The goal of the Digitized Archival Papers group is to design and build a system that will deliver digitized archival objects and their metadata to users in a meaningful way.

Archival collections are inherently different from the materials that are typically managed by libraries, such as books and serials. This inherent difference has necessitated the development of separate standards to describe archival materials. Description of Archives: A Content Standard (DACS) is the archival equivalent of AACR2 and RDA. Encoded Archival Description (EAD) is analogous to MARC. Both standards were developed to address the challenges associated with describing and cataloging archival materials. Archivists use these standards to produce finding aids, which are the archival profession’s primary tool to describe archival collections.

Though an archival collection can be broken down into its individual parts, or components, the relationships between those components as well as between the components and the whole, are integral to understanding archival collections. These relationships are identified and maintained through physical and intellectual arrangement. Therefore, archival description is designed to represent that physical and intellectual arrangement (generally represented as hierarchy) as well as to provide contextualizing information about the collection (context), such as its creator/s, how the materials were created or maintained, and subjects covered. The digitized versions of archival collections also need to be organized, managed, described, and delivered in a way that retains their unique characteristics as archival materials, maximizes the value of their metadata, and maintains their relationship to the original.

In their article, “When Archival Description Meets Digital Object Metadata,” Zhang and Mauney classify existing metadata representations for items digitized from archival collections into three different models: Embedded, Segregated, and Parallel. In the Embedded model, digitized surrogates are organized and discoverable within the finding aid. The Embedded model is contextual, hierarchical, and linear (not faceted) — traditionally, at least. In the Segregated model, digitized surrogates are organized and discoverable outside of the finding aid in a digital library.

---

1 Principle 1 of Describing Archives: A Content Standard (DACS) states “Records in archives possess unique characteristics. Archival materials have traditionally been understood to consist of the documents organically created, accumulated, and/or used by a person or organization in the course of the conduct of affairs and preserved because of their continuing value. They most often consist of aggregations of documents (largely unpublished) and are managed as such, though archival institutions frequently hold discreet items that must be treated consistently with the institution’s descriptive system. In the course of their regular activities, individuals, archival repositories, and other institutions may also consciously acquire and assemble records that do not share a common provenance or origin but that reflect some common characteristics, such as a particular subject, theme, or form. Such collections are part of the holdings in most institutions and must be described in a way that is consistent with the rest of the holdings.” (xii)

system that treats them as items. The segregated model is de-contextualized, flat, and non-linear (facetable).

Previous efforts at Yale Library to deliver digitized archival materials on the Web in a Library-wide system have been unsuccessful because they did not incorporate the two primary elements of archival representation: 1) complex hierarchy and 2) detailed context. In the past, we have followed the Segregated model primarily, shoehorning archival metadata into a library-specific framework while divorcing digital surrogates from the finding aid. These efforts attempted to deliver archival digital objects from a digital library, which was designed to provide access to digitized books, journals, and artworks. This method decontextualized the digitized archival materials, necessitated recataloging, and ultimately failed to provide a satisfactory user experience. One way to describe the effect that the Segregated model has had on digitized archival material is to imagine that we scanned every paragraph separately in a 50-year run of a journal; created individual metadata for each paragraph; dumped each paragraph and its metadata into a large digital library; and, finally, that we opted to let the user figure out how the paragraphs related to one another, to each article, each journal issue, and the series as a whole.

Other efforts, primarily undertaken by archival units, have attempted to use the finding aid as a discovery and access system by embedding static links within finding aids to the digitized surrogates, which are managed by an unconnected system. This method, which follows what Zhang and Mauney have called an Embedded model, has also been unsatisfactory since it is labor intensive (unsustainably so) and sends users away from the finding aid to view images in a larger digital library, whose inadequacy is described above.

A successful system requires a parallel model and parallel tact, one that draws upon the structure of hierarchical relations that inherently characterize representations of archival collections and one that is integrated into and can benefit from inclusion in a robust digital library.

An example of the Parallel model follows: Maintain, manage, and improve two systems that share a common architecture. One system, the Yale Finding Aid Database (YFAD), will store information about archival collections in Encoded Archival Description (EAD). The other system, a Digital Asset Management system (or an assortment of DAMS, potentially), need not always record additional descriptive metadata about the digitized surrogates that are already described

3 The next version of the Yale Finding Aid Database will also include support for Encoded Archival Context – Corporate Bodies, Persons, and Families (EAC-CPF) metadata. These records essentially function like DBpedia datasets for creators of archival materials. Since they will not include any additional digital material for the DAMS to manage (aside from metadata), they are not discussed in this document. Nevertheless, they could be integrated into the Hydra/Blacklight system at a later date. Their ability to enhance search results (similar to Google’s Knowledge Graph) could prove very useful as the database expands.
in EAD. Instead, the DAMS will need to be able to ingest portions of EAD and transform it into a format that will nevertheless retain some of the contextual relationships described and inherited in the source EAD file. For example, the component-level display of digitized surrogates in Yale Library’s installation of Hydra/Blacklight (i.e. Yale’s new DAMS) will need to provide additional context about the relationships of digitized and non-digitized archival materials that are housed within each parent collection.

As this is a new way for YUL to approach digitized archival materials, an entirely new system that encompasses the following is required: Ingest, Storage, Repository Architecture, Metadata, and Search, Discovery, and Access, each of which will be designed to handle the volume and complexities of digitized archival materials.

Several factors make this the right time to develop the system. The library recognizes the value of providing access to digital content, both published and unpublished, to our users and recently committed to a digital strategy. The library is currently developing its digital infrastructure. The library has committed to digitize the Henry Kissinger papers and has the funds to do so (as well as an imminent deadline). Despite the completion of several other archival digitization projects (primarily funded by donors), we currently have no adequate ingest, storage, preservation, discovery or access systems to support them, which is evidenced, painfully so, by the fact that Yale Library has a number of fully digitized archival collections that are not available online. Finally, our current finding aids database is at the end of its life and the library has committed to developing a new one. Combined, these factors make for a perfect opportunity to develop a holistic and integrated system for digitized archival collections.