NATURAL HISTORY MUSEUM

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**

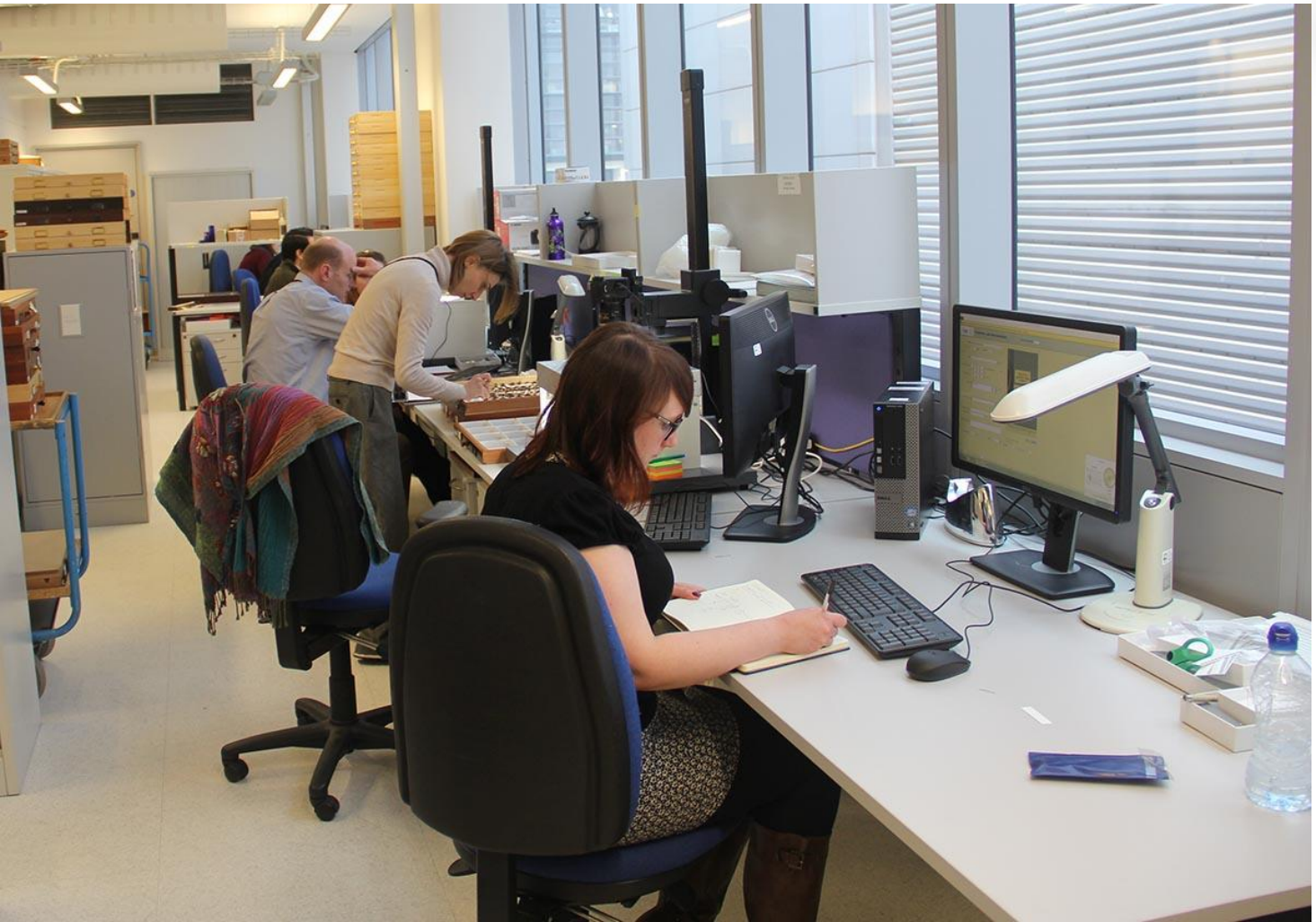
iCollections

- **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**

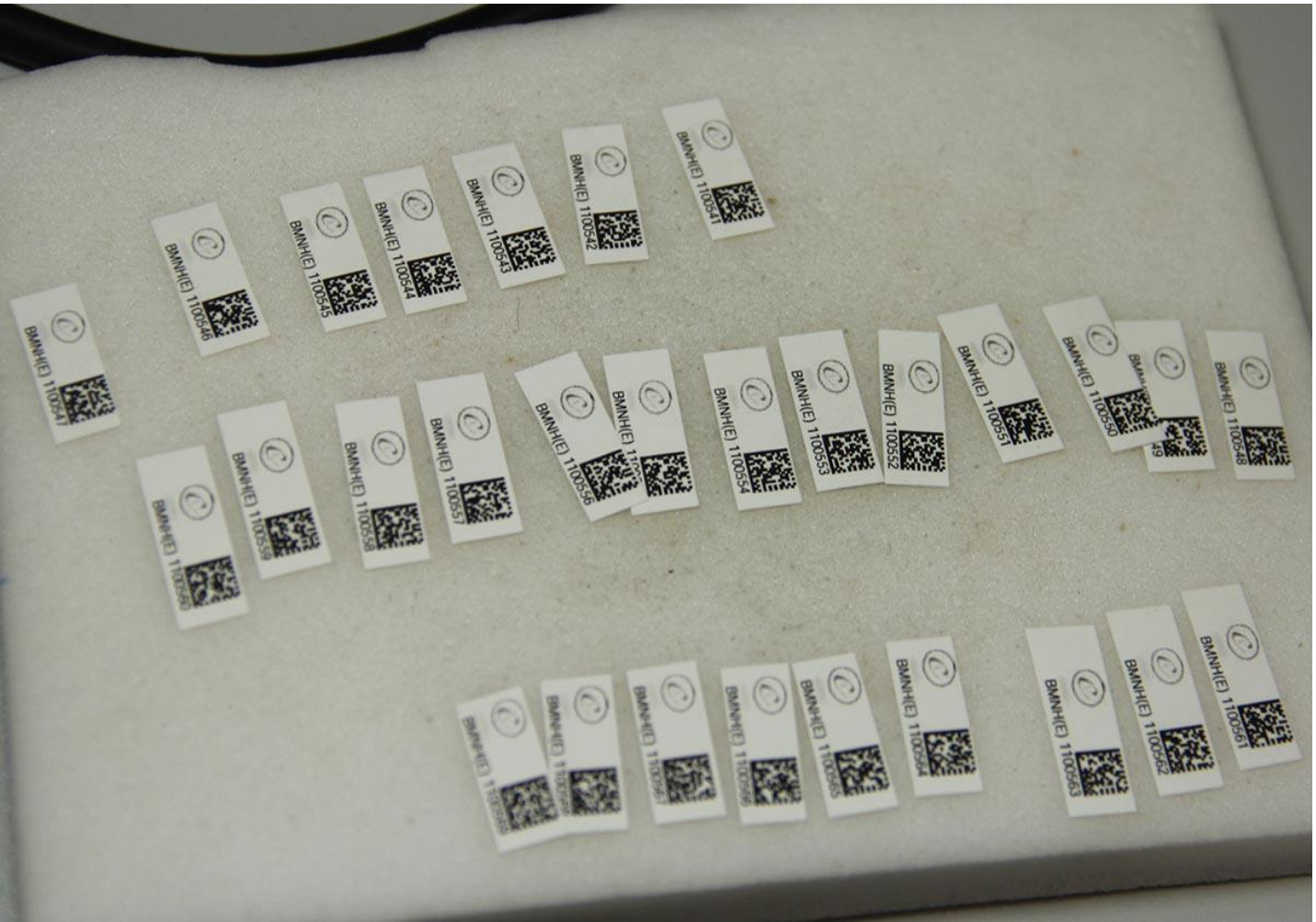
