



# **Data Scoping Project**

**Biological data collections relating to Scotland  
held by  
Dundee Art Galleries & Museums  
and  
The University of Dundee**

## **Final Report**

**Researched and compiled**

**by  
Robert W H Scott**

**For  
BRISC  
2009**

## **Contents:**

<b>Project Report</b>	<b>p.3</b>
<b>Introduction</b>	<b>p.3</b>
<b>Methodology</b>	<b>p.5</b>
<b>Results</b>	<b>p.7</b>
<b>Analysis</b>	<b>p.11</b>
<b>Appendix I: Dataset Key</b>	<b>p.13</b>
<b>Appendix II: Dundee Art Galleries &amp; Museums Datasets</b>	<b>p.16</b>
<b>Dundee Art Galleries &amp; Museums Botany Collection</b>	<b>p.16</b>
<b>Dundee Art Galleries &amp; Museums Invertebrates</b>	<b>p.21</b>
<b>Dundee Art Galleries &amp; Museums Vertebrates</b>	<b>p.40</b>
<b>Dundee Art Galleries &amp; Museums Voucher Specimens</b>	<b>p.67</b>
<b>Appendix III: University of Dundee Datasets</b>	<b>p.122</b>
<b>University of Dundee Herbarium</b>	<b>p.122</b>
<b>University of Dundee Zoology Collection</b>	<b>p.129</b>
<b>Appendix IV: Project Times</b>	<b>p.136</b>

**BRISC gratefully acknowledges matching financial support from Scottish Natural Heritage (SNH) and significant in kind contributions from curators at the Dundee Art Galleries & Museums and the University of Dundee Museum Collection, without which this pilot project could not have been carried out.**

March 2009

# Project Report

## Introduction

BRISC (Biological Recording in Scotland) is a voluntary body of Scottish Charity status with the stated constitutional objective to advance the education of the public about biological diversity and to promote, for the benefit of the public, the conservation of the environment. One specific objective relevant to this project is:

*“To promote the effective gathering and use of biological data and other related information as a means to develop an understanding of Scotland's biodiversity and contribute to the conservation of the natural heritage of Scotland.”*

The present research project was conceived to establish what relevant biodiversity data are held by Scottish museums and universities and to publicise the existence and locations of these datasets, some of which may not currently be part of the recording process (i.e. not sent to national recording schemes, nor published in books or journals), in order to enable wider access to this information.

Historical data are important in establishing baseline information on species, and long term datasets can identify changes in distributions and abundance. The findings would also help to prioritise which datasets should primarily be made digitally available to the wider public.

To support this objective a grant was successfully obtained from Scottish Natural Heritage, and it was decided to run a pilot project as a first step. Staff of Dundee Art Galleries & Museums and the University of Dundee Museum Collections were approached and generously agreed to become partners in the project. A contractor was employed to work with the curators and other staff to draw up metadata for the data collections held by these bodies. The metadata details were to comply with the National Biodiversity Network (NBN) standards, and the outcome to be made available both electronically and as a limited number of printed reports. Choosing just one location initially, also made it easier to find a local contractor.

The lessons learnt through this pilot project can be used to plan the best future approach. It also provides an estimate of the time and costs involved in drawing up similar projects for other biological datasets held by Scottish universities and museums.

Throughout the course of this project, the employed contractor worked for a sum total of 224 hours, between the 8<sup>th</sup> October 2008 and the 9<sup>th</sup> April 2008.

Information of particular interest to this study included: summaries of the types

of species within each collection; information on the geographical location where specimens were collected; date ranges during which the collection was gathered; the size of each collection, and more general information regarding each institution's data-sharing policies regarding their collections.

There are several accessibility limitations currently facing museums. At the core of these is a somewhat paradoxical arrangement, which ties together both the issues of physical access limitations, and the difficulties of increasing public awareness of the collections. It is becoming increasingly apparent that many potentially interested bodies, be it in a professional or purely recreational capacity, are unaware of the true extent of many museum collections. Yet, as there is little information available from outside the institution regarding what is housed within it, many of these potential users are not aware that indeed they would be interested at all. There have been a number of previous efforts which have attempted to resolve this problem. One that is of particular interest to this project was the attempt by H. E. Stace, C. W. A. Pettitt, and C. D. Waterston in 1987 to catalogue all natural history collections held within UK institutions; titled "*Natural Science Collections in Scotland (Botany, Geology, Zoology)*". More recently, there have been strong initiatives towards providing public access to collections catalogues through the internet. It is such initiatives, aided by projects such as this pilot study, which will help to break down such barriers to collections accessibility.

