Digitization Workflow for Natural History Specimens

Collection Manager	EMu System Admin	Digitization Manager	LepNet Tech-Data Capture	LepNet Tech-Photo Capture	Collection Informatics

Specimen Image Geodata Data

1. Pre-digitization Curation
    a. Collection assessment - Follow checklist.
    b. Conservation/preparation for handling.
    c. Rehousing if needed.
    d. Gather field notes and collection ephemera.
    e. Examine, decrypt and update specimen labels and notes.
    f. *Barcode objects.
    g. Develop Digitization Plan:
        i. Parse label/catalog information and map to EMu.
        ii. Workflow design:
            1. Determine data capture interface,
            2. Determine photo views for objects/lots,
            3. Decide whether to georeference.
        iii. Construct data entry interface(s) (e.g., **EMu,** crowdsourcing, Google, etc.).
        iv. Write data entry tutorials and create record templates.
        v. Develop IPTC metadata pre-sets.
    h. Person-in-charge releases groups of specimens for Stage 2. [☐ **checkpoint**]

2. Information & Image Capture
    a. *Barcode objects.
    b. Capture image(s) of object or lot with label:
        i. Appropriate file naming of image,
        ii. Apply metadata pre-sets and add specifics,
        iii. White balance.
    c. Approval that image & metadata standards are met. [☐ **checkpoint**]
    d. Object assessment: [☐ **checkpoint**]
        i. If impaired, return to person-in-charge for conservation/repair.
        ii. If okay, return object to collection storage.
    e. Label transcription using appropriate interface(s).*(EMu; LepNet tech in house)*
    f. Georeferencing using determined method [optional based on 1.g.iii.2].
    g. Continue documenting databasing conventions (controlled vocab, field terms, etc.).

3. Information Processing & Ingestion
    a. Label and georeferenced data verification/validation (on an individual record basis):
        i. Check accuracy of transcription and parsing, correct as needed.
        ii. Correct typos and spelling.
    b. Data manipulation (performed on a batch of records):
        i. Remove unwanted characters (text gremlins).
        ii. Categorize and parse aggregated notes data.
        iii. Ensure consistency among terms & controlled vocabularies.
        iv. Check for duplicated records.
        v. Add EMu backend field names to column headers.
    c. Approval that data and georeferencing standards are met. [☐ **checkpoint**]

d. Image processing:
    i. Develop/refine specific procedures from test set (usually first 500 images),
    ii. *Validate IPTC metadata,
    iii. Generate DNG master & JPG derivatives,
    iv. Vet images for release to crowdsourcing or other platform (if applicable)
        1. Return to step 2.e for crowdsourced data capture. (N/A)
    v. Migrate images to appropriate batch processing folders.
e. Use Multimedia Import script:
    i. Import image derivatives into EMu.
    ii. Migrate all images to *MASTER_COLLECTIONS* folder.
f. Import data into EMu. (N/A)
g. Approval that data, images & geodata are properly imported. **[☐ checkpoint]**
h. Tidy image files & project ephemera, and archive procedure documents.

1. Deployment
    a. Decide scope of records and fields to be exported.
    b. Map EMu field to Darwin Core that are not already coded in EMu.
    c. Export data from EMu:
        i. Create export job (Crystal Reports or other mechanism).
        ii. Check and test export and data format.
        iii. Execute final export.
    d. Approval that data export is ready for public release. **[☐ checkpoint]**
    Note: Ensure EMu has assigned a globally unique identifier to each record.
    e. Upload .csv file to appropriate IPT.
    f. Check Darwin Core mapping in IPT.
    g. Adjust IPT metadata information.
    h. Publish (release) data. **[☐ checkpoint]**

* Flexible process, can occur in another stage if more efficient.