iCollections

- **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

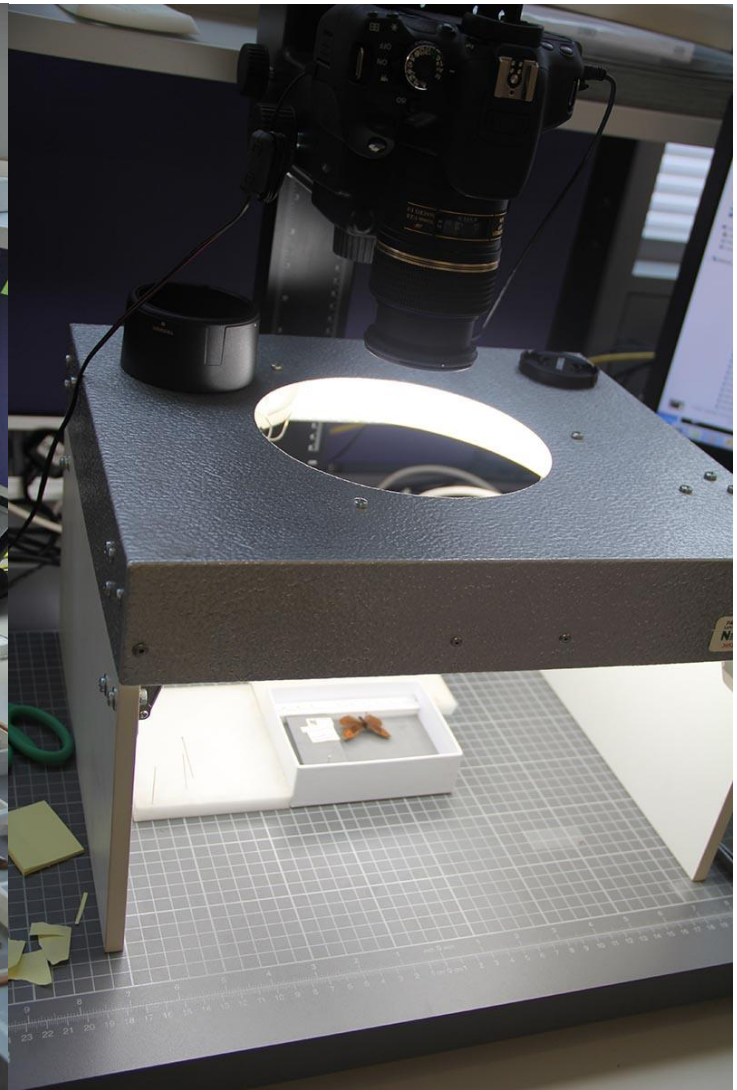


iCollections



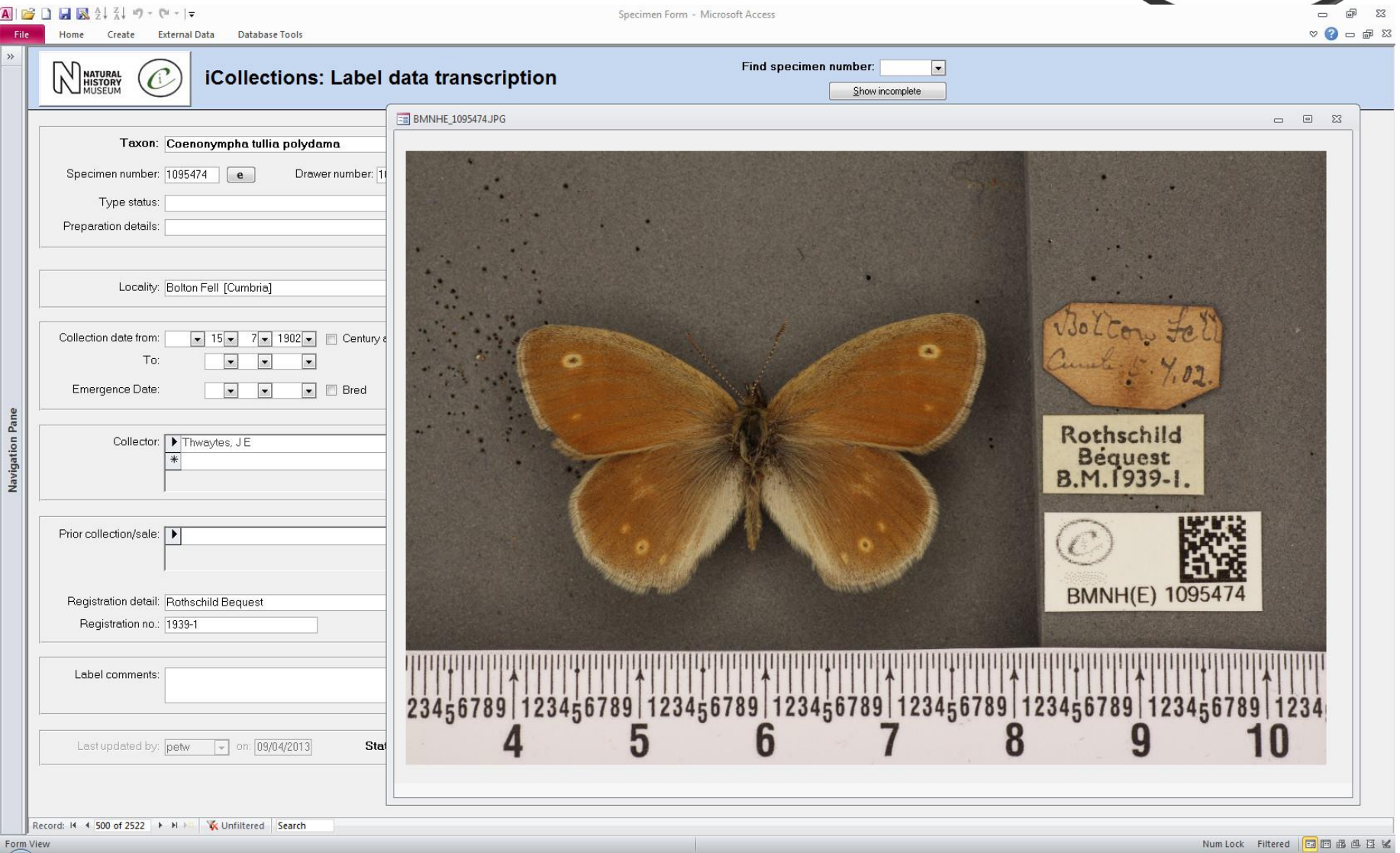
iCollections

iCollections

