In 2002, the University of Virginia Library began working toward the development of a Digital Collections Repository on top of Fedora. The development began with a set of assumptions which framed all work on the Repository. There were six key assumptions about the architecture:

- The Repository will be a part of a global network.
- All media and all content types will be integrated into one Repository collection.
- There will be simple objects and complex objects with many relationships, and we will need to manage the objects, their relationships to one another, and their relationships to their contexts.
- We will be faced with born-digital scholarship incorporating both digital materials and context.
- Any given resource can be associated with and presented in any number of contexts.

There were six key assumptions about services:

- The Repository will be a curated repository, where all content is selected using the same criteria as for the print collections.
- The UVa community is the primary users of the Repository. This does not mean that the Repository will be a closed one, but will support open access to content and public accessibility whenever possible.
- All of the UVa Library's digital collections will eventually be managed and delivered by the Repository.
- The Repository will be part of the solution to create a single point-of-access for the Library’s print and digital collections together.
- The Repository will have a public interface to support discovery and use of the collections by the UVa community.
- The Repository will provide tools for the use of the collections in instruction and research.
Specifications for functionality and delivery were documented in different ways. The content models for the different format types include specifications for behaviors that objects should be able to present, such as delivering subsets of their content or metadata, delivery of static files or on-the-fly transformations (such as raw XML delivery versus styled HTML), or supporting the download of image files. The functional specifications for the searches and the results and presentation formatting were documented in detailed screen-by-screen descriptions. Great care was needed to determine that the desired functionality was matched to behaviors in the underlying content models, and that the correct number and types of media files and metadata were present to support the behaviors.

The local mantra “If digital collections cannot be used, then they have not been preserved” was foremost in our minds when we set out the assumption that the Library had to develop both discovery services and end-user tools for the Repository. In 2003 a "Phase 1" prototype discovery interface was released for review by Library staff. The prototype was not operating on top of Fedora, but was a proof-of-concept to guide the Fedora Repository development. Input was solicited from Library staff on the design, functionality, and usability. The project team collated over 130 comments into 23 recommendations in four broad categories: User Interface, General User Functionality, Image-specific Functionality, and Text-Specific Functionality. The identification of a minimum set of contents was the top priority; developing hierarchical, thesaurus-based searching ranked last. Other highly rated priorities included search limits by format, collection, or topical set; cross-format searching; support for collection browsing in addition to searching; and the availability of both keyword searching and advanced fielded searching with Boolean operators for all discovery options.

The "Phase 2" Repository built on top of Fedora was released to Library staff and selected faculty in 2004 as a beta. What followed was a two-year iterative process where feedback was collected via web forms and email, focus groups were held with faculty and library staff, and faculty taught classes using the service and tools. A key part of the Phase 2 Repository was a digital object collector tool, now named “Collectus,” that allows users to create personal portfolios of objects. This is a Java application for the client machines (updated automatically via Java WebStart) that provides the ability to collect images and texts into personal portfolios and generate slide shows or electronic reserve websites that include pointers to the images and metadata in the Repository. The slide shows and electronic reserves deliver the images wrapped in an ImageViewer that allows zooming, rotation, and other on-the-fly image manipulation. As new formats are added to the collections the tool will be updated, functioning as a sort of combination shopping cart and low-level authoring tool for the Repository. Over the course of that two-year beta period feedback was constantly reviewed to prioritize redevelopment. The interface was updated, functionality was augmented, bugs were fixed, the server infrastructure was upgraded, collections were added, and workflows were stabilized. The production Digital Collections Repository became available to the full University community in 2007.

This process required close working relationships between the project leaders, the implementation team who were part of many Library units, and public services staff. The key skills were good communication – you can never communicate enough, even if it’s to say that there hasn’t been any progress – and the tangential communication skill of “translation” between the public services staff and the programmers. There were many requests for new services that
could not be accomplished in the identified time frame due to technical challenges. Specifications were not always documented at the level that the programmers expected. We often served as translators, iterating through feature request discussion to develop specifications for the programmers, and explaining technical challenges in an accessible way to the public services librarians to see if functional requirements could be changed.

The current UVa Library Digital Collections Repository collections consist of digital images, electronic texts, and EAD finding aids. Digital video, audio, printed music, datasets, and GIS are part of the Library’s collections, and migration of those formats is in various stages of implementation. Many of the collections come from over a decade of internal digital production, the creation of surrogates of the Library’s physical collections. Some are licensed from vendors. Some are born-digital scholarship created by faculty, often integrating Library materials. Some comes from open access sources, such as Federal and state datasets. All the objects, when brought into the Repository, bring relationships with them, whether simple relationships between media files and metadata, more complex relationships, such as that of page images to a text volume transcription, the relationships between issues of a newspaper, or more complex relationships still, such as the organizational context that a scholar overlays onto a digital archive in a web site. The objects and their relationships are part of the Repository.

What were our overall accomplishments? We outlined policies to build collections that increase access and use of our unique materials and provide faculty with what they want and need. We identified a set of circumscribed formats and minimum metadata standards to which all objects must adhere. We created a controlled environment that, in theory, simplifies our preservation tasks by minimizing the classes of objects that we must sustain. We put in place a potentially scaleable architecture with which to manage objects and the relationships among them, operating in a consistent, managed environment that makes the task easier to build discovery and delivery services, and tools for the use of the objects.

The current state of the Digital Collections Repository is one of review and reassessment while continuing to provide production services. An improved indexing and search service is in development using Lucene/Solr, which will provide greatly improved collection discovery and browsing. We are also defining “Levels of Service” for this and future repositories to guide scaleable infrastructure and service development. This will likely lead to a redefinition of existing content models and refactoring of their disseminators.

Success of a repository can only be assessed against the purpose that the repository serves in its operating environment; no repository can rated as successful unless it fulfills its purpose. The collections, services, and tools have been tested by our faculty and we have heard that we are giving them what they want – persistent, trusted collections that contain content that they find useful in their teaching and research, and the tools that they need to use them. These are the foundations for a sustainable repository and collection.