

# White Paper Report

Report ID: 98533

Application Number: HD5031108

Project Director: Paul Jacobs (pfj1@ra.msstate.edu)

Institution: Mississippi State University

Reporting Period: 3/1/2008-12/31/2009

Report Due: 3/31/2010

Date Submitted: 9/15/2010



DATS: Digital Archives Transaction System

# Final Performance Report and White Paper

**HD5031108**

**DATS: Digital Archives Transaction System**

Institution: Mississippi State University

Project Directors: Christopher Holland, Paul F. Jacobs

December 23, 2009





# Narrative

## Project Activities

The proposal arose out of work with databases of archaeological and museum materials at Mississippi State University, as well as from work with other systems of data distribution. Under current conditions research in artifact analysis using on-line resources requires that a dozen (or 50) individual searches must be made across a dozen (or 50 or more) individual sites; and because each site likely will use its own in-house nomenclature and a data structure constructed specifically for that particular collection of artifacts, searches in each of the dozen sites will have to learn the particular name for an artifact type that may be common to all. The need for a tool that works seamlessly with multiple types of databases and data sources became apparent. A “distributed archive transaction system” (DATS) seeks, first, to provide a means for searching simultaneously all twelve (or 50) sites with but a single query to retrieve relevant data from all data sources. Second, because many of the data sources used by active excavations are works in progress, DATS as a “distributed archive” system provides a means for recovering relevant data without requiring original data sets to become fixed or static. DATS works with original data sources only and does not require or possess copies of data sets which would then have to be homogenized to standard data formats and standard nomenclature. The benefit of a “distributed archive” approach is that as necessary additions, deletions, and corrections appear in the data sources these are immediately available to users.

The DATS Project, then, sought to develop communication software that will negotiate retrieval of data from multiple datasets written in multiple database types and using non-standardized nomenclature. DATS provides a rich source of data while protecting the need of data owners to control data sets.

Preparation of a working version of DATS was assigned to PI Holland, while preparation of datasets and increasing numbers of client data owners became the responsibility of PI Jacobs.

## Major Activities

### DATS Server

The basis for the project was the DATS Server: a software system that receives search requests and then distributes those requests to various DATS Agents which, in turn, respond to the DATS Server with relevant data. The DATS Server then packages these multiple responses from the DATS Agents into a single response for the end-user. The DATS Server also tracks available Agents and other DATS Servers that have registered with it, continuously tracking which services are available for searches.

### DATS Agents

DATS Agents are small software applications that interface directly with data providers’ systems and databases. The DATS Agent resides on the same server as the data source and communicates with that data source. These Agents can be written in any

programming language and can interface with whatever database the data provider uses. Data providers can tailor the use of the data and the results given to the DATS system because the DATS Agent software is in the control of the data provider. The Agent software communicates with the DATS Server system via JSON or XML formatted data transactions. The Agent can then communicate directly with the local databases using a native communications method.

### DATS Pseudo-Agents

Once the project was underway we leveraged the fact that both Google and Yahoo! had created programming interfaces into their vast search indexes. By using that technology we created what we call Pseudo-Agents. Pseudo-Agents are standard DATS Agent services with one exception: in addition to interfacing directly with a local database they query the Google or Yahoo! indexes via RESTful webservice API calls. These Pseudo-Agents allow the DATS Server to search sites that don't have native Agent nodes running on their systems. That means that there is effectively no barrier to entry for taking part in the DATS system is; if a site is on the internet it can probably be included into the DATS system through a pseudo-agent.

### DATS Demonstration Site

The DATS Server software is itself not a user-facing technology. That role is filled by other technologies (web servers, web applications, desktop applications) which communicate with the DATS Server. The demonstration site for DATS is a standard web site that communicates with the DATS Server via RESTful requests and JSON-formatted data. These are common practices for today's web-enabled applications and relatively easy to implement on any platform.

The DATS Demonstration Site also has a "blog" section that has project notes, presentation slides, and in-progress screen shots of the project.

### Changes/Omissions from the project

The project had originally planned to recruit many data providers. That proved to be a difficult task for various reasons: lack of actual digital records for data, reluctance to share digital data that does exist, and lack of technical knowledge needed to get a DATS Agent working with the provider's data.