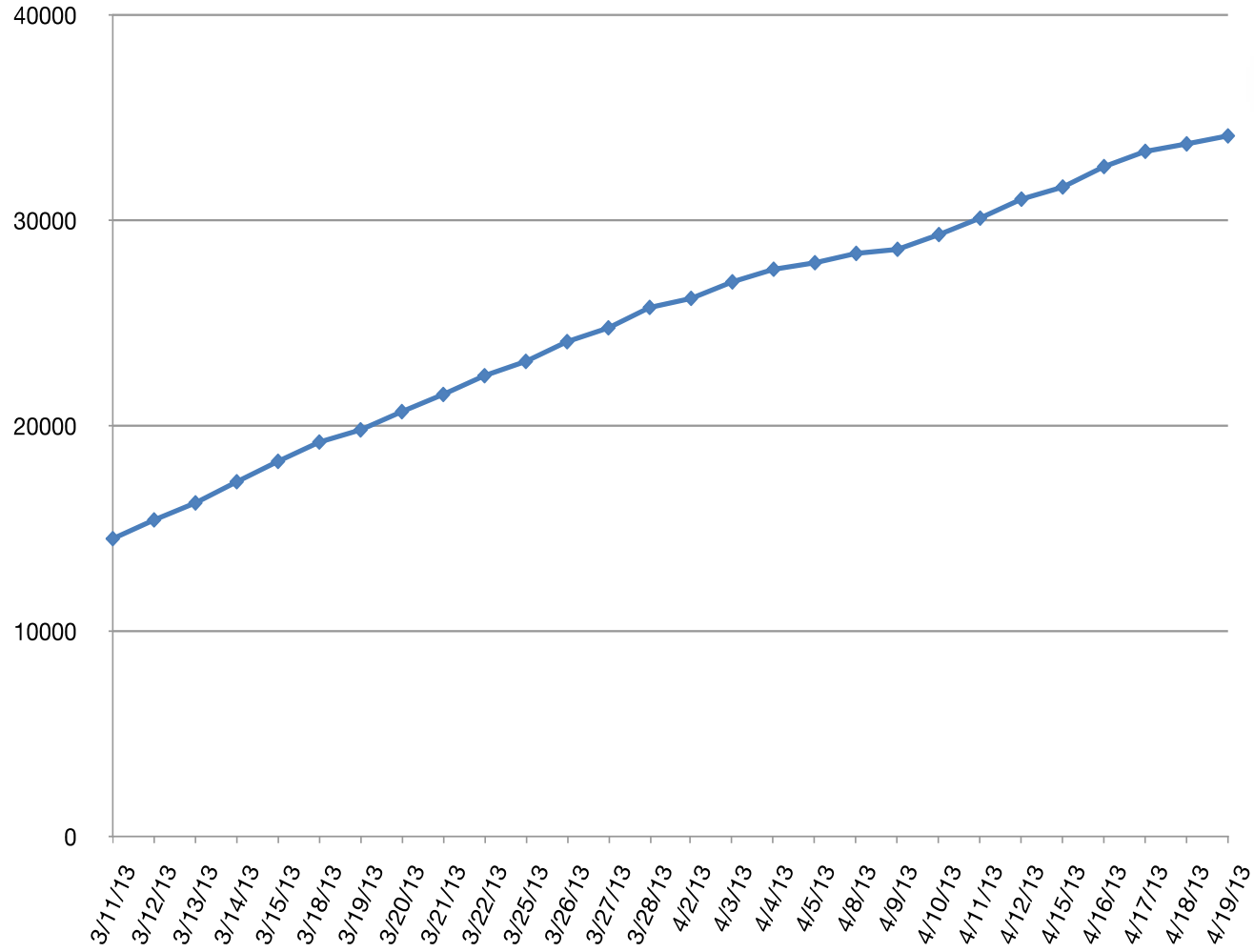


iCollections

iCollections



iCollections

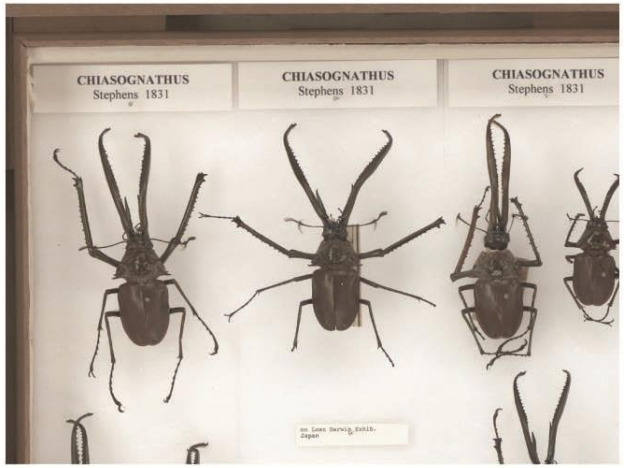


iCollections

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

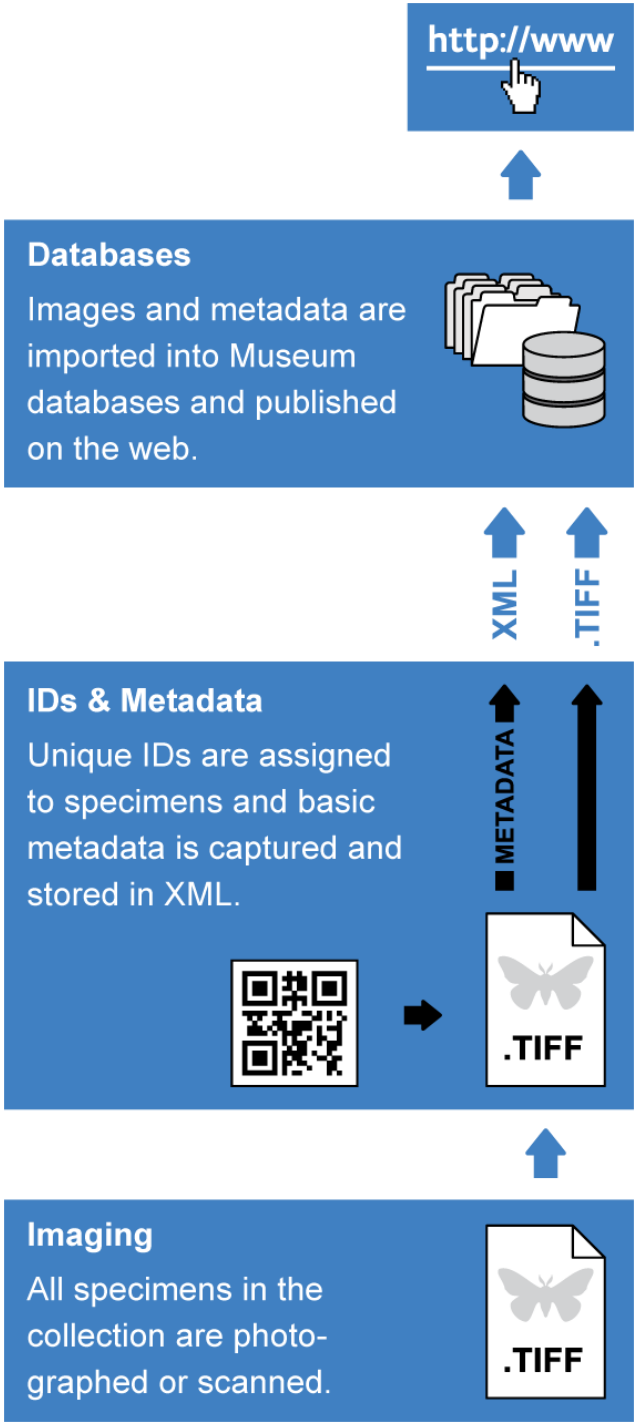


Blagoderov V., Kitching I., Livermore L., Simonsen Th. and Smith, V. 2012. No specimen left behind: industrial scale digitization of natural history collections. *ZooKeys*. 209: 133–146. doi: 10.3897/zookeys.209.3178