The University of Dundee Museum Collections is tasked with managing a diverse array of collections. Some relate to the history of the university, while others have been collected for use as teaching aids throughout the university's development and expansion. Information regarding the collections is available by contacting the curator, and although raw data can be extracted from the collections database, this is limited in its extent. The curator is currently looking into means of making the catalogue available online.

The McManus Galleries & Museum (now named Art Galleries & Museums), which opened in 1867, is a Local Authority museum for the Dundee area. Since 2005, the Dundee Art Galleries & Museums building has been undergoing an extensive redevelopment program, with most services being transferred to their Collections Unit in Barrack Street. Unrefined data from the collections database can be made available by the curatorial staff, although more in-depth collections information is not, necessarily, instantly available. There is a current project in place to transition collections data to an online catalogue.

At present data from both institutions are only available by onsite access.

## Methodology

The initial stage of the project involved extracting the currently digitised raw data from each institution's collections database. This was necessary as, by exporting the specimen records to a spreadsheet application such as Microsoft Excel, calculation of metadata fields such as the "Start Date" and "End Date" would be much faster. Similarly, exporting the data would make it far easier to sort specimens by collector, and thus break the data down into individual collections. As the two institutions utilise different software packages to organise their collections, a different methodology was required in each individual case.

The Dundee Art Galleries & Museums currently utilise Adlib Museum Edition (v6.3.0 build 391) for administrative purposes, which fortunately allows direct exporting of specimen data to Microsoft Excel spreadsheets. In addition, some of their records have been digitised in the past using Recorder (v6.10.4.120), which allows for a similar method of exporting data.

In the case of the University of Dundee Museum Services, the software being used for collections cataloguing is INCA (v2.7). Although INCA technically allowed exporting of raw data via Excel spreadsheets, due to unanticipated computer compatibility issues, it was eventually decided that the data should be exported as CSV (Comma Separated Value) files, which were subsequently imported into Microsoft Excel 2008.

To facilitate the process of exporting data, records were exported from AdLib, and INCA in several separate batches. In the case of Dundee Art Galleries & Museums data, all the data stored in Recorder was exported as a single spreadsheet of records, followed by separate 'Botany', 'Vertebrate' and 'Invertebrate' groupings as separate instances. In the case of the University of Dundee collections data, it was exported from INCA as 'Herbarium' data, and as 'Zoology collection' data. It was decided that in the final report format, the data would remain separately organised into these groups to facilitate the process of data analysis.

Once the initial raw data had been exported as Excel spreadsheets, a template for the final collections datasets was devised utilising the GEMINI v2.1 Metadata standard. Yet, before the final datasets could be constructed, it was necessary to define more specifically the concept of what constituted a 'collection' within the parameters of this project. Eventually it was determined that a collection would be identified by a series of different qualifiers. Firstly, any specimens recorded as being part of a named collection (e.g. the Kenneth Todd collection which is housed within the University of Dundee), were regarded as part of one discrete dataset. Secondly, the remaining specimens were grouped by the name of their original collector, and instances of 10 or more specimens collected by the same individual were regarded as discrete datasets. Finally, all specimens which did not qualify for either of these two defined groups were placed in a single 'general collection' dataset. Although it could be argued that the initial groupings of records that the data was exported in (such as the University of Dundee

Zoology Collection) could be regarded as collections datasets, it was felt that by breaking the collections down further into groupings by collector, the resulting metadata would be far more meaningful, and provide a greater understanding of the data held by the participating institutions.

One of the major issues encountered regarding the methodology of this project involved prioritising the collection of data. It was realised very early on that the time constraints of the project would not allow for complete analysis of all the data associated with the collections housed at both institutions. As a result it was determined that by prioritising already digitised data, over written records, the highest return of information for time would be attained within the project time-constraints.

The majority of other obstacles which were encountered related to technical problems caused by the design variations between the software packages being used. One such problem was an intercompatibility issue resulting in Microsoft Excel misinterpreting the structure that INCA and AdLib use to denote specimen catalogue numbers. In the case of INCA, individual parts of a single object are recorded under the same number, differentiated by a '/', followed by a part number. Similarly, Dundee Art Galleries & Museums utilise a '-' to separate specimen number and part number. The result of this practice is that Excel interprets the data as a mathematical function upon import, so for example, specimen number 1234/2 in INCA becomes number 617 in Excel. Unfortunately, a solution to this problem was not forthcoming during the course of this pilot study, and consequently this would be a significant factor which needs to be addressed before undertaking future projects of a similar nature.

