

Module 2: Data entry, Fluid-preserved

Module 2: Data Entry from Ledger, Card, Label, or Catalog Images

Task ID	Task Name	Explanations and Comments	Resources
T1	Navigate to the file folder in which image files are stored and load image of label to be transcribed.		
T2	Create verbatim data record from image file.	<p>Focus of data entry should be on preservation of verbatim data even when interpreted data are entered.</p> <p>In many institutions, data entry is governed by institutionally specific guidelines that address the fields to be captured and rules governing flexibility of technician interpretation. Some institutions opt for recording complete verbatim records, requiring technicians to populate all fields and transcribe all label data. Other institutions opt for incomplete or skeleton records that include a subset of the most important label data, to include:</p> <ul style="list-style-type: none"> • collection date, • collector name, • field number, • current determination, • higher geography, • geographic coordinates and datum, • locality name. 	<ul style="list-style-type: none"> • Technician, • Technology, • Standards-based database management system that includes taxonomic, geographic, and person authority files. • Institutionally prescribed data entry standards, • Clearly written and precise data entry protocols, • Standards for verbatim data entry, • Guidelines for data interpretation and in situ data cleaning, • Standards for data to be cataloged, • Measures for determining data content completeness.

		<p>Methods for data entry vary. Many institutions use keyboard entry. Others use voice recognition software. Optical Character Recognition (OCR) techniques are sometimes used, though sparingly in wet collections. Improvements to and use of OCR extraction and parsing are under study and are improving, especially for typewritten labels or cards that provide clearly visible text with sufficient resolution and text size. Strategies for improving the parsing typewritten label data is also underway.</p> <p>Data entry may happen directly into the database or through externally prepared datasets (typically spreadsheets) that are subsequently imported into the database using pre-existing database features.</p>	
T3	Clean/verify data, initiate escalations and notifications.	Variations of this activity might appear in more than one task, in more than one workflow, and/or in a separate data cleaning workflow. At a minimum, data entry technicians should immediately review data entry for each newly entered record to ensure accuracy against the label data and should correct as needed.	<ul style="list-style-type: none"> • Data entry technician. • Knowledgeable expert. • Institutionally specific data entry protocol.

		<p>Provisions should also be made to accommodate data correction and improvement at any time errors are noted.</p> <p>Methods for escalating data anomalies or technician questions and misunderstandings should be provided to allow technicians to notify collections managers, curators, or other knowledgeable personnel of problems as they arise or in batches, as specified in the institution's data entry protocol. Wherever possible, notifications should be escalated through electronic means to ensure efficiency and preserve a history of the notifications, e.g., as through e-mail or other electronic tracking system. In some instances, notifications may be made through personal communication.</p> <p>Immediate data quality reviews and editing might also be made by an available expert (curator, researcher, collection manager, or other knowledgeable person).</p>	
T4	Enrich data.	In some instances, data enrichment might be included wholly or partially here, or as part of one or more separate workflows devoted to other processes, e.g., Georeferencing.	<p>Technician.</p> <p>Clearly defined rules for enriching data (included within</p>

		<p>Data enrichment involves inserting or enhancing specimen data with additional data from related sources not stored with the specimen, e.g., field notebooks, monographs, published papers, etc. Enriched data that are not part of the original specimen record label(s) should be clearly noted. Enriched data might include:</p> <ul style="list-style-type: none"> • interpreted data (localities, field sites, etc., when known and not specifically included on the label), • data from field notebooks, monographs, reports, etc. • notations of misspellings of scientific, geographic, or personal names as shown on the label. 	institutionally specific protocol).
T8	Spot check for data quality at end of data entry session.	<p>Some institutions include a cursory data quality review by the entry technician or a proof technician at the end of or within a reasonable period following each data entry session. Selected data fields of a pre-determined percentage of randomly selected records are checked for accuracy and completeness. Reviewed fields usually include those important for search and display.</p>	