Tackling these problems actually became a source of great discovery when PI Holland created the Pseudo-Agent idea. A rough yet successful Pseudo-Agent was then tested to support the idea. The implementation of Pseudo-Agents became even easier when both Google and Yahoo! released open interfaces to their data. At that point the project was able to add data sources with the click of a button. Pseudo-Agents do not give the kind of direct access to data that genuine Agents can provide, but they certainly do fill the need for accessible data when doing homogenized comparative searches.

### Publicity Efforts

Project members presented papers at the international meeting of the Archaeological Institute of America (AIA) in Philadelphia, PA, USA and at Xanthi University, Xanthi, Greece. The project was also featured in two articles for the *CSIG News of the Coroplastic Studies Interest Group of the AIA* and was featured in the refereed journal *Ancient Near East*, 70. See appendix for references to these presentations and publications.

## Accomplishments

### Comparisons to Goals

One of the main aims of the DATS project was simply to understand whether or not the goal of real-time distributed searching of multiple datasets could work. It seems that it is not only feasible but has become fairly mainstream and various search companies such as Google and Yahoo! are now offering product APIs that allow distributed searches. Just judging by this indicator alone it is safe to say that we were/are on the right track.

An important goal of the DATS project was the involvement of data owners in the concepts of freely sharing data and access to databases. Opportunities to “sell” these ideas to data owners were exploited in public presentations, in published papers, and in private communication. In principle, data owners generally welcomed the ideas. And in fact, DATS has been given access to several data sets which include artifacts in the Lois Dowdle Cobb Museum, Pierides Museum, Stinking Water (Mississippi) excavation project, Lahav Research Project excavations, Maresha figurine corpus, Tell Halif figurines database, Tell Gezer artifacts of Field IV, database of Freshwater Mussels retrieved from excavation sites, photographic archives of Near Eastern archaeological sites. Nonetheless, in practice there continues to be noticeable reluctance by many data owners to release data to public access. Hence, though negotiations with data owners often begin on a clear note of agreement about the need to distribute data, often the request for access to specific sets of data bring about delay and hesitancy.

The DATS project has been in conversation with several data owners for more than a year in efforts to include their data sets with the DATS experiment; though two have agreed to participate, actual access to the databases has not yet occurred. In other instances, data owners who had at first expressed enthusiasm have either shied away completely or have pushed the idea to the background of their priorities. Hence, one of the conditions which led to the DATS project—namely, the need to make data more readily and easily available—remains an obstacle to moving the project forward.

On the other hand, in part as compensation for the fact that data owners have not readily granted access to their data, the DATS Project turned a part of its attention to data retrieval through “Pseudo-Agents” and search engines such as Google and Yahoo. Although this aspect of the DATS site was not an original part of the proposal, it has opened the DATS service to significant sources of information.

### Unachieved Goals

The project achieved every major technical goal that was originally set out in the proposal. The project team plans to continue to pursue and to improve those achievements and to provide an even better product by increasing the number of data partners as well as leveraging the power of the Pseudo-Agents that were created by PI Holland. Technical improvements to the demonstration site will also be undertaken.

### Audiences

Because the DATS site/product is freely available via the Web it is open to use by all audiences, although current content of the system is targeted at an audience of professional archaeologists and art historians. Feedback by those that have used the system has been positive and has mentioned that it is much easier to use than traditional methods of comparative research within these fields. We are looking forward to expanding the system to handle other facets of research.

## Evaluation

An evaluation of the DATS Project was formally carried out as required by the terms of the grant proposal. Four people had been identified as evaluators of the software after it had become accessible to the public. These four were Dr. Adi Erlich, Department of Art History, University of Haifa, Israel, Dr. W. Jack Bennett, CEO of Archaeological Assessments, Inc., Dr. Joe D. Seger, Director, Cobb Institute of Archaeology, Mississippi State University, Dr. Nestor Tsirliganis, Director of Cultural and Educational Technology Institute/Athena R.C., Xanthi, Greece, and Dr. Vassilios Tsioukas, Department of Architectural Engineering, The Democritus University of Thrace, Greece. These evaluators were selected on the bases of their interest in the subject matters in the initial DATS databases, because of their “ownership” of one or more of the databases, and because of their experience in writing software programs for the manipulation of data.

