CAA 2009 Williamsburg

The VERA Information Environment

Hugo Mills[1] and Mark Baker[1]

[1] h.r.mills@rdg.ac.uk & mark.baker@computer.org. ACET Middleware Group, School of Systems Engineering, University of Reading, UK.

Abstract

We present three components of a virtual research environment developed for the ongoing Roman excavation at Silchester. These components ─ Recycle Bridge, XDB cross-database search, and Arch3D ─ provide additional services around the existing core of the system, run on the Integrated Archaeological Database (IADB). They provide, respectively, embedding of legacy applications into portals, cross-database searching, and 3D visualisation of stratigraphic information.

Key words*: 3D visualisation, archival search, portals, and virtual research environment.*

1 INTRODUCTION

The VERA project is a cross-disciplinary collaboration between archaeology and computer science to develop components of a Virtual Research Environment (VRE) for archaeology. The archaeological excavation of the Iron Age and Roman town at Silchester[1], Hampshire (Calleva Atrebatum), which is one of the largest and best preserved sites of its kind in the UK, has been underway since 1997. The Silchester Town Life Project is concerned with a diachronic study of a large sample of the Roman town through excavation and subsequent analysis of all the associated finds. The principal aim is to characterise the changing nature of urban life from origins in the pre-Roman period in the first century BC to its demise in the early medieval period in the fifth/sixth century AD.

One major aspect of the VERA project was to investigate methods of acquiring data electronically in the field, streamlining the overall process of the excavation, and allowing the post-excavation analysis to start much earlier. This work has been discussed elsewhere[2]. Beyond the data acquisition elements of the project, however, there are several other strands of work within VERA, which concentrate on the aspects of the *analysis* of the data. It is these aspects that together form the VERA information environment, and which we will discuss within this paper.

There are three main software components which have been developed under the VERA project: First is the *Recycle Bridge*, which is an infrastructure tool used to assist in the deployment of legacy web applications within a portal environment. Also there are two further tools, the *XDB* cross-database search and the *Arch3D* visualisation tool, which offer alternative ways of accessing and viewing the data from the excavation.

2 RECYCLE-BRIDGE

1 Silchester Web Site, http://www.silchester.rdg.ac.uk/

2 O'Riordan, Emma Jane, Clarke, Amanda Sarah and Fulford, Michael. "Managing Change: Introducing Innovation into well-established systems." 37th Annual CAA, Williamsburg, VA, 2009.

The Integrated Archaeological Database (IADB) is a PHP-based web application for managing data from archaeological excavations. However, the IADB has limitations in its supported authentication mechanisms, and also has a number of known security issues. In order to address these security issues, we use a single-sign-on process through a standard JSR-168 portal environment.

Due to the nature and structure of the IADB/Silchester code, including the extensive use of IFrames, a new portlet bridge has been developed, called the Recycle Bridge[3], which is JSR-168 compliant. The idea behind the Recycle Bridge is that the source code of the original Web applications does not need to be fundamentally changed, so that this means that the code tree is not forked and that a project does not need to support and maintain the code that has been inserted in the portlet. The only changes necessary when inserting a Web application into Recycle Bridge is to provide Single Sign On (SSO), so that a portal user does not need to firstly login to the portal, and then login to the Web application. The Recycle Bridge uses an IFrame to display the embedded application inside the portal and a cookie to share authentication information between the portal and the application (see figure 1). It is necessary to write an authentication plugin or patch for the application that can read a cookie (and suppress the applications login screen). The Recycle Bridge has been tested not only with the Silchester/IADB code, but also with MediaWiki and Wordpress (both have APIs that can be used to integrate a security plugin). The VERA project have worked with the OSS Watch project to package up the Recycle Bridge and release it under the correct open source license (GPLv3). There has been significant interest in the Recycle Bridge, in particular by the other VRE 2 projects, and other groups developing portal-based systems including a large NHS project (NIHR Portal).