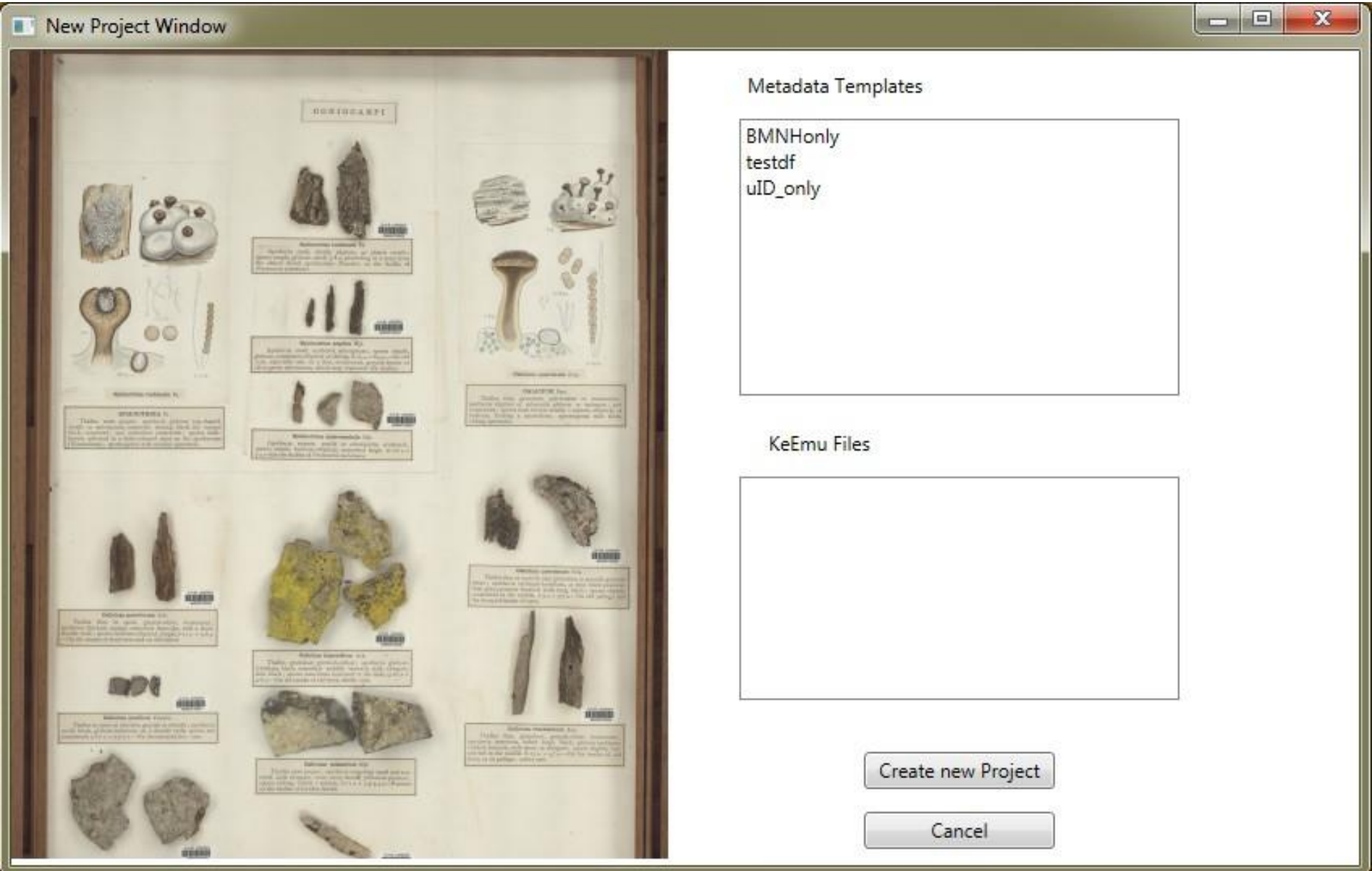
SatScan

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

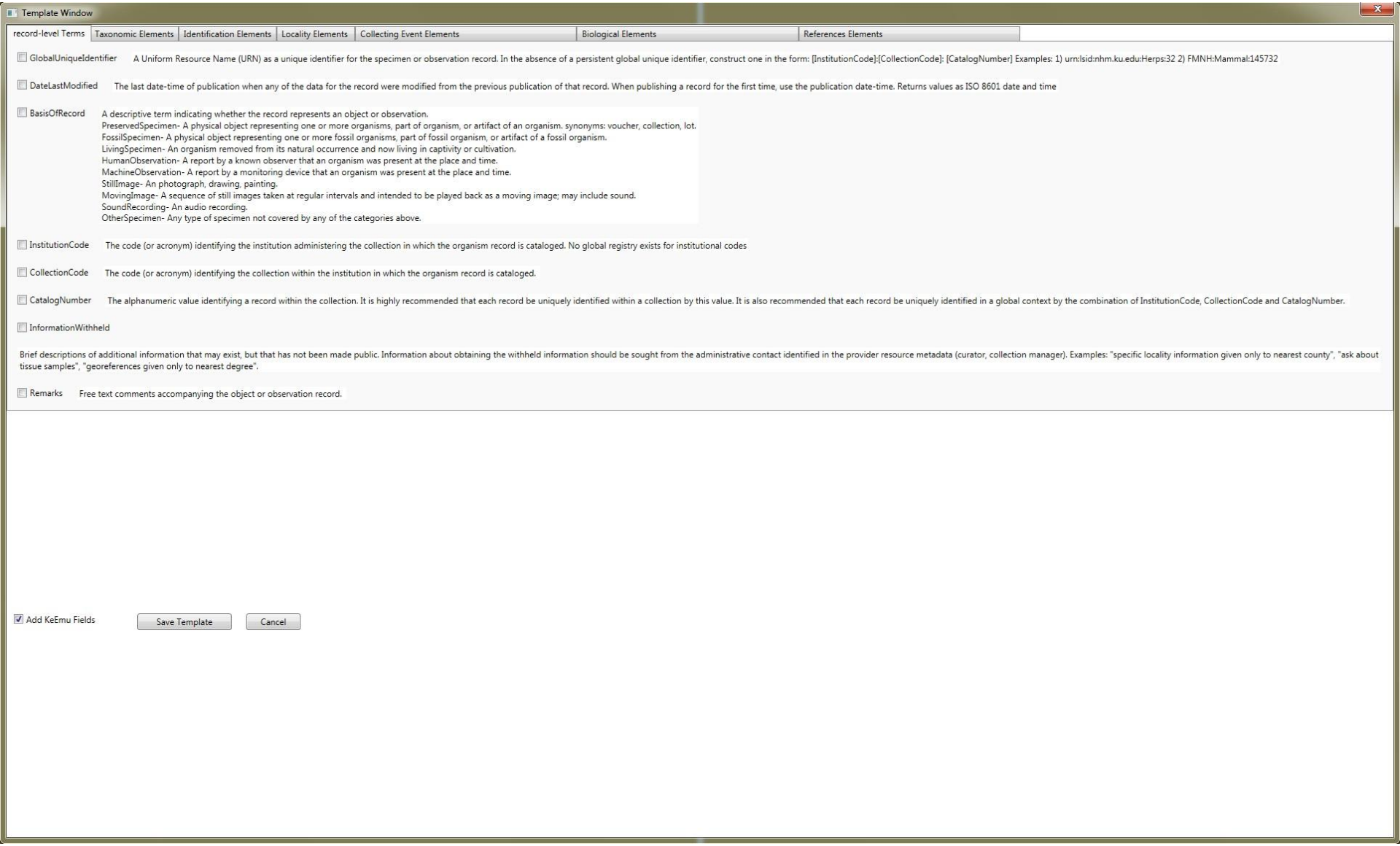
Can we quickly extract images of individual