iCollections :Digitising the past for the future
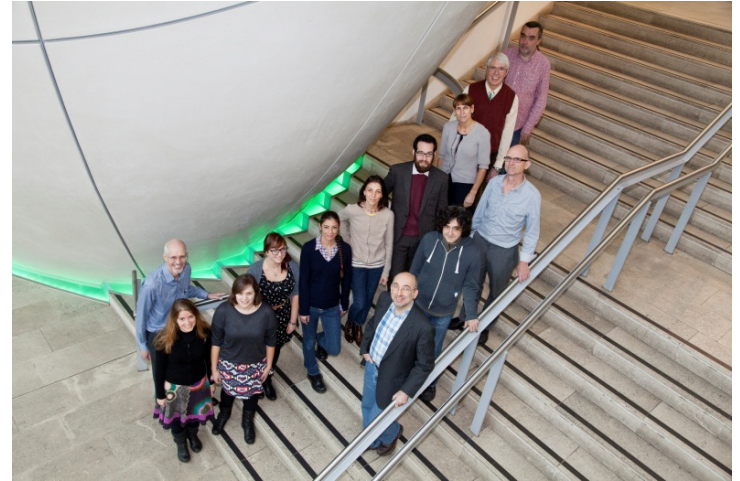
Lyndsey Douglas, Joanna Durant,

Elisa Cane, Gerardo Mazzetta,

Flavia Toloni, Peter Wing

Introduction

- iCollections team at NHM

- Trialling mass digitisation pipelines

- 170,000 British and Irish butterflies imaged so far

- Working on small collections including British orchids

Why digitise?

- Making collections relevant to modern science
- Increasing the accessibility of the data held in the collections
- Creating an international resource
- Assessing size and identifying any

  gaps in the collections

Choosing coherent collections

How we chose the collection to digitise:

- Entire collection
- Collection with high research potential –
    Phenology

    Morphometrics

    Wing pattern recognition

    Migration changes over time
- Recognised value of collection
- Collection with no existing pipeline

Preparation of collections

- Preparation done by curators before digitisation

- Estimation of collection size

- This optimises workflow and consistency

- A small amount of re-curation is done by the digitising team

Digitisation of collections

- Advantages for the management:
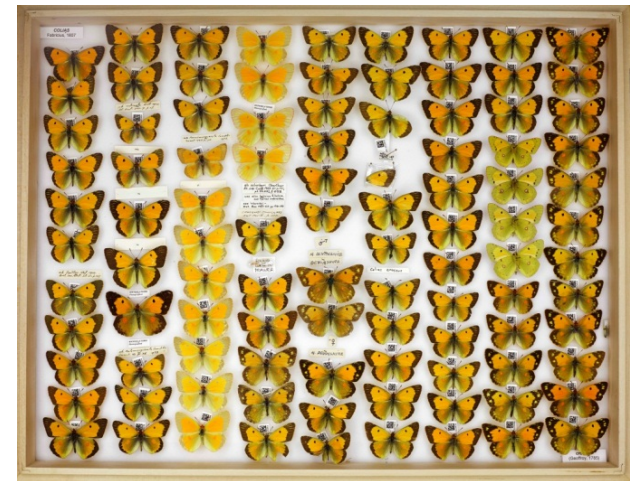
    Efficiency

    Preservation

    External requests

    Unique barcodes

    Updated taxonomy

    Re-curation

    Re-housing

Workflows

- Team works in pairs to maximise efficiency
- Preparation – imaging – re-housing – transcription

Resources

- I.T. is a finite resource!
- Technology choice: Bespoke MS Access 2010

  -Flexible and standard technology

  -Simple and flat interfaces - high imput - Intergrate web tools

  -Effective and consistant

- Enable transfer of images and metadata to the CMS KeEmu data repository

Minimise the impact on the digitising team and reduce potential errors

Maximise efficiency and quality of data capture workflow

Normalisation

- Conforming data to a "normal" structure
- Interfaces for taxonomic names and localities

Geo-referencing

- Providing latitude and longitude co-ordinates
- Aims to rationalise locality information
- Currently 7,500 site localities
- Enables GIS software to map species distribution over time
- Facilitates study of relationships between species distribution and changing environmental conditions