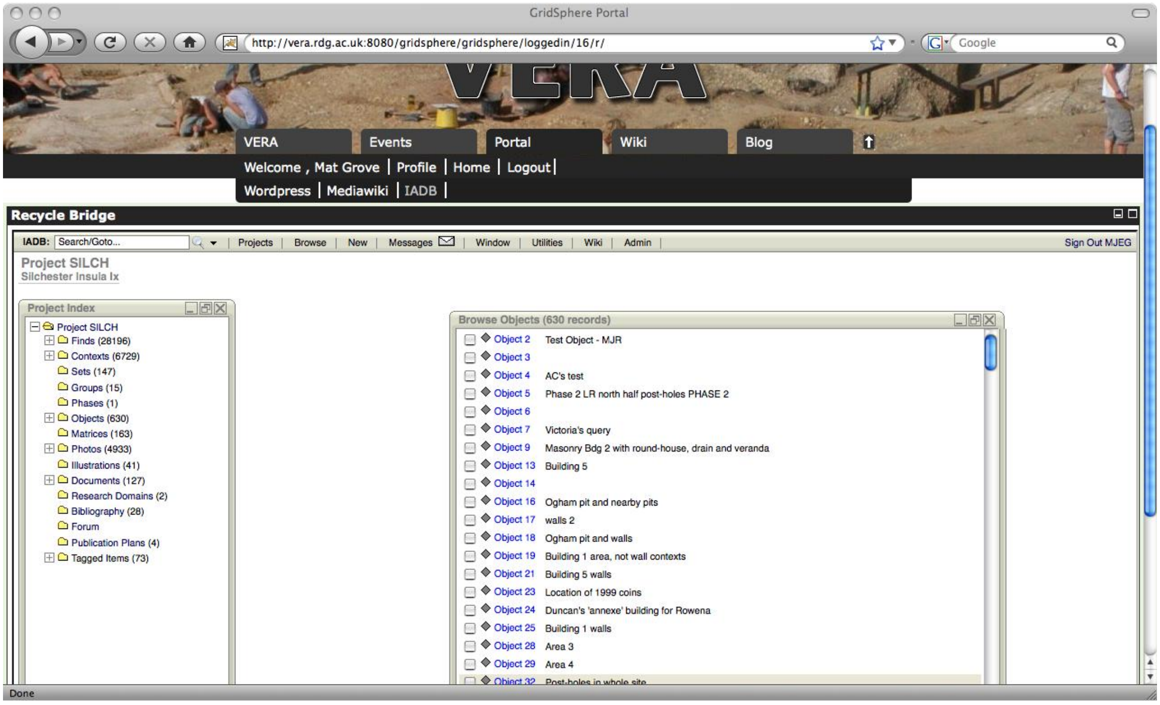


Figure 1. The Recycle Bridge used in GridSphere showing the IADB

3 XDB CROSS DATABASE SEARCH

The XDB service is a system for doing cross-database searches. It aims to create a generic and easy to use Web-based system (see figure 2) that can be utilised by various communities to search through multiple

3 Recycle Bridge, https://vera.rdg.ac.uk/tools/trac/portlets/

existing databases and potentially find matches between the artefacts or finds being studied. For example, it may be the case that an archaeologist has a piece of pottery with a particular stamp or graffiti mark on it; from their perspective it would be useful to gather more information about the stamp or graffiti, to help date the pottery, identify who made it or verify where it was produced. In order to do this, it is entirely possible that multiple databases would have to be searched, with many different user interfaces and schemas. The XDB offers database owners a system infrastructure which allows them to easily expose their data sets for search through a common interface.