Another problematic idiosyncrasy in Excel's programmed functionality is that it forcibly attempts to implement a pre-defined formatting structure (of day-month-year) upon any cell that it interprets as a date. This was potentially problematic, as not all date entries within each system proved to have been entered using the same date format. Therefore in rare instances, data could occasionally be scrambled by Excel when copied and pasted directly. This issue was resolved by typing in the date entries manually, along with verifying the format settings of spreadsheet cells.

## Results

This study resulted in a sum total of 129 datasets, representing 13 collections housed at the University of Dundee, and 116 collections housed by Dundee Art Galleries & Museums. The finalised datasets produced by this pilot study are included, in their entirety, in the appendices of this report. They follow the format outlined by the GEMINI v2.1 Metadata standard. Definitions of each field's contents, alongside the corresponding NBN field headings are provided in Appendix I: Dataset Key.

In some cases, particularly those of the University of Dundee Herbarium collections, not every record relates to a single specimen, as these collections are not yet recorded to a specimen level. This has been noted in the abstract of relevant collections.

Several resources at the Dundee Art Galleries & Museums could not be digitised due to project time constraints, thus making it impossible to provide metadata regarding these collections. For the purposes of the time estimates to digitise these resources, it is assumed that each record will require 10 minutes to transfer into the institution's database. Of those collections housed by Dundee Museums & Art Galleries, the 'James Wardrope Collection' and the 24 cabinets of undigitised insect specimens are probably the highest priority collections to be digitised, as they contain information of actual specimens housed by Dundee Museums & Art Galleries, as opposed to simple records of survey data.

- The 'James Wardrope Collection' of insects are housed in 4 cabinets, containing a sum total of 72 drawers. The collection is believed to contain in the region of 9624 specimens, which should require an estimated 1604 hours to digitise.
- An additional 24 cabinets of undigitised insects of various origins are housed alongside the Wardrope Collection. The 159 drawers within this collection should hold an estimated 12,720 specimens, which should require in the region of 2,120 hours to digitise.
- The 'Dundee Urban Wildlife Project Site Review', carried out between 1992 and 2000, constitutes to data for a total of 55 discrete sites. Within these folders are an estimated 5717 individual species records. Digitisation of this data should require an estimated 952 hours.
- The 'Scottish Wildlife Trust Site Surveys' of Angus carried out between 1993 – 2000 identify a total of 31 sites, estimated to contain 6722 records of individual species. This will require an estimated 1119 hours of work to digitise.
- A series of 'Dundee City Council Urban Habitat Surveys' carried out in 2000 provide an additional 1680 individual records. These should require 280 hours to digitise.
- A series of butterfly survey records collected between 1970 and 1999 were estimated to contain 411 records. These should require roughly 68 hours to digitise.
- A series of records from surveys of insects between 1970 and 1989 yield a total of 310 individual records, which should require a total of 52 hours

to digitise.

- The 'Sidlaw Hills Invertebrate Pitfall Trapping Project' of 1995, carried out by Dundee Museum & Art Galleries contains an estimated 715 individual records, which should require roughly 119 hours to digitise.
- The 'Balgavies Loch Invertebrate Survey', carried out in 1992 contains an estimated 5919 records. It should require in the region of 986 hours to digitise all this information.
- The 'Report of the Survey of Beetles (Coleoptera) recorded from sites in Angus and Dundee' was carried out between 1991 and 1992 in Montrose Basin LNR, Angus. It contains an estimated 121 records, which should require roughly 20 hours to digitise.
- The 'Lepidoptera collected in Kirkinch, Perthshire 1981 – 87' Collection contains an estimated 663 discrete records, which should require roughly 110 hours to digitise.
- The 'Nature Conservancy Council Deeside survey (1970 – 1984)' contains 44 records, requiring an estimated 7 hours to digitise.
- The 'Nature Conservancy Council Aberdeenshire survey (1978 – 1985)' contains a total of 11 records, requiring roughly 2 hours to digitise.

