Journal of Natural Science Collections

Title: Using web-based tools to transform the Bivalvia collection database in the KwaZulu-Natal Museum

Author(s): Ziganira, M.

Source: Ziganira, M. (2018). Using web-based tools to transform the Bivalvia collection database in the KwaZulu-Natal Museum. *Journal of Natural Science Collections*, *Volume 5*, 66 - 77.

URL: http://www.natsca.org/article/2447

Using web-based tools to transform the Bivalvia collection database in the KwaZulu-Natal Museum

Matabaro Ziganira

KwaZulu-Natal Museum, 237 Jabu Ndlovu Street, Pietermaritzburg, South Africa

mziganira@nmsa.org.za

Received: 02/03/2017

Accepted: 07/12/2017

Citation: Ziganira, M., 2018. Using web-based tools to transform the Bivalvia collection database in the KwaZulu-Natal Museum. *Journal of Natural Science Collections,* 5, pp.66-77.

Abstract

The initial steps towards digitising the KwaZulu-Natal Museum's Mollusca collection were taken in 1994. This involved the creation of a Microsoft Access database with a relatively small number of fields designed to capture the essential details of specimen provenance. South Africa has funded national institutions to create metadata which will lead to digitisation (including databasing, digital imaging, and georeferencing), by promoting and increasing access to natural history collection data to a much broader user base. However, at KwaZulu-Natal Museum, the initial progress was very slow, due to problems with database design and lack of expertise. In 2014, a pilot project was initiated to use GEOLocate Web Application to georeference collection records of the Bivalvia database which already have locality descriptions but lack geographic information. Subsequently, the digitised Bivalvia data have been supplied to help big science projects in South Africa. It is anticipated that the records will ultimately be linked to other databases, and used to update coordinates to these other datasets.

Keywords: biodiversity data, digitisation, georeferencing, natural science collections

Introduction

The two research departments of the KwaZulu-Natal Museum (Human Sciences and Natural Sciences) have received funds from the National Research Fund (NRF) to digitise all of their collections. The Natural Science Digitisation Project (NSDP) was developed as part of this national initiative with the aim to digitise, standardise, and clean all of its collection databases. This initiative follows the model of The Global Plants Initiative (GPI) project, which has increased plant collector research and the compilation of such data (Penn et al, 2018). This initiative aims to help national institutions to create metadata which will lead to digitisation (including databasing, digital imaging, and georeferencing), by promoting and increasing access to natural history collection data to a much broader user base (Berent et al, 2010). The objectives are thus to digitise and mobilise biodiversity data stored in museums, herbaria, and research institutions in South Africa towards creating one research infrastructure. To achieve this, it was noted that the digitised data include species name, georeferenced location, collector and collection date, and other specimen data recorded on the label by the collector (Paterson *et al.,* 2016).

The collections management system SPECIFY was adopted by this national initiative as standard for all animal specimens, to ensure sustainable preservation of the collections and that data meet the Darwin Core Standards. However, because of limited 'in-house'

human capacity, NSDP has targeted the training of interns and volunteers to perform some tasks, although the inevitable turnover presents set-backs. Because of the size of some of the collections, migration to SPECIFY might take a long time and thus compromise other 'in-house' research priorities. In addition to this, the cost of extracting data from open-sources has not yet been evaluated by the KwaZulu-Natal Museum curators who have, to date, used the data in its current format in Microsoft Access, and are happy with that format. Because of this, it has been difficult to measure the success of capacity-building training, due to lack of application of these tools. Certain tools, however, such as georeferencing, have been adding value to the collections.

The first practical phase of the digitisation project was in the year 2014-2015. Georeferencing tools were used to add geographic coordinates into the Bivalvia database, and promising results were achieved through a short staff training programme. Digitising the Bivalvia database was a pilot project because this database was incomplete; many of its records were not yet databased and of those that were, many records lacked geographic coordinates. The aim of this work was to use the GEOLocate Web Application (Rios and Bart, n.d.) to georeference collection records of the Bivalvia database of the KwaZulu-Natal Museum that already had locality descriptions but lacked geographic information. It is believed that, once complete, this exercise will provide guidelines for cleaning and improving the quality of data for end-users, thus saving time and money in repeating similar tasks with other databases at KwaZulu-Natal Museum and other South African institutions in general.