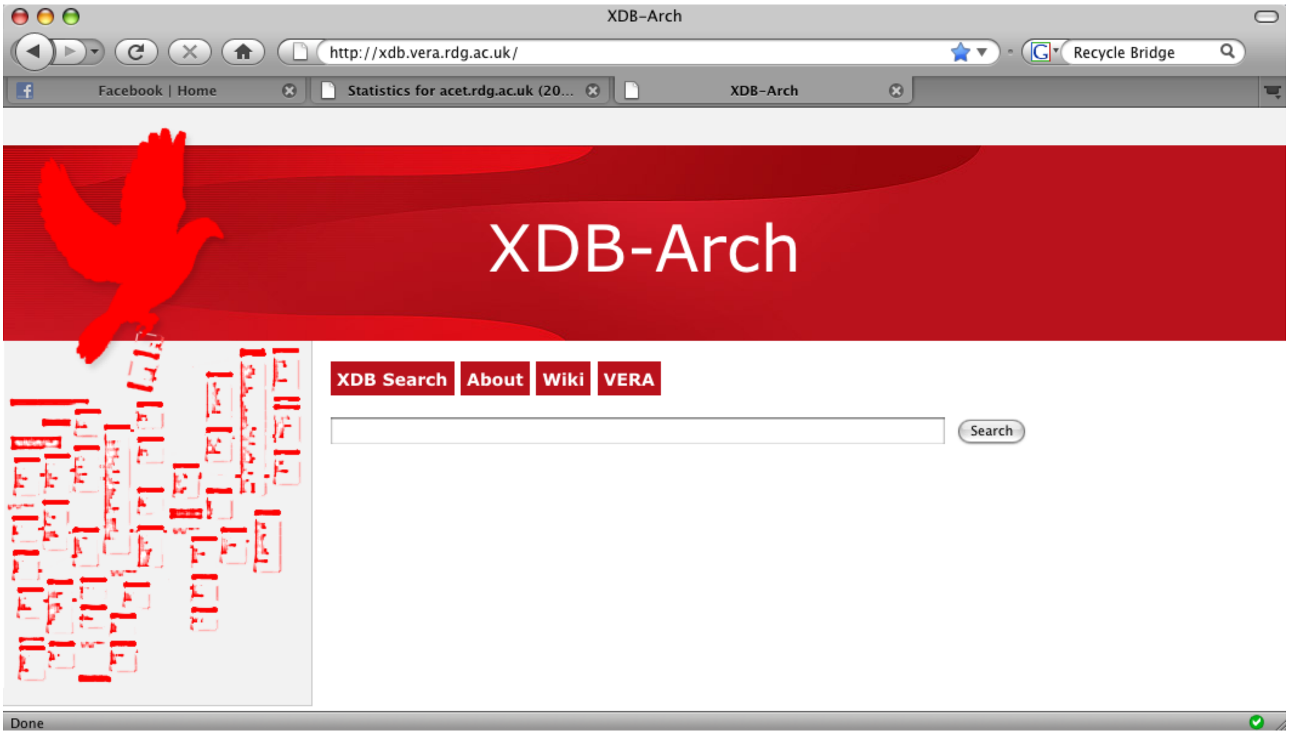


Figure 2. The XDB Interface

The high-level architecture of the XDB is shown in figure 3. It is based on a peer-to-peer architecture, using a P2P application framework called Tycho[4]. Tycho implements four main components: *consumers* (C, in figure 3), which generate requests, *producers* (P, in the diagram), which service requests, *mediators* (M, in the diagram), which act as routers between producers and consumers and implement the peer-to-peer connectivity, and a *virtual registry* (or VR), which is a simple distributed database spread across all of the mediators in the network. Every producer and consumer which joins the network has a record describing it inserted into the virtual registry. The producer or consumer may then add additional material, in the form of a large text entry, to that record. The entire VR, including the additional material, can be queried *en masse* by any component of the system, using a single query.

Built on top of the Tycho architecture is the XDB system. The XDB implements its user interface for searching the databases as a very simple web application, which acts as a Tycho consumer. On the other side of the coin is the *database connector*, which is a Tycho producer component. The database connector is responsible for:

- Publishing metadata to the virtual registry about the repository it services.
- Parsing queries sent to it.
- Performing queries against a specific database.

---

4 Baker, M.A., and Grove, Matthew. “Tycho: A Wide-area Messaging Framework with an Integrated Virtual Registry”, in Special Issue on Grid Technology of the International Journal of Supercomputing, (eds)

The peer-to-peer architecture of the XDB ensures that there is no single point of failure for the network – if a database or producer fails, or the network connection fails, the remainder of the network will still operate fully. The other advantage of the P2P design is that it permits multiple search interfaces to be implemented and attached to the network. Again, this ensures that there need be no single point of failure at the search interface, and allows alternative interfaces to be developed without affecting existing ones.

