The title of this paper, "Managing the Digital Dig" came from our ongoing discussions on how to normalize and preserve large archaeological data sets. Site materials are emerging in abundance and often in a chaotic state. Some sites and objects are described in extensive excavation reports; others, separated from the original context, have only cursory, object-based metadata. Metadata and contextualization have become as much buzzwords at AIA conferences as they are in digital humanities meetings. A reason for this interest in archaeological metadata may be economic. A good summary of the state of publication in archaeology, at least in the UK, can be found in the publication *The Virtual Representation of the Past* in the chapter called "Stepping Back from the Trench Edge: An Archaeological Perspective on the Development of Standards for Recording and Publication". The
authors, Julian Richards and Catherine Hardman, propose that the rise of digital archaeology, i.e. a publishing model that moves excavation finds and reports quickly to the web, has changed the traditional means of sharing archeological information and parallels the dwindling of extensive archaeological publications (Richards and Hardman, 108).

**Preservation of Site Metadata (fig.2)**

Richards and Hardman are interested in the collection of archaeological data at the site and its immediate delivery. What of the archaeological data already languishing in legacy collections that could contribute to knowledge of the sites currently under excavation? In 2005, Heritage Preservation, in partnership IMLS, produced the Heritage Health Index Report ([http://www.heritagepreservation.org/HHI/HHIfull.pdf](http://www.heritagepreservation.org/HHI/HHIfull.pdf)), the first comprehensive survey conducted of the condition and preservation needs of cultural heritage collections in the United States. The survey encompassed more than 4.8 billion artifacts held in the public trust by more than 30,000 archives, historical societies, libraries, museums, scientific research collections, and archaeological repositories throughout the country. According to the report, approximately 30% of artifacts in these collections are in unknown condition. The percentage is even higher (46%) for bulk catalogued archaeological collections, such as potsherds and other small items (Heritage Preservation, 47).
Legacy and Image Collection (fig.3)

Obviously, legacy archaeological collections are often large and consist of multiple formats, including field notebooks, field drawings, artifacts, photographs, and media. Furthermore, within large institutions, archaeological collections can become dispersed, with paper records being deposited in an archive, specimens and artifacts entering museum collections, drawings and photographs being acquired by special collections, and slides and images becoming part of visual resource collections.

**tDAR – the Data Repository**

The call for a centralized system to aid in the development of artifact cataloguing has come from many quarters, but debate on standardization of archaeological data is of two minds. The first one is that the collection of data at a site is almost personal and analysis is at the discretion of the director of the time. This attitude towards data collection and dissemination is visible in recent projects such as tDAR, the Archaeological Record, a Mellon-funded initiative to build a national repository for archaeology based on pdfs of raw data. The material in tDAR is provided with open access by the archaeologist at his/or her own expense.
Open Context – Sifting the Objects

A second response to the challenge is Open Context. The Open Context site provides a data dissemination tool that has underneath an XML schema called ArchaeoML. ArchaeoML has recently been funded by an NEH grant to aid in the creation of an open repository for well-developed site data.

Projects such as tDAR and ArchaeoML assume that the archaeological data based on artifact collection is in a state ready for dissemination and readily available for inclusion in the data collection tool. This is rarely the case with legacy collections inherited by institutions. Cataloguing of site information is declared to be “the most crucial variable for archaeologists” and “also the only attribute which is uniquely a product of the excavation, and which cannot be ascertained from the artifact itself.” Yet, this information is often missing from legacy collections that have been “broken up” and whose artifacts have been privileged over other materials such as notes or images. Many institutions are then pressured to “re-assemble” that valuable patrimony and publish their excavation history digitally.

What is ArchaeoCore? (fig.4)
How can the site and any objects found on site during different stages of excavation be related back to each other? In particular, what is the relationship of fieldwork to institutional collection? With our ArchaeoCore project, we hope to provide a schema for the normalization of image metadata that will bring to the surface more finds and tie those finds to their site (fig.8). The answer lies in developing a model for archaeology metadata that foregrounds the site over the objects found.
ArchaeoCore is designed as a metadata schema for normalizing mounds of archaeological image data. The schema re-orient the description of data around the governing principle of context rather than item.

The CIDOC-CRM provides the most useful and comprehensive conceptual model for archaeology images. By its own definition, the CIDOC “CRM Core is a proposal of metadata elements for resource discovery both more general and more precise than Dublin Core.” Selected CRM Core entities were used as the basic entities to establish the structure of ArchaeoCore.

**Linking Past to Present and Back**

An obvious debt is also owed to VRA Core on which our beta version was based. A few observations can be made on variation in VRACore and ArchaeoCore in the cataloguing of ancient works by describing elements of an easy example, the Parthenon. The object Parthenon cannot be divorced from the site Acropolis and even then it has a region (thus, Greece, Attica, Athenian Acropolis, and Parthenon). VraCore refers to a creator and a creation site, neither of which are applicable in an archaeological context. Creation site differs from repository site in archaeology given the history of the transport of finds (Parthenon sculptures). Dates cannot
always be fixed as they depend on which building are you talking about, the archaic temple, the classical temple, the church, or the mosque. In ArchaeoCore we use the term persistent name an entity from the CIDOC-CRM. Parthenon, itself, is a persistent identifier not the name of the building. Thus, defined, we reached the foundation of ArchaeoCore and its deviation from VRA Core.