As the project reached the point of offering a working version of DATS each of the evaluators was given brief instructions on the workings of DATS, on how to navigate, and on the benefits anticipated. Each was asked to spend several hours and/or several sessions in using DATS to try to gather data relevant to imagined or real search needs. Each evaluator was asked to write a report on the sessions working with DATS, to highlight what worked well, what failed to work, and what might be done to improve the software delivery.

Dr. Erlich is a recognized expert in coroplast studies and has published extensively on the analyses of terracotta figurines; her analysis of DATS was to determine whether the location of comparanda via DATS aided significantly in her studies figurines found in excavations at Tel Maresha, Tel Dor, and other sites. Dr. Bennett is a professional archaeologist whose company depends on accessible data for the reports on surveys and excavations required by clients, particularly for providing contexts and comparative materials. Dr. Tsioukas has written numerous software programs for the use and collection of data, assuring useful insights into the workings of DATS by an expert in computer languages. Dr. Seger, as Director of a research institute, brings an interest in data source integrity to the evaluation. Dr. Tsirliganis, as Director of C.E.T.I. , also is interested in the accurate and rapid retrieval of data relevant to cultural heritage.

All of the evaluators concluded that the project concept itself is a worthwhile effort and shows promise for their individual needs. Dr. Tsirliganis commented that the speed of delivery was “very good,” Dr. Erlich that DATS is “very helpful and effective,” Dr. Tsioukas that DATS has a “nice interface and ease of use,” Dr. Bennett that his “experience working with the initial version the DATS project has been most positive,” and Dr. Seger that the results of working with DATS was “somewhat satisfying.” Dr. Bennett’s estimation of the intent of the DATS project speaks to the need for the service DATS is attempting to supply: “assisting scholars world-wide to gain access to the information about and contained in archeological collections is extremely important for the future of our discipline. In a world in which field investigations have severely declined and storehouses are crammed with unanalyzed materials the development of strategies and technologies to make ancient materials accessible to modern scholars should be one of our very highest priorities.”

The evaluators commented with criticisms of DATS in its initial form. Dr. Tsirliganis noted especially that in its initial form DATS suffered from a lack of instructions for a user. Furthermore, he noted that searches sometimes result in results that appear to be extraneous to the search query. Dr. Erlich also noted non-relevant results to search queries. Dr. Tsioukas, to the contrary, commented on the apparent 100% accuracy of responses to his queries. He also commented on the usefulness of DATS in querying subjects not within the realms of the data sources subscribed. However, Dr. Bennett returned the point of view to a “more focused search results,” a reference also to the fact that among a list of useful results are extraneous items. Dr. Bennett

stated the criticism succinctly: “I found the system was not as rigorous as it might be in narrowing responses to queries that are fairly specific. I found that if I entered a category of artifact like ‘spear point’ I was returned a long set of examples. As I tried to narrow this search by asking for ‘bronze spear point’ or ‘trefoil spear point’ or ‘bronze trefoil spear point’ I received much the same set of results. There needs to be a way in which these modifiers would narrow the results.” This same critique was offered by Dr. Seger, as well.

The evaluators also suggested some specific corrections or alterations to DATS in its next form. Dr. Tsirliganis requested a ranking of search results from most to least relevant finds, as well as access to more than the first ten results from each of the various data sources. Dr. Tsirliganis also suggested the ability of DATS to perform searches within results of a search, as well as searches with delimiters such as NOT and XOR conditions. Dr. Erlich suggested that thumbnail photos appear with all of the responses to a query. She also pointed to the difficulty, e.g., of different spellings and different names for commonly searched terms causing only partial recovery of relevant data. Both Dr Erlich and Dr. Tsirliganis pointed out the interface problem in which an attempted return to search results after having reviewed an individual element of that result is by way of the “BACK” button, which unfortunately leads the user to a new search and not to the desired results of the previous one.

The reviews were incorporated into the planning for the second version of DATS; some of the criticism of the working of DATS has already been addressed. Others are currently under construction.

## Continuation of the Project

The DATS system will continue to be supported by the Cobb Institute of Archaeology at Mississippi State University.

Continuation of the DATS project is planned both in terms of technical improvements as well as by the addition of new data resources. Since the start of the project we have entered into cooperative work with SUNY-Buffalo, Wilson College, and the membership of the Coroplastic Studies Interest Group (CSIG) of the Archaeological Institute of America (AIA) in attempts to increase participation in the DATS project. Near-term technical enhancements will ease the technical knowledge needed to make data accessible and allow more data providers to take part in the project.