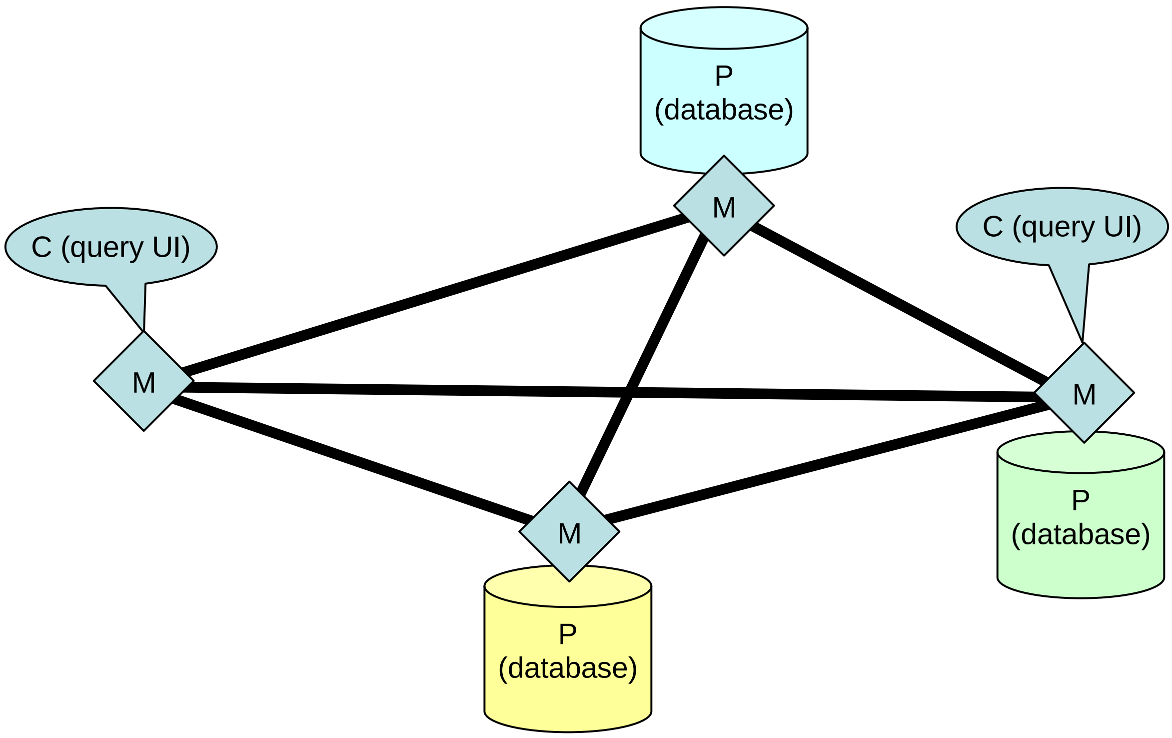
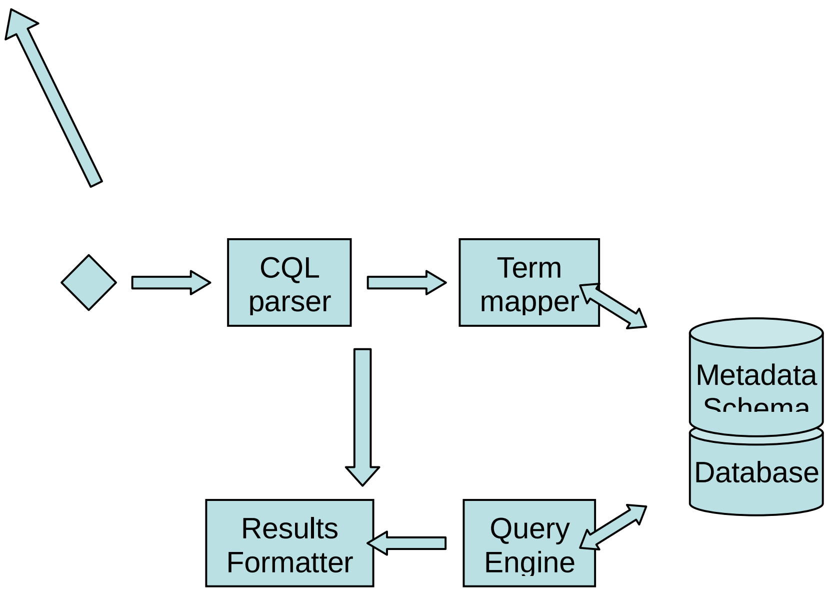
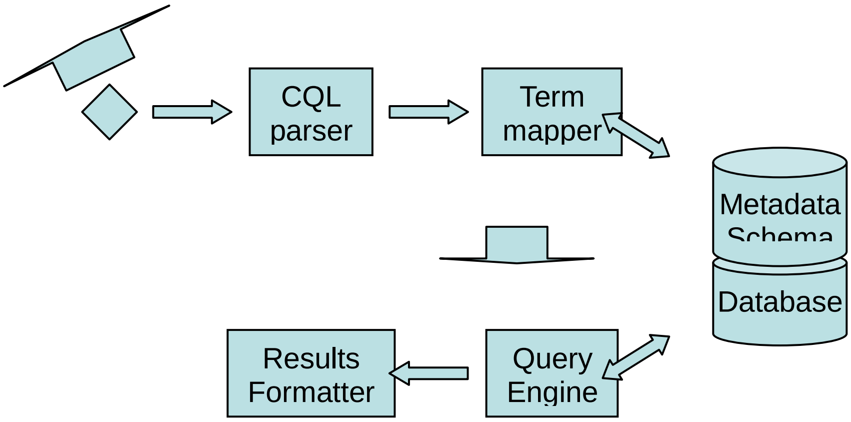


Figure 3. High-level peer-to-peer architecture of XDB



4The internal architecture of the XDB database connector is shown in figure 4. The life-cycle of a query involves several stages: *parsing* to determine search terms and relationships, *term mapping*, which attempts to find alternative search terms suitable for the vocabularies used in each data store, the *query* itself, and *results processing*, before the results are returned to the user in a suitable form.



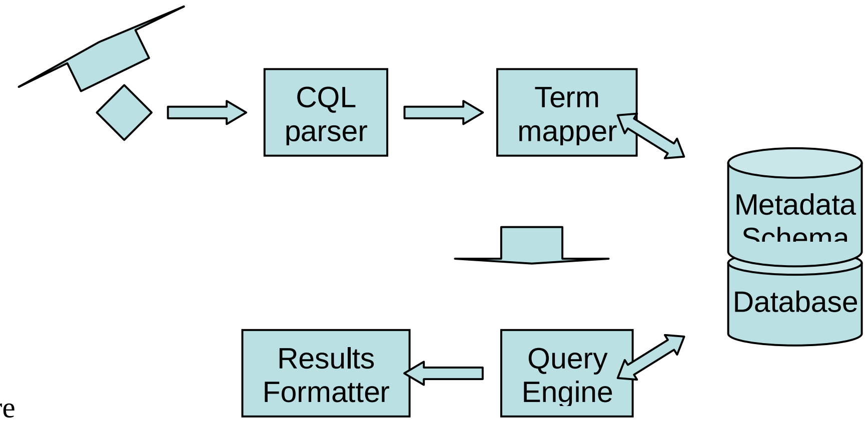


Figure 5. Architecture of XDB database connector

In parsing the query, we use the Common (or Contextual) Query Language[6], or CQL, which is a search language based on Z39.50 semantics. It aims to be a rich yet simple search language – effectively mid-way between the ease-of-use of, say, Google's search language, and the power and complexity of SQL, the structured query language used in most relational databases. CQL supports simple keyword search, boolean relationships (AND, OR, NOT), and more complex relations, such as proximity search ("word A is near to word B") and schema restrictions ("find this word, appearing in this context within the database schema"). The outcome of parsing the CQL search expression is a *parse tree*: a data structure describing the terms and relationships in the expression and indicating which words are search terms and which are search relations.

