Identify “areas of interest”: splitting up primary labels from annotations, etc.

Methods: Digital or Human Cropping

Image Processing: Increase contrast, remove distortions, etc.
Method: Digital Processing

OCR (Optical Character Recognition)
one or more engines
ABBYY, Tesseract, etc.

Parse/Excavate the image into regions that correspond to DwC fields

NLP (Natural Language Parsing)

Create/Use Teaching Libraries, Data Dictionaries

OCR
one or more engines
ABBYY, Tesseract, etc.
Use Existing Collection Data or Data Dictionaries/libraries to help standardize parsed data

Digitized Label Data
DwC field, Value
DwC field, Value
DwC field, Value
DwC field, Value
Associated with original image(s)

Final Review before import into Collections Data Management