base.

Alignment

ORDS was created as a service in the Higher Education Sector in the UK funded by JISC. The ORDS team believes that the move to a properly managed open source solution would be facilitated by the use of a formalized 'incubator'. This would provide the project with a tried-and-tested governance structure, set of processes, and a model for due diligence.

Known Risks

Orphaned products

ORDS will not be orphaned. The primary backer to this project currently runs it as an ongoing university wide service with support and maintenance. It is not envisaged that this would stop. In fact they wish to make this software available to the wider community. This is also strongly encouraged by the original funders of the project.

Inexperience with Open Source

Two of the developers designated as committers on this project already have a history of supporting open source through their involvement as committers with other Apache projects, namely Apache Wookie and Apache Rave. All developers also have a history of commitment to open source ideals in their use of open source repositories for all previous project work.

Homogenous Developers

Initially this will be collaboration between the developers at the University of Oxford and two of the partners of CETIS LLP.

Reliance on Salaried Developers

Whilst the Oxford developers are salaried developers there is a commitment to the open source philosophy and also the continuation and growth of ORDs. The CETIS partners are contractual developers, but already have a history of supporting open source development with Apache through changes in their own roles and circumstances. Initially we will seek new developers, designers, testers and documentors from the institutions who have expressed an interest in using the service. As stated above we already have several interested parties who wish to use ORDs for their own researchers.

Relationships with Other Apache Products

Several components of the current code base rely on apache commons. As part of the incubation process the team would be looking to move the database access to OpenJPA and user authentication to Shiro. It may also be useful to utilize Wink for REST services though this would probably be at a later date.

One of the main design goals for the project is to virtualise Microsoft Access databases, which also aligns it with Apache Poi.

As an SQL database project it may also find alignment within Apache DB alongside DDLUtils and Derby.

Excessive Fascination with the Apache Brand

Apache has a recognized track record of supporting and encouraging the growth of open source projects. The attraction for the ORDS team is in supporting the ongoing development and growth of the ORDS project recognizing the Apache have a solid foundation for doing this because of its tradition of upholding good open source practice and philosophy.

Documentation

Developer and Administration documentation is to be created as part of the incubation process.

Currently overview documentation can be found at: <http://ords.ox.ac.uk>

The ORDS user documentation can be found at: <http://ords.ox.ac.uk/documentation.xml>

Initial Source

The initial source is comprised of code written by IT Services at the University of Oxford. It is currently hosted in house on a university SVN server. Work is being undertaken to put all the source under apache 2 licenses, this includes the due diligence on dependencies already mentioned and making sure correct license notices are included in all sources.

Source and Intellectual Property Submission Plan

Work is currently being undertaken preparing the source for open source distribution making sure the build system operates in multiple environments. The IP rights to the code can be transferred from Oxford University (donated) to Apache at the outset or when required - this has already be agreed with the university.

External Dependencies

Tomcat

Maven

OpenHMS Jackcess

Apache Commons

Slf4j

Hibernate – to be replace with OpenJPA?

Postgresql

JQuery

Spring Framework

Jetty

Amqp-client

HttpClient

Logback-classic

Far32-lib

Tigra_calendar

[RedQueryBuilder](#) (EPL)

[JsPlumb](#) (MIT)

...

Required Resources

- Mailing lists
 - ords-private (with moderated subscriptions)
 - ords-dev
 - Subversion Directory
- Subversion directory
 - <https://svn.apache.org/repos/asf/incubator/ords>
- Issue tracker
 - JIRA ords (ORDS)
- Wiki
 - <http://cwiki.apache.org/ords>

Initial Committers and Affiliations

- David Paine, University of Oxford
- James Wilson, University of Oxford
- Meriel Patrick, University of Oxford
- Scott Wilson, CETIS LLP, apache id - scottwilson@apache.org
- Kris Popat, CETIS LLP, apache id - krispopat@apache.org
- Atri Sharma, Data Systems and Insights - Barclays

Sponsors

Champion

TBD

Nominated Mentors

TBD

Sponsoring Entity

Incubator