Geo-referencing
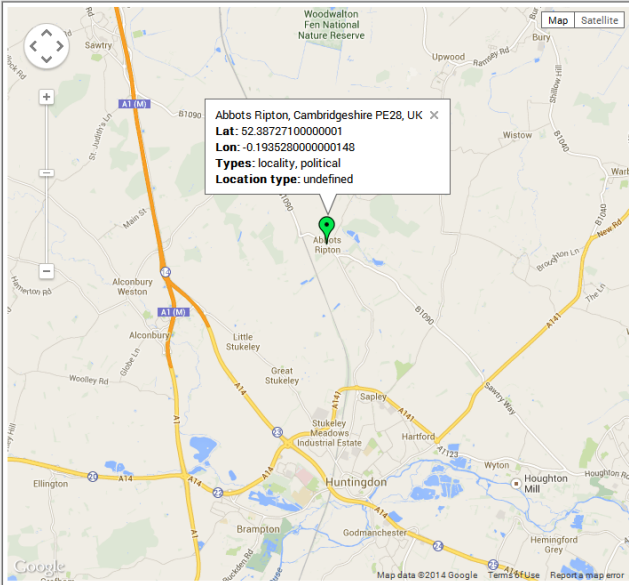
Communication

- Digitisation is a team effort
- Weekly project meetings
- Digitisers are central!
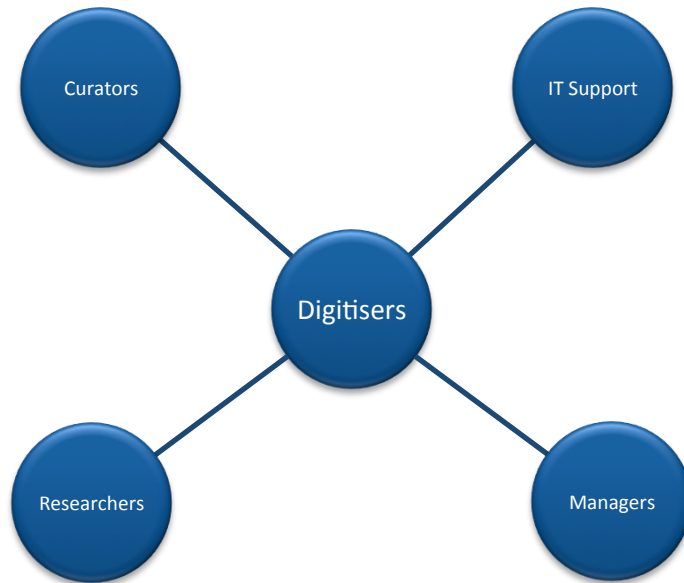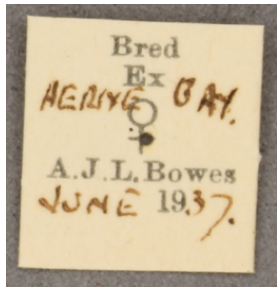
Communication
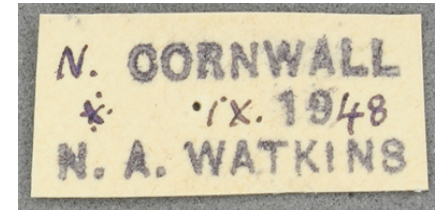
- Bred specimens flagged during transcription
- Appropriate field to record this information
- Can be removed from final dataset (phenology)
- Label interpretation lies with curators

Quality control

- QA processes for taxonomy and localities
- The project aims to set up a minimum standard for transcription and imaging
- What rate of error do we consider acceptable?
- Some conflictions are in t[ ]themselves!

B.3.7.35
G.H.Sale
CoxBench.
Torquoy. O.

BMNH(E) 1291106

Lessons learned

- Digitisation process should be documented with written protocols for each stage

- A unique back-end database and transcription interface is vital

- Identifying and solving problems

- Space for the digitisers to work!

Going forward

- Time management and communication are the biggest considerations for this type of project

- One digitisation team responsible for carrying out several projects across the NHM

- How to carry out more than one project at a time... redistributing the work force?

Conclusions

- Prioritise coherent, scientifically valuable collections
- Don't underestimate the importance of:
  - Geo-referencing
  - Curatorial support
  - QA for locality and taxonomy
  - Communication
  - Time management

Any Questions?