## Long Term Impact

The Distributed Archives Transaction System (DATS) was constructed to resolve difficulties in archaeological and museum object-based research by providing a web site and software, which will enable a single search across numerous data sources. The significance of the project rests in the fact that recovery of desired information will be simpler (via single query to address all client data sources) and faster (as relevant data to a query will be uniformly formatted to meet the request of the user). Whether the data is requested by a professional museum curator, an art historian, an archaeologist, or by a casual browser DATS will return relevant information rapidly and efficiently; and since DATS provides thumbnail views of items found in its search a user will be empowered to select items that actually resemble the search interest.

One external reviewer of the DATS project, Dr. W. Jack Benett wrote of the project’s significance and potential impact, “Not only do I think that this is an extremely important system ... for the retrieval of material now largely inaccessible, it will provide an impetus for the institutions which house these materials to develop and employ cataloguing systems that will be more user

friendly to those who wish to study these materials. And, with field opportunities rapidly closing, this will surely be the venue in which our discipline will make its most important new insights.” As additional data sources are added to the list of clients this long-term impact will come closer to reality.

The invited article “Sharing Archaeological Data: The Distributed Archives Transaction System” in the “Open Forum” section of the journal *Near Eastern Archaeology* vol. 70 demonstrates that the discussion among Near Eastern archaeologists has come to include the basic views and purposes set forth by DATS. The invited paper (“Getting Data into the Hands of Archaeologists: DATS”) at the conference on transfer of digital technologies at the Democritus University of Thrace also indicates the interest in the project by scholars outside of Mississippi State University. Similarly, the articles “A Distributed Archive for Coroplastic Research” and “Coroplastic Studies, An Argument for Total Publication” in the newsletter of the Coroplast Studies Interest Group have reached an audience whose academic research is closely tied to several of the data sources used by DATS.

## Grant Products

The primary consumer product of the DATS project is a freely accessible demonstration website.

<http://www.datsproject.org/>

The DATS Server and Agent language specification is based on, though not identical to, the format used by the Yahoo! Boss API. The DATS API specification is freely available via the DATS website.

In addition there is a development blog that includes presentation slides and transcripts, development notes, journal papers, and feature announcements for the duration of the project.

