

Storage Practices at IU

The Scholarly Data Archive (SDA)

The Scholarly Data Archive (SDA) is the storage backend most heavily utilized by Indiana University Libraries and similar memory units within the IU system. SDA is a disk-cache front-end, with tape as the main long-term storage of files. It provides secure, well-managed long-term storage for inactive ("finished") content.

For technical information on SDA, see [this KB article](#) or contact [UITS Research Technologies](#). For assistance in actually depositing or accessing content on SDA, contact your local system administrator. **This page is specifically intended for non-technical content managers as a way to better explain the uses and potential challenges of relying on SDA for long-term storage.**

Accessing Content on SDA

Because the main storage mechanism for the Scholarly Data Archive is tape, SDA is not intended to be used for active storage. Files that you regularly use or want to access should be stored locally on your hard drive or departmental server. This section describes how to access content on SDA, in order to better elucidate how best to deposit content for discoverability and ongoing access.

In most cases to access content on SDA, you need to contact your local administrator for a copy. You will have to provide them with a filename so that they are able to retrieve the item. You or your administrator are not able to search SDA in the same ways that you can locate files on your local computer. Because of this, file-naming (see below) is extremely important. Accessing content also often takes a fair amount of time and computing power, so SDA is not optimal for a lot of small files. See [Organizing and Compressing Files](#) for more information on this.

Once an item is in SDA, you can easily access a few key pieces of information about it: SDA number (a number assigned based on deposit), MD5 checksum (to ensure fixity), date and time deposited, date and time last updated, and date and time of the last failure. This final date should be B:1969-12-31 19:00:00; if it is not, contact your administrator about the file.

Optimizing Deposit

File-Naming

Because SDA is not a bright archive intended for discoverability, it is not at all easy to search for files. In this case, a unique identifier should be given to each item being deposited into SDA (e.g., collection identifier + date; see [this page](#) for an example). This should be stored in a spreadsheet, database, or repository separate from SDA with other meaningful descriptive metadata that will enable you - or a future preservationist or user - to locate the content on SDA.

Organizing and Compressing Files

You can store items in SDA in a few different ways, depending on your anticipated future preservation /access needs and what type of content you are depositing. If you are managing large files, such as audiovisual content, you may wish to package all versions of the same audiovisual file (e.g., preservation, production, and access copies of a digitized media) into one compressed file to be deposited in SDA.

If you are managing small files, such as images, you may wish to organize them as a collection or as subcollections rather than depositing each individual file on SDA. This optimizes use of SDA and allows you to maintain collection context of your digital objects. If you decide to organize content in this way, however, you should keep good descriptions of all content contained within the final compressed file put in SDA, as it is not possible to search SDA for individual files contained within a deposited object. This information should again be maintained in a separate spreadsheet, database, or repository along with the other descriptive metadata that will enable you or your administrator to find and retrieve files easily.

Creating an Archival Information Package (AIP)

Based on the OAIS model and standard digital preservation practice, you should create an archival information package (AIP) for deposit into SDA. This provides a standard approach to the types of information that have been deemed necessary for long-term preservation of digital objects. We recommend using software like [Library of Congress' Bagger](#) to create bags, as they will be easier to check and integrate into new systems later on.