specimens and annotate them?



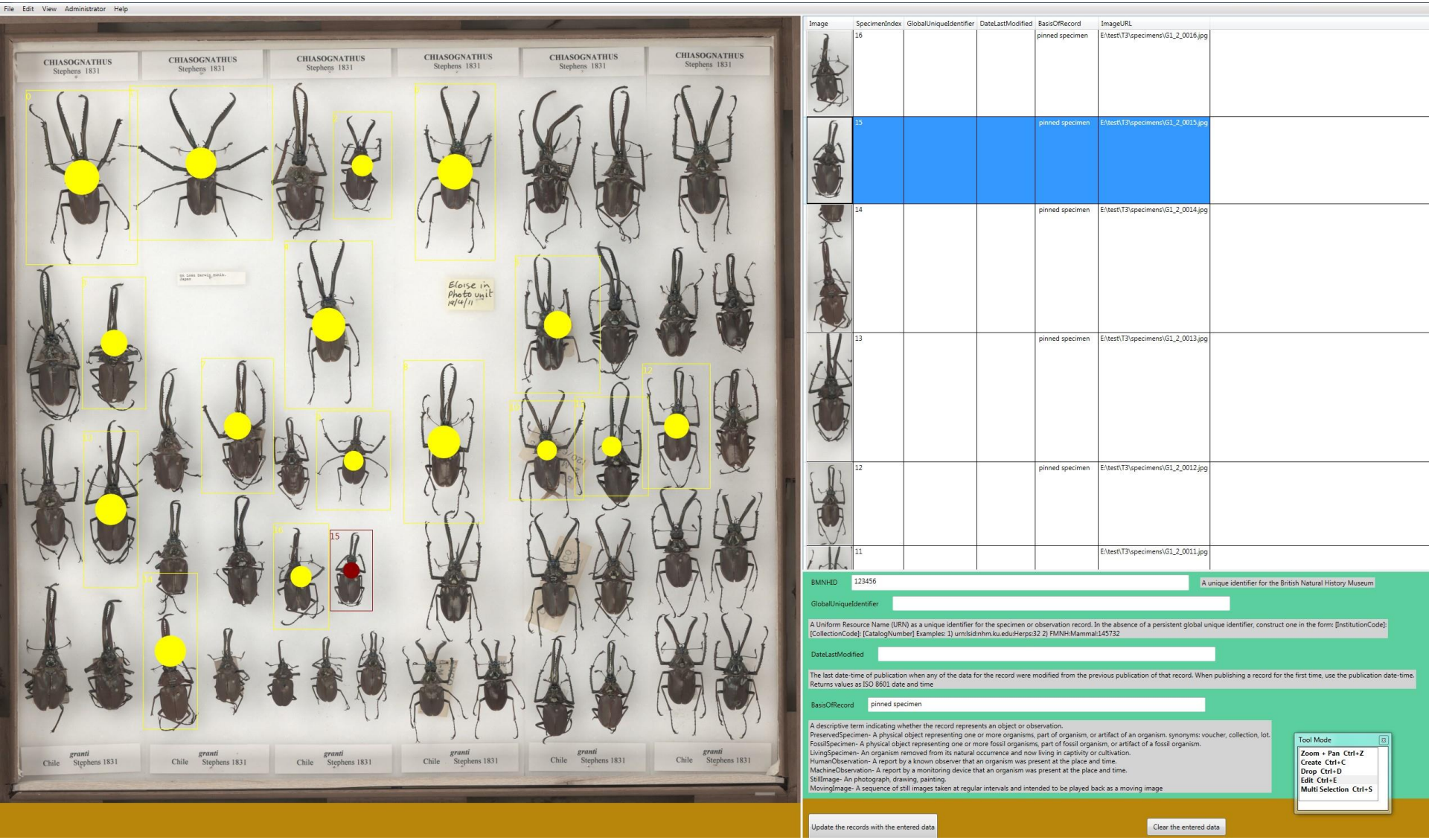
SatScan


New Project Window
Metadata Templates
BMNHonly
testdf
uID_only
KeEmu Files
Create new Project
Cancel