It was estimated that the remaining collections at Dundee Museums & Art Galleries would result in roughly 120 more datasets after completion of the digitisation process (assuming that each individual site within the survey data constitutes to a discrete dataset). Assuming that it requires 15 minutes to collect the corresponding metadata for each dataset, roughly 30 hours would be necessary to fully analyse the remaining Dundee Museums & Art Galleries collections.

Of the, as yet, undigitised collections at the University of Dundee, the most significant is probably the 'J. L. Colville Collection', housed in the University Herbarium. Primarily of Scottish origin, the Colville collection constitutes to roughly 50 folders of specimens, each containing on average 30 sheets of specimens. Of these 50 folders, 32 have had their corresponding information transcribed onto computer, although none of this data has been incorporated into the University's collections database. A point of interest is that a number of specimens within the Colville Collection appear to have originated from the earlier 'E. C. Crapper Collection', which was gathered in Fife and Tayside. They were both collected in the early-mid 20th century. Despite this, these two collections are not regarded as independent, as they had been merged prior to coming into the possession of the University of Dundee. Curatorial estimates currently suggest that it would take in the area of 125 hours of work to complete digitisation of the J. L. Colville Collection.

The remaining collection considered a priority for digitisation within the University of Dundee herbarium is the "Arbroath Collection". Although the collection is currently recorded on INCA, this is not to an individual specimen level. INCA currently holds 317 records representing on average 15 specimens per record. Current curatorial estimates suggest that 375 working hours would be necessary to complete digitisation to a specimen level.

Since the initiation of this project, significant progress has been made in the digitisation of the 'R. A. Taylor Collection'. These specimens were collected by a former University of Dundee Botany lecturer during the 1940s and 50s from the areas surrounding Rescobie and Balgavies Lochs, in Angus. All 263 specimens within this collection should have been added to the INCA database by the time that this report is published.

There are several collections of specimens identified as being of Scottish origin housed within the Herbarium stores that have not been officially accessioned into the University Collections and may not be retained. They have been listed below for the sake of completeness:

- Approximately 50 sheets of plant Voucher specimens collected by Brian S. Brookes, Dun Moss, Forest of Alyth, 1970.
- Approximately 50 sheets of plant specimens collected by Hugh Ingram between the 1960s and 70s.
- Approximately 12 sheets of plant specimens collected by Hugh Ingram and Ursula Duncan, Roscobie Loch, Angus, 1970s.
- Approximately 40 plant pathological specimens, collected in 1955.
- Approximately 70 specimens of 'Caithness Mosses', collected between 1983 and 1984.
- Approximately 20 Mosses collected by Hugh Ingram, Tulach Hill, 1983.
- A collection of prepared slides of Diatoms collected by Dr George Dickie, mostly dating to the 1870s. Only some of these are of Scottish origin. Stored in 9 boxes, the collection is estimated to consist of 40 slides per box, or 360 slides in total.
- Approximately 30 Insect cases, most of which are in very poor condition, and of which those specimens that cannot be conserved will be disposed of. These are predominantly of unknown geographic origin, although 1 box is known to be Scottish, while other specimens appear to have previously belonged to the 'Kenneth Todd collection'.

In addition, the University possesses a collection of 1,009 root nodule specimens in spirits that is currently on loan. Collected by Tom Corby c.1968, the collection almost entirely originates from Zimbabwe. None of the specimens are known to be of Scottish origin. Digitisation of collection data would require in the region of 168 hours.

The vast majority of the University of Dundee zoology collections have now been catalogued in INCA, although in the case of insect collections these are not yet to specimen level. These include a drawer of approximately 170 Diptera (flies) from Glen Clova, which should require 12 hours to catalogue digitally, and a currently uncatalogued collection of freshwater and terrestrial mollusca, of which very few are believed to be of Scottish origin.

Assuming that all collections listed above were to be accessioned into the overall University of Dundee collections, they should constitute to an extra 14 datasets. Assuming that all relevant data had been entered into INCA previously, construction of the corresponding metadata should take roughly 3 ½ hours,

excluding the time required to extract the data from INCA.

## Analysis

One of the most valuable resources to this study has proved to be the personal knowledge that the curatorial staff have of their collections. Specifically written, and to a greater extent, digitised records do not necessarily provide contextual understanding of the specimens in relation to the collection, and they do not always make clear the relationships between objects within museum collections records. This lack of information can be attributed to the predominant usage of collections databases as an organisational tool, as opposed to a generalised depository of data regarding each specimen. On the other hand, it could be argued that this assumption is now changing, particularly considering the more recent moves towards providing online resources that are easily understood by the non-museum professional. Such a process requires incorporating exhibition-style interpretation with more standard catalogue data.