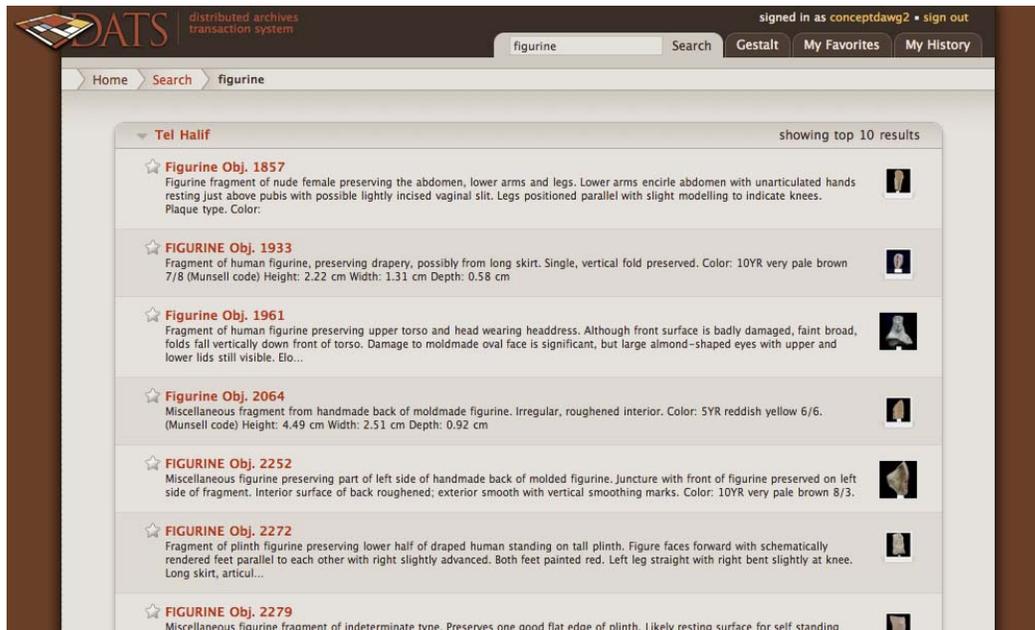
<http://www.datsproject.org/blog>

# Appendix

## Presentation and Publication References

1. Jacobs and Holland, "Sharing Archaeological Data: The Distributed Archives Transaction System" (invited article) in *Near Eastern Archaeology* 70, 2007 (appeared in 2008).
2. Jacobs, Invited paper "Getting Data into the Hands of Archaeologists: DATS" presented at University of Xanthi, Greece. May 2008, at Conference "Co-operation networks for the transfer of know-how in 3D digitization applications" at the Cultural and Educational Technology Institute/"Athena" R.C., Xanthi, Greece.
3. Jacobs, "Getting Data into the Hands of Archaeologists: DATS" in Proceedings of "Co-operation networks for the transfer of know-how in 3D digitization applications" at the Cultural and Educational Technology Institute/"Athena" R.C., Xanthi, Greece.
4. Jacobs, "Coroplastic Studies, An Argument for Total Publication" in Newsletter of the Coroplastic Studies Interest Group (CSIG), No. 2, Summer 2009. [http://www.coroplasticstudies.org/images/csig\\_newsletter\\_02\\_0001.pdf](http://www.coroplasticstudies.org/images/csig_newsletter_02_0001.pdf)
5. Holland, "A Distributed Archive for Coroplastic Research: [www.datsproject.org](http://www.datsproject.org)" in Newsletter of the Coroplastic Studies Interest Group (CSIG), No. 2, Summer 2009. [http://www.coroplasticstudies.org/images/csig\\_newsletter\\_02\\_0001.pdf](http://www.coroplasticstudies.org/images/csig_newsletter_02_0001.pdf)
6. Jacobs and Holland. 110th Annual Meeting of the Archaeological Institute of America, held in Philadelphia, PA January 8–11, 2009. Paper: "A Distributed System of Archives for Coroplastic Studies."

## Samples of Demonstration Site



British Museumshowing top 10 results

---

☆ **'Violin' figurine**  
 Throughout the Early Bronze Age in the Cyclades (approximately the third millennium BC) marble figurines were produced in two basic types: schematic and naturalistic. Schematic types, like this example, are most common in the earliest period of ...

🔍 **Related search terms**

figurine • violin • naturalistic • Early Bronze Age • Cyclades • marble figurines • third millennium BC • The British Museum • characteristic • outline • British Museum • legs • Room 11 • Cycladic Islands • GR • Sculpture • 2800 BC • island • Amorgos • Aegean Sea

---

☆ **Terracotta female figurine**  
 This fired clay female figurine comes from a tomb in the rich Phoenician cemetery at Tharros. ... Terracotta female figurine. Phoenician, 6th century BC. From Tharros, Sardinia. Probably representing the goddess Astarte ...

🔍 **Related search terms**

Phoenician • Tharros • Phoenician art • female figurine • tombs • cemetery • Terracotta • goddess Astarte • Egyptian influences • clay • fertility goddess • wears • The British Museum • Sardinia • wig • fragments • paint • painted black • integral element • British Museum

---

☆ **Phallic figurine**  
 The natural shape of a calcite cobble has been used to represent the outline of the lovers. ... Sculpted figurine of two lovers, from Ain Sakhri. Natufian, about 10,000 years old. Probably from the cave of Ain Sakhri, Wadi Khareitoun, Judea ...

🔍 **Related search terms**

figurine • phallic • arms • Sakhri • Wadi • Judea • natural shape • calcite • cobble • outline • heads • British Museum • two lovers • Natufian • cave • legs • stone point • chisel • one another • one hug

---

☆ **Ivory figurine**  
 This beautiful figurine of a young Hittite girl is made of ivory. Her hair is drawn back and hangs in a pig-tail down her back. ... The function of the figurine is unknown, but ivory was widely used in the ancient world to decorate furniture The dowel hole in the square base ...

🔍 **Related search terms**

ivory • figurine • Hittite • hair • drawn back • hangs • pig-tail • hands • sides • proportions • curves • treatment • Hittite king • The British Museum • naturalistic • stylized • ears • triangles • knees • pubic triangle