The *term mapper* is potentially the most complex component of the search architecture. Its function is to examine search terms and try alternative searches or terms which may be more suited to a particular database. In most subject-specific data repositories, there will be a vocabulary of technical terms in common use within that repository. However, across closely related disciplines, or even within disciplines but across different repositories, terminology may vary. For example, the terms "writing", "graffiti" and "inscription" may, be used interchangeably; be used to describe distinct concepts; or one may be used for a single concept in preference to the others, depending on the precise subject area and the culture of the repository owners. It is the function of the term mapper to understand these differences in terminology and to suggest alternatives. For example, searching for "inscription" in a database known to use that term, as well as the requested

6 Contextual Query Language v1.2, http://www.loc.gov/standards/sru/specs/cql.html

"writing" term. We are currently using a very simple term mapper, and the set of techniques that could be used is still an open problem. We are following up this aspect of the research within a new project, called Linksphere[7].

The *query* phase of the search simply uses the parse tree (and any alternative parse trees suggested by the term mapper) to perform a query on the underlying database. This component of the system must be customised for each data store in use. We currently have query engines for searching the small finds and context descriptions within the Integrated Archaeological Database, and for searching the Vindolanda inscriptions database from Oxford University's Classics department.

The final phase of the search, *results processing*, formats all the results in a similar manner regardless of their source, returns them from each data source to the requesting user interface, and combines the results for presentation to the user. The results are presented in a simple form, with a deep link to each found item in its respective database, a short section of textual context for the search terms, and potentially thumbnails of a related image or images, if the data store's query engine returned any.

As mentioned above, we currently have two database connectors operational with the XDB: one for the IADB, specifically for the database of the Roman era Silchester excavation, held at the University of Reading. The second is for a database of inscriptions on wooden tablets found at the fort of Vindolanda on Hadrian's Wall.

4 ARCH3D

The prototype data visualisation service of the VERA environment, called Arch3D, allows the integration and investigation of the multi-dimensional datasets obtained from an excavation. Most previous efforts at 3D visualisation of archaeological data have either been in the form of reconstructions of the site as it might have appeared above ground[8], or have worked on the principle of showing the exact positions of finds and other archaeological entities within the ground[9,10]. For sites such as Silchester with no effective visual or descriptive record of the original town, the former method is more a matter of archaeological interpretation and artistry, effectively using the computing facility merely as a canvas to draw the reconstruction on. On the other hand, the latter approach could potentially be of use in analysing the history of the site. However, a view showing a strict positional interpretation of the site is potentially very confusing: we see only the view of the site through the distortions of time and movements of earth of site under human occupation. Our approach with Arch3D has been to combine two forms of data into a single display: the X-Y horizontal position of a context and its related finds, and the stratigraphy normally expressed in the form of a Harris matrix.

As shown in figure 5, our visualisation of the site in Arch3D shows contexts as flat "plates", representing the outline of the context in plan. The third dimension is used to indicate the position within the stratigraphy of the site: contexts shown higher up are more recent, those shown lower down are older. This offers a good

7 Linksphere project, http://linksphere.org/

8 Beacham, Richard and Hugh Denard "The Pompey Project: Digital Research and Virtual Reconstruction of Rome's First Theatre" Refereed Proceedings of the ACH/ALLC Conference: 'Digital Media and Humanities Research' Journal of Computers and the Humanities **37** No.1 (2003), 129-140.

9 Losier, LM., Pouliot, J., Fortin, M. (2007), "3D geometric modeling of excavation units at the archaeological site of Tell Acharneh (Syria)", *Journal of Archaeological Science* **34** (2007), 272-288

10 Katsianais, M., Tsipidis, S., Kotsakis, K., Kousoulakou, A., (2008), "A 3D digital workflow for archaeological intrasite research using GIS". *Journal of Archaeological Science* **35** (2008), 655-667

conception of the shapes and relative positions of the contexts in the site — something missing in a normal Harris matrix — and also of the stratigraphic relationships between the contexts. The system allows the display to be rotated, panned, and zoomed, affording views of the model from any angle and viewpoint. See figures 6 and 7 for examples. Figure 8 shows a further feature of the viewing system: the ability to "flip" from any current view to a simple bottom-up (or top-down) view, and back again. This can be used, for example, to re-orient the viewer in a complex environment, or to examine the exact plan relationships of a set of contexts, without losing the settings for a side-on or diagonal view.