One of the other factors that became clear throughout the life of the project is that there was a considerable dependency on the accuracy, thoroughness, and formatting consistency by both the specimen collector, and later recorders of the data. Unfortunately, to some degree, it is difficult always to identify errors in recorded data, while in other cases, it is merely a case that factors such as the grid reference of the collection site was not recorded. In a similar vein it was found that even minor formatting changes in the data could significantly increase the time needed for analysis of the data.

Another factor which needs to be made clear is that as the remit of this project deals solely with Scottish specimens, the assembled datasets do not reflect the contents of each collection in their entirety. This is particularly the case regarding the University of Dundee Zoology collection datasets, of which only 142 out of a total 2,866 records represent specimens of Scottish origin; less than 5% of the full collection.

In terms of the project itself, although it was not possible to analyse every collection within both institutions, a significant proportion of the data was able to be collected and analysed within the projected timescale. An important factor is that the speed at which this could be done was heavily influenced by the degree to which digitisation of specimen records had been carried out by the participating institutions. Although the datasets analysed throughout the duration of this project do not cover every specimen which may potentially fall within the project remit, they do provide data on all of the specimens that have been identified as being of Scottish origin on the participating institutions databases.

One other major challenge originated from the difficulty in ascertaining whether a group of specimens were, in actuality, a discrete collection. This arose from the fact that museum records do not always deal in smaller 'collections', but rather over-arching fields such as 'zoology' or 'botany'. In such cases, it is often only specimens that were accessioned together as a discrete, prior named, collection that have such information recorded. It was under such circumstances that curatorial advice was most needed, either to identify specimens that constituted

to a collection but were not recorded as such, or to determine criteria for when an individual's contributions to the museum collection could also be regarded as a discrete collection.

Regarding the timescale of the project, in retrospect the original estimates might be considered slightly optimistic, although throughout the course of the project it became evident that a more accurate estimate would have required a far higher level of familiarity with the collections than would have been feasible.

Regardless, this pilot study should still be considered a useful tool for estimating the timescale of future projects. It has also identified factors which most heavily impact the time required, in particular the degree to which collections records have been digitised.

There are several key lessons to be taken away from this study that would be of value to future projects of this nature. Firstly, that writing up the finalised metadata forms should be started as early as possible. The data extracted from either database system is usually extracted in a rather raw and initially inaccessible format, yet attempts to restructure the tables to a more readable format proved to be considerably more time-consuming than initially anticipated, and in the end merely detracted from the time available to complete the final datasets. The second major lesson learned is that very little of the information which identifies a single specimen as part of a collection is actually contained within collections databases. Therefore close collaboration and cross-checking of metadata with curatorial staff can speed progress up significantly, as they often know more about the collections and specimens than is stored in collections records.

Limitations identified in the course of this study are mostly technological ones, and are specific to every institution and its choice of database software. While extracting data from AdLib proved to be a reasonably simple process, extracting data from INCA proved to be less intuitive, and required curatorial assistance. The ease of extracting such data in a format compatible with Microsoft Excel (or a similar spreadsheet system) has a significant impact on overall project efficiency, and will be an important factor for consideration in future projects.

Aside from technology-based complications, the other major limiting factor proved to be the ambiguity of how collections within a museum are determined. It seems highly probable that the exact definition varies from museum to museum, based on how they have chosen to organise their collection.

## **Appendix I: Dataset Key**

## **Appendix II: Dundee Art Galleries & Museums Datasets**

### **Dundee Art Galleries & Museums Botany Collection**

## **Appendix II: Dundee Art Galleries & Museums Datasets**

### **Dundee Art Galleries & Museums Invertebrates**

## **Appendix II: Dundee Art Galleries & Museums Datasets**

### **Dundee Art Galleries & Museums Vertebrates**

## **Appendix II: Dundee Art Galleries & Museums Datasets**

### **Dundee Art Galleries & Museums Voucher Specimens**

## **Appendix III: University of Dundee Datasets**

### **University of Dundee Herbarium**

## **Appendix III: University of Dundee Datasets**

### **University of Dundee Zoology Collection**