Progress Report
Technical and Financial

Grants for Digital Data
Specimens Thematic Network

Prepared By:

Elizabeth Morrison

Zoologist

Natural History Museum of Jamaica

Institute of Jamaica

24/06/2010

1. Executive Summary

The Inter-American Biodiversity Information Network (hereinafter "IABIN") and the Natural History museum of Jamaica (NHMJ) of the Institute of Jamaica have entered into agreement to establish through the SPECIES AND SPECIMENS DATABSE DEVELOPMENT PROJECT the regulatory framework to digitize specimen data of 10,000 specimens of insects; 2,500 other invertebrate specimens; 1,000 vertebrate specimens currently housed in the zoological collections of the NHMJ. The specimen data includes the required DarwinCore fields for each specimen: Global unique identifier, locality, date collected, basis of record, institution code, accession number, collector, collection code, catalog number and scientific name.

The data for 10, 157 specimens is currently stored in MS-excel spreadsheets and is currently being verified as to its accuracy and completeness. The main problems encountered included missing data for necessary fields such as GPS coordinates and some misplaced data include collector data that required physical checks of the housed specimens to ensure veracity of the data. The main issue of missing GPS coordinates is being addressed as coordinates are being assigned to localities with the use of maps and GIS software. To date a total of 346 common collection localities across Jamaica have been assigned GPS coordinates. The intent is to upload the verified data once complete into ARA system with the assistance of INBIO technicians who designed the ARA software within the next month. The biodiversity content thus digitized will be made freely available to the public.

2. Results of Planned Outputs and Scope of the Project

The primary objective of the project was to develop and provide access to metadata catalogs and datasets of information on faunal specimens from Jamaica, in a format that is interoperable with data made available through other institutions. In order to achieve this objective it was be necessary to convert current locality data (available as descriptions) into GIS coordinates that would allow compatibility with IABIN databases. This information will compliment the Wildlife and Protected Areas Act as well as the National Biodiversity Strategy and Action Plan for the island and will allow for ease of access by researchers worldwide and at all levels to information on species data.

Up until June (month 6) of the project 10,000 insect specimens, 1250 other invertebrate specimens and 500 vertebrate specimens should have been digitized. (See table 1)

Table 1: **Quantifiable Performance Indicators**

Activities	Month 3	Month 6	Month 9	Month 12
Digitize 20,000 specimens of insects	5,000 specimens digitized (25%)	10,000 specimens digitized (50%)	15,000 specimens digitized (75%)	20,000 specimens digitized (100%)
Digitize 2,500 other invertebrate specimens	625 specimens digitized (25%)	1250 specimens digitized (50%)	1875 specimens digitized (75%)	2,500 specimens digitized (100%)
Digitize 1,000 vertebrate specimens	250 specimens digitized (25%)	500 specimens digitized (50%)	750 specimens digitized (75%)	1,000 specimens digitized (100%)

2.1 Results

To date a total of 10810 have been compiled into an excel dataset ready for upload into the database. (See Attached Datasets.) The verification process is proceeding concurrently as the MS-excel datasets are being created. Some fields are missing due to the fact that the collections are old and data for such fields were not collected. E.g. the majority of localities are represented in a descriptive form as the use as the assigning of coordinates to each specimen was not the practice before the 1990s. GPS coordinates are now being assigned using maps and GIS technology. To date a total of 346 common collection localities across Jamaica have been assigned GPS coordinates (See attached Annex)

Table 2: Specimens Entered To Date

Category	Total Number Entered	Number of specimens with complete information	Number of specimens with Missing Information
Insects	8685	383	8302
Invertebrates	1463	262	1201
Vertebrates	662	8	654

*There is no catalogue number for the collection. There is also no collector code for the specimens used above. Only a small number of specimens from a marine collection have a collector code and it was not used in the dataset.

3. Methodology employed and activities carried out to achieve the planned outputs

3.1 Purchase of computer system

A project computer was identified and purchased according to government of Jamaica (GOJ) purchasing guidelines where quotations of three similar systems are submitted for review.

3.2 Hiring of Project Assistant

The project assistant, Miss Judeen Meikle was hired and began work on May 17th. The process was conducted according to hiring and selection guidelines of the Government Of Jamaica through the Human Resource Management Department of the Institute of Jamaica. As such, a job description was developed for the position (see Annex 1), the job was publicly advertised for a period of two weeks, and applicants were shortlisted according to their qualifications and final selection of the project assistant made by the IOJ selection committee. Unfortunately hiring was delayed for a week as under new government guidelines permission to hire a new employee had to be acquired from the Ministry of Finance. The Project Assistant works three times a week from 8:30am – 3:30 pm.

3.3 Training

Training of the newly recruited project assistant included an overview of the NHMJ Zoology Biolink database (database currently used by the NHMJ zoology department) and the ARA system. Training was conducted by Mr. Damian Jackson.

3.4 Identify complete data sets

An initial assessment of the sets to be provided has been completed and the data is being added to the excel database that will be supplied to the Ara software managers for upload. In an effort to give data sets that are as complete and accurate as possible, verification of the data provided is being conducted on an ongoing basis by the project assistant and the zoologist.