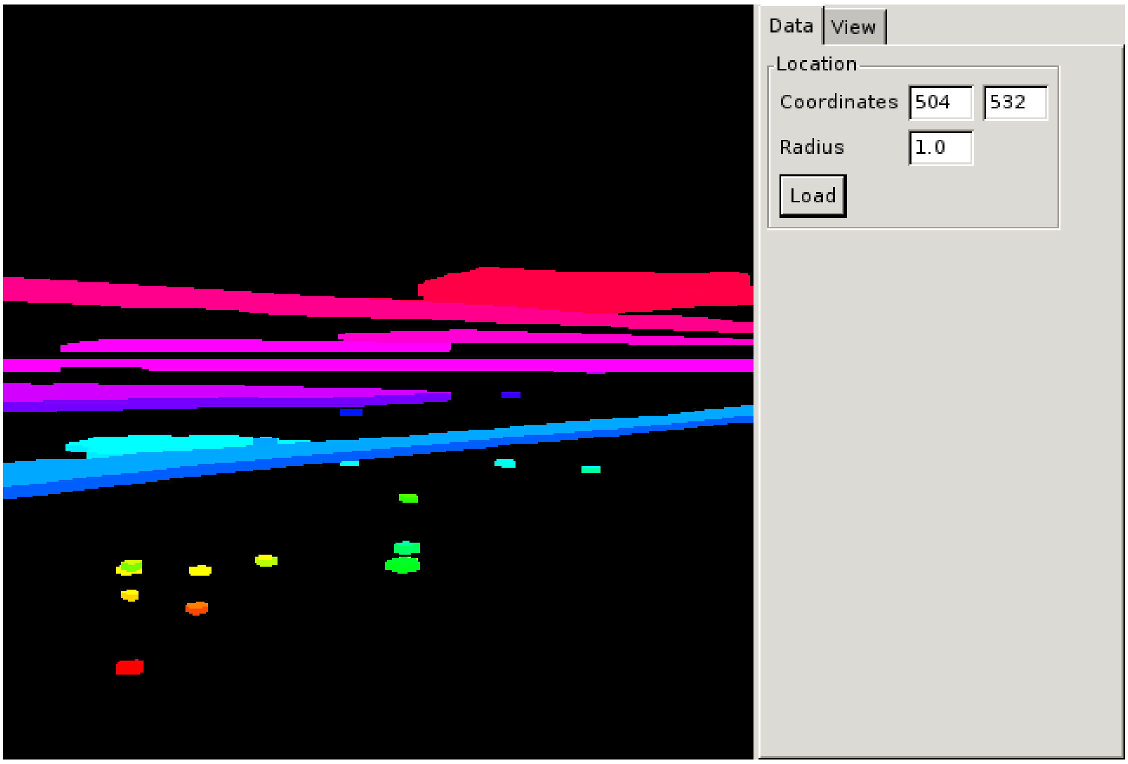


Figure 6. The XDB Interface: side view of visualisation

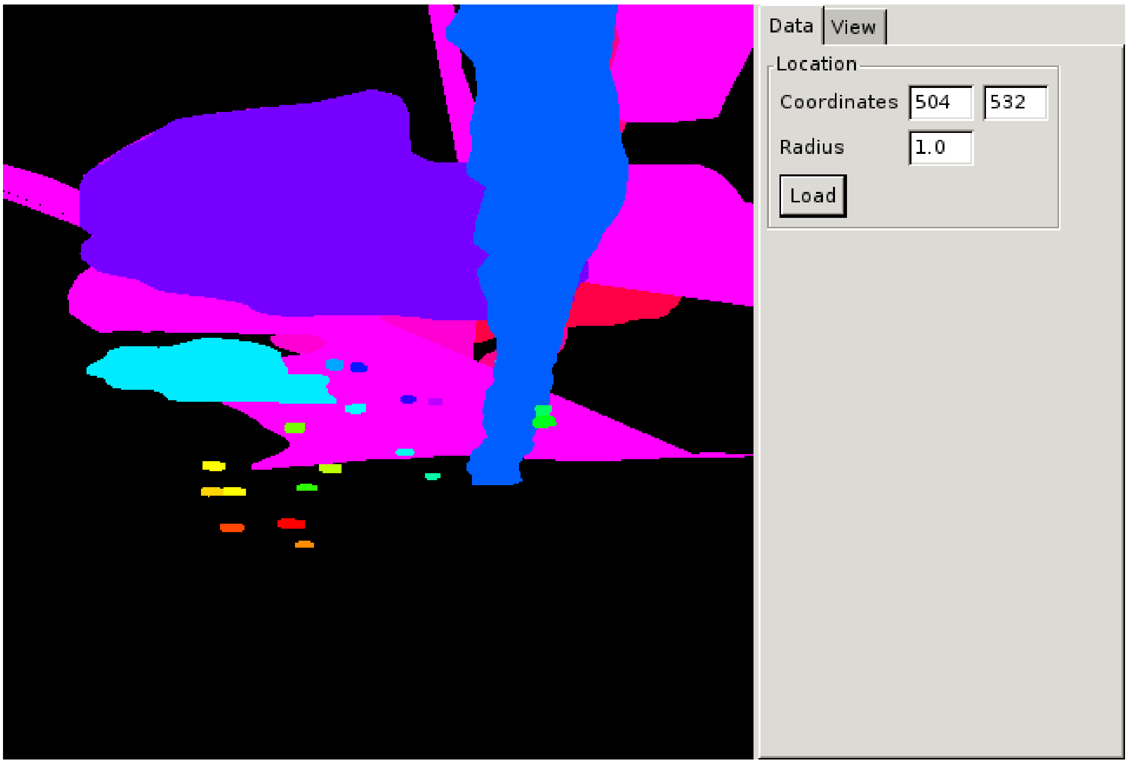


Figure 7. The XDB Interface: rotated view

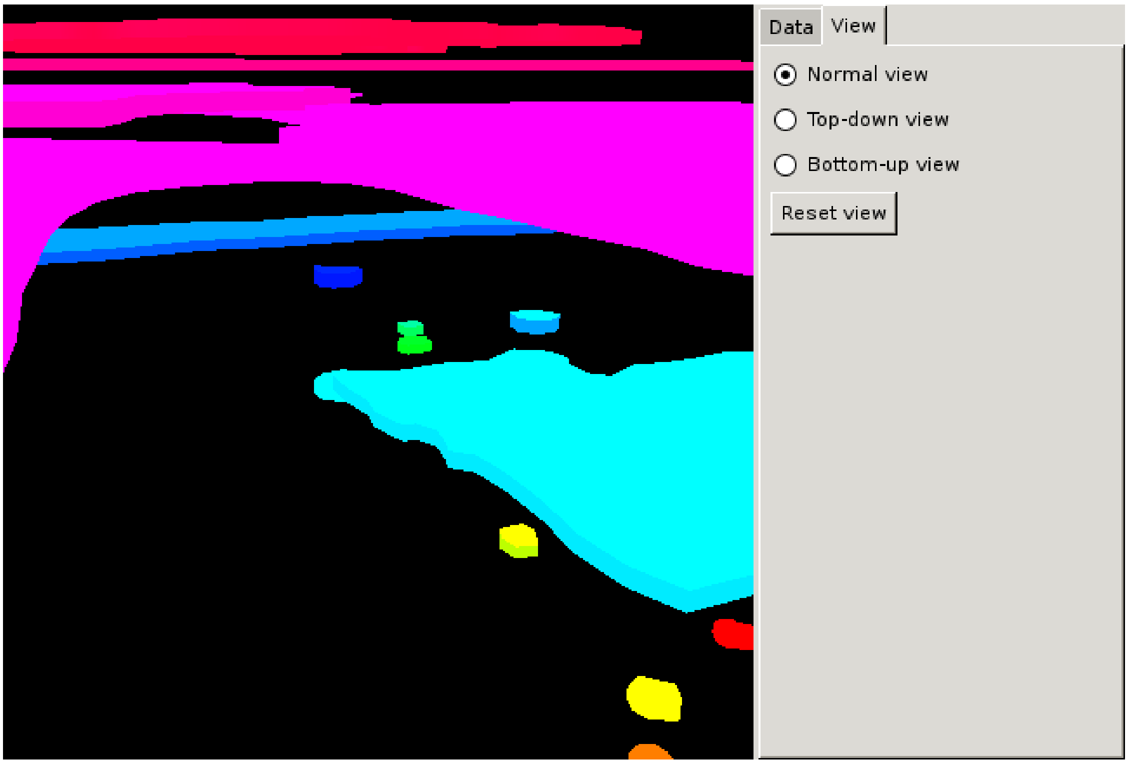


Figure 8. The XDB Interface: rotated and zoomed view

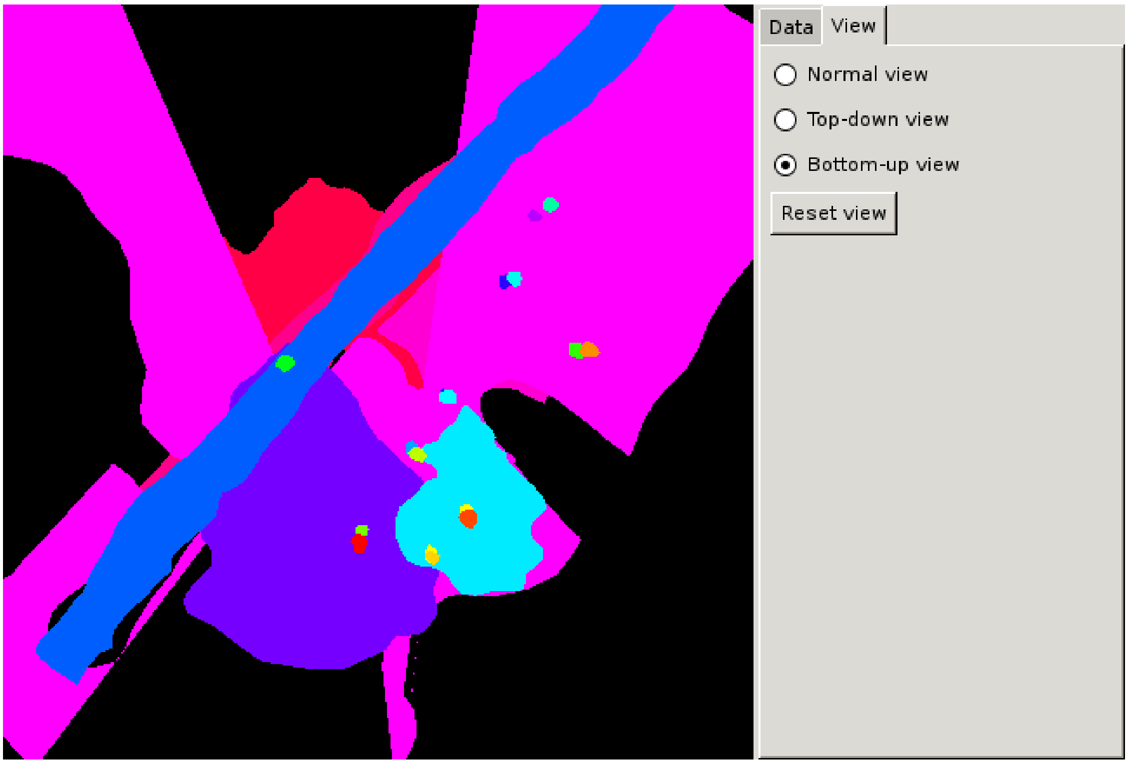


Figure 9. The XDB Interface: bottom-up view

The selection of contexts to display is currently based around picking those contexts, which have centroids within a given distance of a centre point. It is possible to make multiple selections of contexts, in which case the Arch3D system will integrate those separate context sets into a single display. The system uses a simple machine-readable HTTP interface to retrieve its data directly from the IADB over the Internet.

5 CONCLUSION

The VERA information environment offers a range of tools for managing, finding, and analysing data from archaeological excavations. In addition to the Integrated Archaeological Database, which acts as a data entry and repository system, we have developed a portlet component, the Recycle Bridge, for hosting arbitrary legacy services within a JSR-168 compliant portal environment, which we use for hosting the IADB in