The ArchaeoCore Community (fig.7)

What remained was to find partners to test our theoretical metadata schema on real images. We found those partners in our Shared Shelf and ARLIS/VRA communities respectively. Our first partners in the development of ArchaeoCore were colleagues from the ICFA at the Dumbarton Oaks Research Institute who helped articulate the need for ArchaeoCore.
In the summer of 2012, a founding group consisting of VR professionals, archaeologists, metadata librarians, and archival librarians from UVA, Dumbarton Oaks, Princeton, and the IFA met at the Princeton Summer Summit and created a baseline schema for ArchaeoCore. Working from our two-part ArchaeoCore 1.5 schema which was based roughly on a VRACore work type/image type model, the group expanded the earlier model into a four-part prototype for a relational database. We hope to have that four-part schema go live soon at UVA as the KORE Archaeology Collection.
In ArchaeoCore, we define Site as “Description of a place (or group of physical sites) in which evidence of past human activity is preserved. Sites may range from those with few or no remains visible above ground, to buildings and other structures still in use. (Wikipedia) Must include name, modern location and date ranges; may include human cultures represented and physical description of location.”

At UVA, we have begun to catalogue in KORE images of our collection of artifacts from a 17th century site at Flowerdew Hundred on the James River in Virginia. We are working on just three sites from the more that 90 registered sites on the Flowerdew Plantation. The following screen shots will give you an idea of our progress from observing the archaeological metadata for just one potsherd, the 17th century North Italian Marbleware shown previously. The Marbleware is from the Stone House Foundation site.
Campaign Data (fig.10)

Defining what we meant by “campaign” has become our most recent struggle. Our present definition was refined at Providence with the help of archaeologists, VRA professionals, librarians, and library school students and is the following “Triangulation between institution, personnel, place, and time period. Field must encompass institution/sponsor, location, and date and may also include people involved, such as field director, excavators, photographers, and specific excavation units.” Our Campaign title is Piersey’s Manor House, the 17th century structure.

Artifact Data (fig.11)
The artifact data is perhaps the easiest to define as “a movable object representing human activity recovered by some archaeological endeavor. Must include artifact type, origin and date range; may include description, materials and repository information.” In our example, this is the screen to describe the brown marbleware sherd which moved by ship from its creation site in Northern Italy and was probably sold by a middleman to a planter in Tidewater Virginia.

**Technical/Administrative Data (fig. 12)**

The final screen contains the information about the pictorial representation embedded in the image and also includes the image rights information. This screen has morphed into the File Data information screen. It will become of greater use as we begin a project to photograph our artifacts from Flowerdew Hundred and add them to their respective metadata record.

The Archaeocore community is now informally affiliated with VRA and will meet again in Princeton in August of 2013. The screens just displayed will be ready shortly as a template in Shared Shelf for our partners and available on the University of Virginia Fiske Kimball Fine Arts Library blog (http://pages.shanti.virginia.edu/Fine_Arts_Library_News/) for anyone who wants to use them. In addition, a version of ArchaeoCore will be in use in the field in Greece and other locations this summer. We would also welcome any newcomers to the community.
Images

Fig.1  Kadmeion, Thebes, Assemblage of pot sherds (Stylianopoulos photo, 2010)
Fig.2  Flowerdew Hundred, Virginia, Site 44PG64, Metal fork (Photo courtesy of the Rector and Board of Visitors of the University of Virginia)
Fig.3  Flowerdew Hundred, Virginia, Site 44PG64, Iberian Pot sherds (Photo courtesy of the Rector and Board of Visitors of the University of Virginia)
Fig.4  Flowerdew Hundred, Virginia, Site 44PG64, North Italian Marbleware (Photo courtesy of the Rector and Board of Visitors of the University of Virginia)
Fig.5  Kadmeion, Thebes, Venetian Marbleware (Stylianopoulos photo, 2010)
Fig.6  Kadmeion, Thebes, Turkish domestic ware (Stylianopoulos photo, 2010)
Fig.7  Flowerdew Hundred, Virginia, Site 44PG64, Fragment of a 17c. plate (Photo courtesy of the Rector and Board of Visitors of the University of Virginia)
Fig.8  First ArchaeoCore Princeton Summit, screen discussions (Glendon photo, 2012)
Fig.9  ArchaeoCore Site screen capture (August, 2013)
Fig.10  ArchaeoCore Campaign screen capture (August, 2013)
Fig.11  ArchaeoCore Artifact screen capture (August, 2013)
Fig.12  ArchaeoCore Technical/Administrative screen capture (August, 2013)