The specimen data was exported from Biolink into an MS-Excel file as Biolink does not export the data in TAPIR and DIGIR format. This MS-Excel file will be uploaded to ARA and posted on the worldwide web by IABIN. The software which will be used to provide data to IABIN is the ARA system which is a free data capture tool developed by the Species and Specimens Thematic Network (SSTN). After discussing several options the ARA data tool was selected for the following reasons:

Mr. Damian Jackson, Systems Administrator, IOJ had been previously trained in the use of the ARA system at a IABIN funded SSTN Software Tools Workshop held in Costa Rica in August 2008. As such it was the tool that we were familiar with.

The ARA system is user friendly and has in built in functionality to export the data in DIGIR and TAPIR standards.

It was the most cost effective.

3.5 Verification Process - Addressing Gaps in locality data

GPS coordinates acquired from localities mapped in Arc GIS and added to each specimen in the Excel file.

3.5.1 Missing Data

The MS-Excel dataset were reviewed for completeness and missing fields identified. The main missing fields included GPS coordinates data and collector data. Steps were taken to address this issue.

3.5.1.1 To geo-reference each specimen:

The location in the database was located on a road map (soft copy) for the general area.

Then it was then located on the ICONOS image in ArcGIS that was correctly projected.

The coordinates were then extracted

Each coordinate was added to its corresponding specimen

3.5.1.2 To assign collectors:

The locality, year and accession number was recorded.

The actual specimen was pulled from the collection and each specimen was search to locate the specimens that corresponded with the data in (1).

The collector was then assigned to each individual specimen.

3.6 Data Accessibility via the Internet

An attempt to import the data from the Excel file into the ARA system which has the compatible TAPIR and DIGIR protocol was not possible as their import feature has not yet been developed. Contact was made with the developers of the ARA system David Solano de Unidad de Desarrollos Informáticos Instituto Nacional de Biodiversidad (INBio), Santo Domingo de Heredia, Costa Rica and they agreed to import the MS- Excel data we have accumulated so far into the ARA system as they can manually import the data for us until the import feature is completed. This will allow for the project to be completed on time. The technician at INBIO who will be doing the data importation will not be available until July to upload the data. All data accumulated in MS-Excel until that time will be sent to them via email for upload.

Project Activities that will be completed each month as included in accepted proposal document are in Table 3 below. Status of activities are included in Table 4 below.

Table 3: Project Activity Timeline

Activity vs. Month	1	2	3	4	5	6	7	8	9	10	11	12
Identify complete data sets	x	x										
Hire part- time project assistant	x											
Training		x										x
Convert descriptive data to GPS coordinates		x	x	x	x	x	x	x	x	x	x	
Add coordinates to Biolink		x	x	x	x	x	x	x	x	x	x	
Formatting of Biolink data to TAPIR and DiGIR for IABIN and upload			x	x	x	x	x	x	x	x	x	
Production for Pub. Ed. Materials on database								x	x			
Plan and host end of project workshop										x	x	x
Prepare and Submit quarterly technical and financial reports			x			x			x			x

Table 4: Status of Activity Table

Activity vs. Month	Status
Identify complete data sets	Ongoing
Hire part- time project assistant	Complete
Training	Training of Assistant is complete. Training associated with the use of the database for data users is to be completed in the next phase of the project.
Convert descriptive data to GPS coordinates	Ongoing
Add coordinates to Biolink	Ongoing
Formatting of Biolink data to TAPIR and DiGIR for IABIN and upload	To be completed next month
Production for Pub. Ed. Materials on database	To be implemented
Plan and host end of project workshop	To be implemented
Prepare and Submit quarterly technical and financial reports	Ongoing

4. Lessons learned, problems and viable solutions

The project was also delayed due to time consuming government policy and guidelines regarding hiring of employees and the purchasing of equipment. For future similar projects adequate and more realistic timelines must be proposed at the initial stages of the project.

The project was delayed for several days due to social and political unrest in the capital city of Kingston in May.

Many of specimens stored in the zoological collections remain unidentified to species. This therefore reduces the number of specimens that can be placed in the database. There is a need for identification

by experts in the field with possible international assistance to address the taxonomic impediment that exists.

It is important to collect all the necessary metadata fields associated with a specimen and to ensure that localities are specific and accurate as possible.

The use of compatible interoperable data capture software would reduce time and error encountered during the transfer of data.

A faster computer to operate ArcGIS on e.g. an IMac is needed. The current computer's screen and speed does not allow for maximum efficiency when reviewing images in ARCGIS. There is also a great need for a more detailed soft copy of maps of Jamaica, these maps should include places had name changes.

Additional personnel are required to go through the catalogues to check the specimens for missing data.

5. Counterpart Funds (herewith the report of expenditure counterpart Excel table)

Please see Report submitted online at http://cofinance.iabin.net/

6. Financial Report

Please see attached Expense report in MS- Excel

Please note that expenses are lower than budgeted for the time period as:

- The project assistant was hired in the latter part of the period and as such has used less funds than budgeted and
- The computer system purchased was at a lower cost than budgeted

However work had begun prior to the hiring of the project assistant by the zoologist and the IT Manager mostly on the identification of the datasets to be provided and the testing of the interoperability of Biolink and ARA system. These costs were covered by co-financing provided by the Institute of Jamaica.

7. **Annexes**

Please see attached:

Annex 1a and Annex 1b: Job Description and Resume of Project Assistant.

Annex 2: MS- Excel datasheets.

Annex 3: Common GPS localities.

Annex 4: Copy of Invoice – purchase of computer