SatScan


Template Window
record-level Terms
Taxonomic Elements
Identification Elements
Locality Elements
Collecting Event Elements
Biological Elements
References Elements
GlobalUniqueIdentifier
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
DateLastModified
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
BasisOfRecord
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. synonyms: 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.
MachineObservation- A report by a monitoring device that an organism was present at the place and time.
StillImage- An 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.
InstitutionCode
The code (or acronym) identifying the institution administering the collection in which the organism record is cataloged. No global registry exists for institutional codes
CollectionCode
The code (or acronym) identifying the collection within the institution in which the organism record is cataloged.
CatalogNumber
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.
InformationWithheld
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 samples", "georeferences given only to nearest degree".
Remarks
Free text comments accompanying the object or observation record.
Add KeEmu Fields
Save Template
Cancel

SatScan


File Edit View Administrator Help
CHIASOGNATHUS Stephens 1831
granti Chile Stephens 1831
Image SpecimenIndex GlobalUniqueIdentifier DateLastModified BasisOfRecord ImageURL
16 pinned specimen E:\test\T3\specimens\G1_2_0016.jpg
15 pinned specimen E:\test\T3\specimens\G1_2_0015.jpg
14 pinned specimen E:\test\T3\specimens\G1_2_0014.jpg
13 pinned specimen E:\test\T3\specimens\G1_2_0013.jpg
12 pinned specimen E:\test\T3\specimens\G1_2_0012.jpg
11 E:\test\T3\specimens\G1_2_0011.jpg
BMNHID 123456
A unique identifier for the British Natural History Museum
GlobalUniqueIdentifier
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
DateLastModified
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
BasisOfRecord pinned specimen
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. synonyms: 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.
MachineObservation- A report by a monitoring device that an organism was present at the place and time.
StillImage- An photograph, drawing, painting.
MovingImage- A sequence of still images taken at regular intervals and intended to be played back as a moving image
Tool Mode
Zoom + Pan Ctrl+Z
Create Ctrl+C
Drop Ctrl+D
Edit Ctrl+E
Multi Selection Ctrl+S
Update the records with the entered data
Clear the entered data

SatScan

<?xml version="1.0" encoding="utf-8"?>
<Project xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<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> <Specimen>
.....
....
</Specimen> </Specimens> </Project>