a single-sign-on environment. We have also developed a system, the XDB, for search across multiple databases in a semantically-aware manner, which allows researchers to find data distributed across several repositories without having to search in each repository individually. Finally, we have developed a prototype 3D visualisation environment for examining the logical and spatial stratigraphy of an archaeological site in a manner similar to that of a Harris matrix, but which gives benefits over the traditional Harris matrix by showing the outlines and shapes of the contexts being displayed.

BIBLIOGRAPHY

Baker, Mark, Smith, Garry, Grove, Matthew, Lakhoo, Rahim, Mills, Hugo, and Albing, Carl. "Using a RESTful messaging and registry system to support a range a distributed applications", accepted for publication, 8th International Symposium on Parallel and Distributed Computing (ISPDC 09), Lisbon, Portugal, June 30 to July 4, 2009.

Baker, M.A., and Grove, Matthew. "Tycho: A Wide-area Messaging Framework with an Integrated Virtual Registry", in *Special Issue on Grid Technology of the International Journal of Supercomputing*, (eds) Gravvanis, George A., Morrison, John P. and Fox, Geoffrey C., Springer, Volume **42**, 83-106, March 23, 2007, ISSN: 1573-0484

Beacham, Richard and Denard, Hugh "The Pompey Project: Digital Research and Virtual Reconstruction of Rome's First Theatre" Refereed Proceedings of the ACH/ALLC Conference: 'Digital Media and Humanities Research' Journal of Computers and the Humanities **37** No.1 (2003), 129-140

Contextual Query Language v1.2, http://www.loc.gov/standards/sru/specs/cql.html Retrieved June 2009

Katsianais, M., Tsipidis, S., Kotsakis, K., Kousoulakou, A., (2008), "A 3D digital workflow for archaeological intrasite research using GIS". *Journal of Archaeological Science* **35** (2008), 655-667

Linksphere project, http://linksphere.org/ Retrieved June 2009

Losier, LM., Pouliot, J., Fortin, M. (2007), "3D geometric modeling of excavation units at the archaeological site of Tell Acharneh (Syria)", *Journal of Archaeological Science* **34** (2007), 272-288

O'Riordan, Emma Jane, Clarke, Amanda Sarah and Fulford, Michael. "Managing Change: Introducing Innovation into well-established systems." 37th Annual CAA, Williamsburg, VA, 2009

Recycle Bridge, https://vera.rdg.ac.uk/tools/trac/portlets/ Retrieved June 2009

Silchester Web Site, http://www.silchester.rdg.ac.uk/ Retrieved June 2009