Pennsylvania State Archives
PENNSYLVANIA HISTORICAL AND MUSEUM COMMISSION

What Metadata Should I Collect?
Collecting Information about information.

Metadata is descriptive data that helps you quickly retrieve a scan when you need it. The process of collecting metadata is often called Indexing.

How you index your scanned documents will determine how easily you are able to search for and find them later. Indexing makes use of *metadata*, a set of data that describes or gives information about other data. The metadata you collect depends on the type of document you scanned and the ways you want to be able to retrieve it. For example, you might collect the following metadata for a birth record: name of newborn, date of birth, names of parents, and such.

You, or your scanning vendor, can collect as many different types of metadata as you would like, though the more metadata you collect, the more labor-intensive indexing you'll be doing.

Which Metadata are the Most Important?

Often the most valuable result of an imaging project is the ability to search and retrieve a document quickly and easily, so it's critical to identify the best metadata for each document type since that's what you'll use to search for documents. The most important index values are the ones that you'll find easiest to search by. Be sure to think about different search scenarios and not just your immediate needs. For example, in daily work you may simply search for student records by name or ID number of the student; years later, though, you may want to find all the records by date of graduation or birth year if, for example, the records may be destroyed based on the age of the student or their date of graduation. You'll only be able to search that way if you added that metadata when you scanned the records. Here are some examples of metadata you might collect at the time of scanning:

Student Records: student ID, name, date of birth, social security number, graduation date

Insurance Claims: document number, unique claim number, or policy number

Employee Records: employee name, social security number, or employee ID

Work Orders & Invoices: PO number, vendor ID, invoice date, or work order number