SatScan

- **Specimen image**
- **Determination**

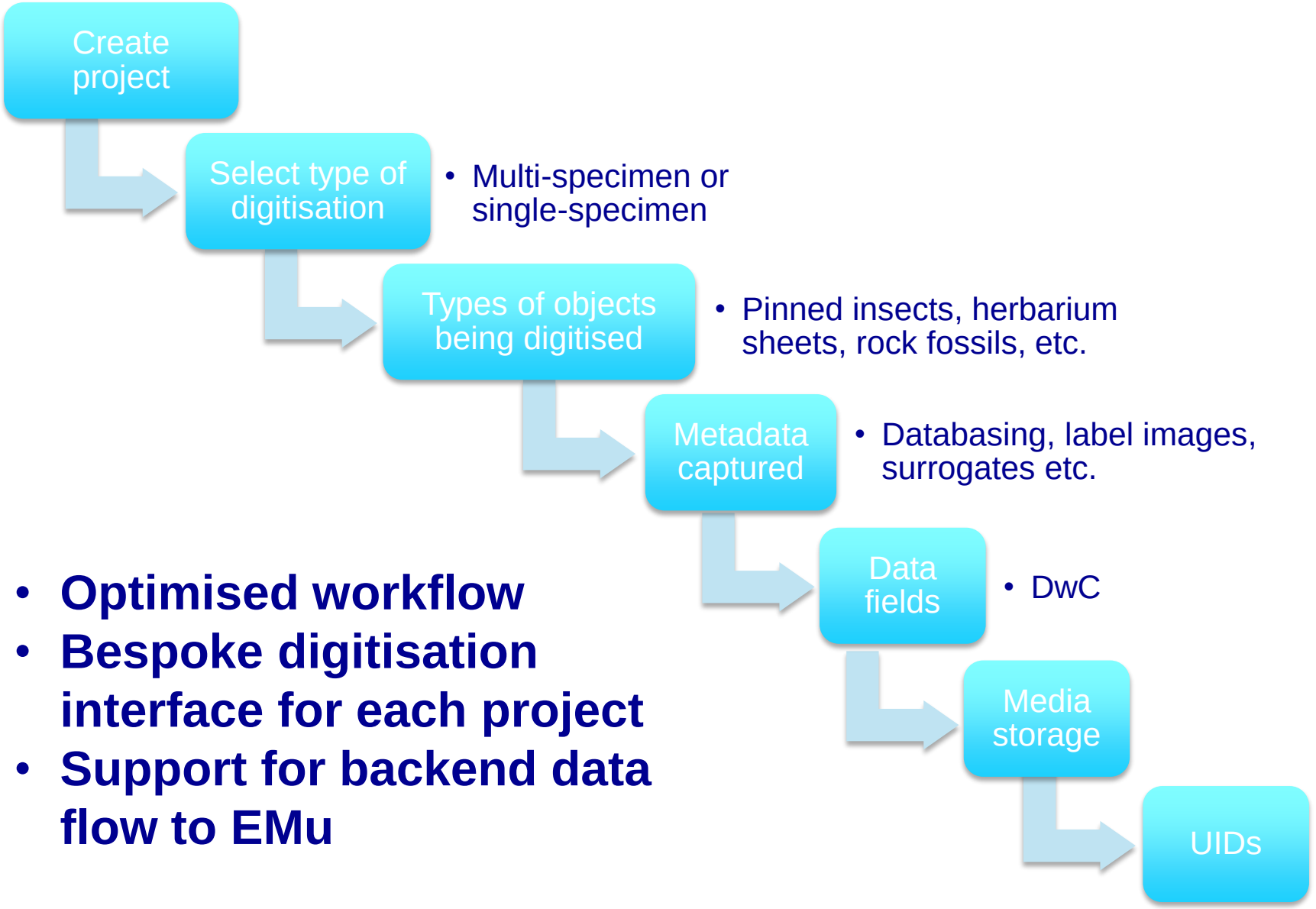
? **Geographic region/country**

? **Sex**

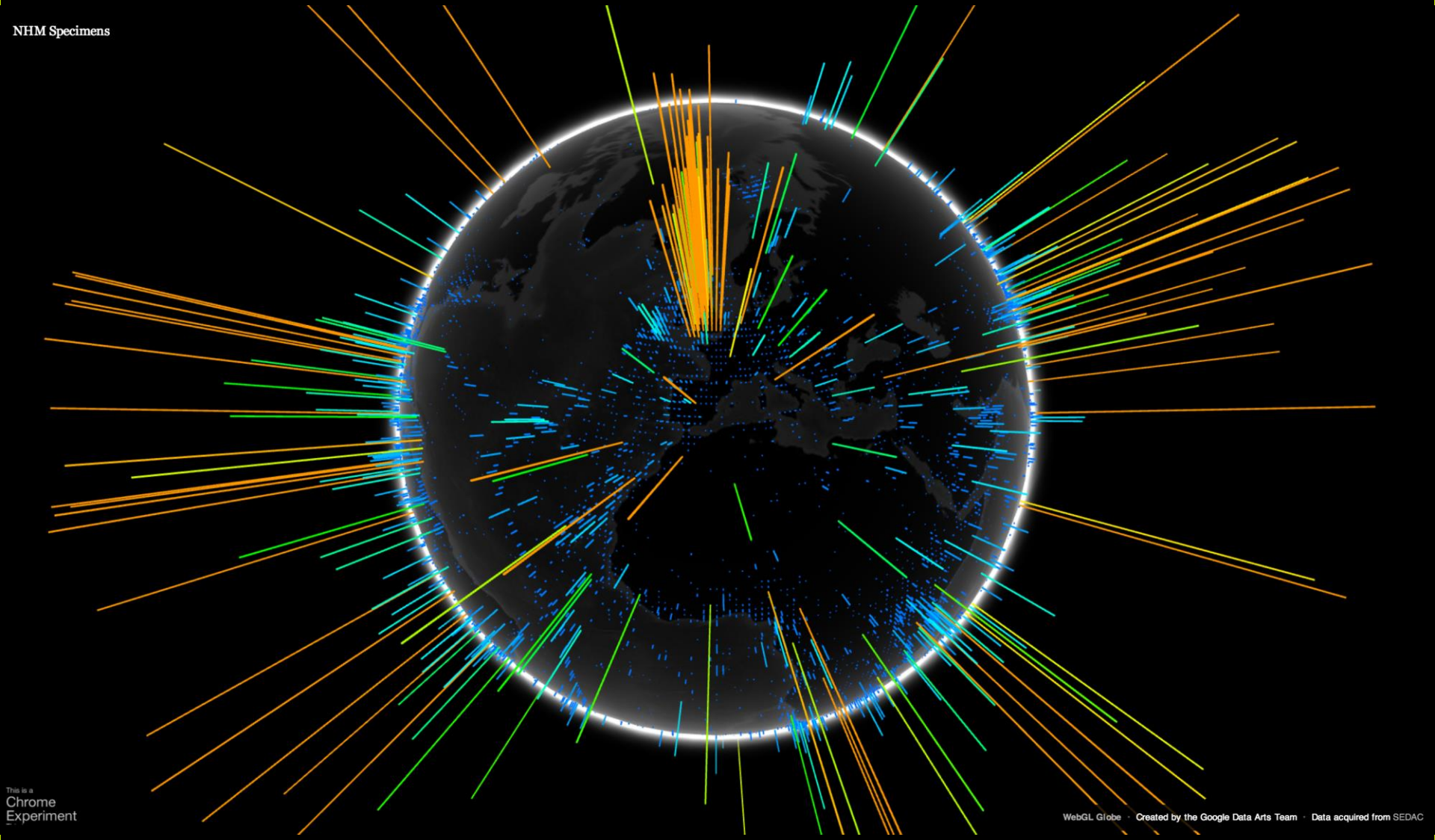
? **Older/historical collections**

! **UIDs**

Generalised workflow application



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


NHM Specimens
This is a
Chrome
Experiment
WebGL Globe · Created by the Google Data Arts Team · Data acquired from SEDAC

Questions?

NATURAL HISTORY MUSEUM