Digitisation of insect collection in NHM

Vladimir Blagoderov
Digitisation

• ≠ Imaging

• Most research projects need spatio-temporal records

• Different requirements for different purposes
NHM collection

- ~30,000,000 insect specimens
- 120,000 drawers
- 700,762 species
- >300,000 specimen records in KE Emu
- ~11,000 georeferenced specimen records
- >500 years to complete
- NHM digitisation challenge: 20,000,000 specimens in 5 years
Tier approach

- Taxon/unit tray/lot/specimen
- Species index – completed
- Prioritisation
- Research, loans: complete digitisation
- Specimens to be used: minimal information to be updated when the specimen is handled
- UIDs to identify physical objects
iCollections

- Digitising British Isles Lepidoptera collection
- ~500,000 specimens, 5,000 drawers
- Re-curation
- Specimen imaging
- Complete label information
- Georeferencing
- For use in Climate Change initiative
iCollections

Preparation
- Specimens and labels are put in individual trays

Imaging
- Canon DSRL and macro lens

Rehousing
- New drawers

Databasing
- Rapid Data Entry Application

Georeferencing
iCollections
iCollections

Flavia
Johanna
Elisa
Peter
Lyndsey
Sara
iCollections

- 4-6 people over 3 years
- Small tasks
- Team work
- Average imaging rate 163 specimen/day*person
- Average time for preparation, imaging and databasing slightly >3 min
- >£1/specimen
- Less then 0.1% of damaged specimens
- 6,800 person*years for the entire collection
SatScan

SatScan

+ Very fast imaging
+ No specimen handling
- Not individual specimens
- Just one view
- No label information
± Some information extracted from drawer image

Can we quickly extract images of individual specimens and annotate them?
SatScan
SatScan

Template Window

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlobalUniqueIdentifier</td>
<td>A Uniform Resource Name (URN) as a unique identifier for the specimen or observation record. In the absence of a persistent global unique identifier, construct one in the form: <a href="CollectionCode">InstitutionCode</a>.(CatalogNumber) Example: 1) umtisdmhmhuherp(12 2) FMNHMammal145712</td>
</tr>
<tr>
<td>DateLastModified</td>
<td>The last date-time of publication when any of the data for the record were modified from the previous publication of that record. When publishing a record for the first time, use the publication date-time. Returns values as ISO 8601 date and time</td>
</tr>
<tr>
<td>BasisOfRecord</td>
<td>A descriptive term indicating whether the record represents an object or observation. PreservedSpecimen - A physical object representing one or more organisms, part of organism, or artifact of an organism, synonym: voucher, collection, lot. FossilSpecimen - A physical object representing one or more fossil organisms, part of fossil organism, or artifact of a fossil organism. LivingSpecimen - An organism removed from its natural occurrence and now living in captivity or cultivation. HumanObservation - A report by a known observer that an organism was present at the place and time. StillImage - A photograph, drawing, painting. MovingImage - A sequence of still images taken at regular intervals and intended to be played back as a moving image; may include sound. SoundRecording - An audio recording. OtherSpecimen - Any type of specimen not covered by any of the categories above.</td>
</tr>
<tr>
<td>InstitutionCode</td>
<td>The code (or acronym) identifying the institution administering the collection in which the organism record is cataloged. No global registry exists for institutional codes.</td>
</tr>
<tr>
<td>CollectionCode</td>
<td>The code (or acronym) identifying the collection within the institution in which the organism record is cataloged.</td>
</tr>
<tr>
<td>CatalogNumber</td>
<td>The alphanumeric value identifying a record within the collection. It is highly recommended that each record be uniquely identified within a collection by this value. It is also recommended that each record be uniquely identified in a global context by the combination of InstitutionCode, CollectionCode and CatalogNumber.</td>
</tr>
<tr>
<td>InformationWitheld</td>
<td>Brief descriptions of additional information that may exist, but that has not been made public. Information about obtaining the withheld information should be sought from the administrative contact identified in the provider resource metadata (curator, collection manager). Examples: “specific locality information given only to nearest county”, “ask about tissue sample”, “geolocations given only to nearest degree”.</td>
</tr>
<tr>
<td>Remarks</td>
<td>Free text comments accompanying the object or observation record.</td>
</tr>
</tbody>
</table>

Add Key Fields: Save Template | Cancel
<?xml version="1.0" encoding="utf-8"?>
<Templates>
  <key>BMNHID</key>
  <value>A unique identifier for the British Natural History Museum</value>
  <key>GlobalUniqueIdentifier</key>
  <value>A Uniform Resource Name (URN) as a unique identifier for the specimen or observation record. In the absence of a persistent global unique identifier, construct one in the form: [InstitutionCode]:[CollectionCode]: [CatalogNumber] Examples: 1) urn:lsid:nhm.ku.edu:Herps:32 2) FMNH:Mammal:145732</value>
  <key>DateLastModified</key>
  <value>The last date-time of publication when any of the data for the record were modified from the previous publication of that record. When publishing a record for the first time, use the publication date-time. Returns values as ISO 8601 date and time</value>
  <key>BasisOfRecord</key>
  <value>A descriptive term indicating whether the record represents an object or observation.</value>
</Templates>
<Specimens>
  <Specimen>
    <DarwinCoreData>
      <key>ImageURL</key>
      <value>E:\test\T3\specimens\G1_2_0000.jpg</value>
      <key>BMNHID</key>
      <value></value>
      <key>GlobalUniqueIdentifier</key>
      <value></value>
      <key>DateLastModified</key>
      <value></value>
      <key>BasisOfRecord</key>
      <value>preserved specimen</value>
    </DarwinCoreData>
    <SpecimenIndex>0</SpecimenIndex>
    <ImageDimensions>
      <Left>519.635599159075</Left>
      <Top>1490.562857142857</Top>
      <Width>2247.9887876664334</Width>
      <Height>3511.8564285714283</Height>
    </ImageDimensions>
  </Specimen>
</Specimens>
</Project>
SatScan

- Specimen image
- Determination
  - Geographic region/country
  - Sex
  - Older/historical collections
  - UIDs
Generalised workflow application

- Create project
- Select type of digitisation
  - Multi-specimen or single-specimen
- Types of objects being digitised
  - Pinned insects, herbarium sheets, rock fossils, etc.
- Metadata captured
  - Databasing, label images, surrogates etc.
- Data fields
  - DwC
- Media storage
- UIDs

- Optimised workflow
- Bespoke digitisation interface for each project
- Support for backend data flow to EMu
Questions?