Mollusca Collection

The mollusc collection consists predominantly of dry shells, but, where possible, wet samples of each species are preserved in ethanol for anatomical examination. The Mollusca collection has benefited greatly from the shell collection and library of Henry Cliften Burnup (1852 - 1928), who was Honorary Curator of Mollusca. After his death in 1928, his collection was incorporated into the Mollusca collection and significant expansion occurred through field work, donation, and exchange, as well as purchase (Kilburn and Herbert, 1994). Fieldwork is usually conducted on an annual basis in order to build up the collection, as well as to improve the taxonomic and ecological data associated with specimens. One of the biggest programmes was the Natal Museum Dredging Programme (NMDP), which began in 1981 and continued until 1997, on annual 10 days cruises (Kilburn and Herbert, 1994). 1,100 stations were sampled, ranging between off KwaZulu-Natal to south-western Cape, as well as on the Agulhas Bank (Kilburn and Herbert, 1994). This programme enriched the KwaZulu-Natal Museum mollusc collection with the most extensive and accurately documented samples. These samples include many rare and unusual species such as *Nassarius eusulcatus* (G.B. Sowerby III, 1902), *Anadara africana* (G.B. Sowerby III, 1904) (now a synonym of *Anadara pygmaea* (H. Adams, 1872)), *Anatoma yaroni* Herbert, 1986, and *Puncturella voraginosa* Herbert & Kilburn, 1986, to mention very few (Kilburn and Herbert, 1994). The Mollusca collection ranks among the 15 largest in the world, and is certainly the largest in both Africa and the Indian Ocean rim. Currently, this collection houses more than 160,000 specimens, many of which have been fully databased in MS Access.

Mollusca databasing

The initial steps toward digitising the museum's Mollusca collection were taken in 1994. This involved the creation of an MS Access database with a relatively small number of fields, designed to capture the essential details of specimen provenance. Initial progress was very slow, due to problems with database design and lack of staff expertise. In 1996, Ntombi Mkhize was employed on a part-time basis and she began to input data for the non-marine component of the collection. In 1999-2000, additional funding was accessed through SA-ISIS/BioMAP (South African Integrated Spatial Information System / Biodiversity Mapping and Assessment Programme), initiated by the Department of Arts and Culture, and Science & Technology. This allowed the employment of a dedicated databasing technician for circa two years, before the funding ceased. Subsequently, at its own expense, the museum employed Ntombi Mkhize again on a full-time basis to continue the Mollusca databasing work. After her resignation in 2014, there was a brief hiatus until Matabaro Ziganira was appointed and, finally, the databasing backlog was eliminated in 2016.

The databasing of the collection was initiated primarily as a research tool, facilitating rapid access to distribution and inventory data, and to make spatial data available to potential stakeholders who might require such information (e.g. KZN Wildlife). For this reason, data entry was initially restricted to records

from southern Africa and the south-western Indian Ocean. Only when this was completed, was databasing expanded to include our holdings from other parts of the world, by which time, the specter of GRAP 103 compliance was also looming large. GRAP 103 is an accounting standard that prescribes the uniform accounting for classifying and recording Heritage Assets, and regulates related disclosure requirements. The standard requires that institutions have records of their collections that are fit-for-purpose, and which contain basic information about objects, including: identification, ownership, location, condition, and value. Public Entities reporting to the Department of Arts and Culture must comply with the requirements set out in the standard. On its own, GRAP 103 has no scientific value. Only when the goals of potential stakeholders such as the South African Biodiversity Institute (SANBI) are brought in, does the exercise become one of scientific value.

The Bivalvia database

The Bivalvia database was created in early 2000 using MS Access, a commonly-known and widely utilised programme for museum collection management. This database contains 25,000 records, many of which are old specimens, collected many years ago. All the information stored in this database is organised in a spreadsheet containing only available and pertinent data for the collected specimens (eg. taxonomic determination, locality description, collection date, etc.) (see Table 1 in Appendix I). The locality information primarily describes the place where specimen data were recorded at the time of collection. However, some of these records lack geographic locations, or the locality description might be ambiguous or inaccurate, or simply not correspond to current geographic location due to anthropological changes (Chapman, 2005; Chapman and Wieczorek, 2006). This limitation makes it difficult to validate the coordinates, and errors are usually difficult to detect. In addition, the extent to which validation can occur depends on how well the locality information describes the same place (Chapman and Wieczorek, 2006). Thus, the process of georeferencing the Bivalvia database also aimed at cleaning the data, and normalising/harmonising ambiguous records to unambiguous master records, through selection and import of unique records only into the GEOLocate Web Application (http://www.museum.tulane.edu). However, there were instances where records did not provide coordinates because of ambiguous or erroneous locality descriptions. In such cases, the 'County' column in the spreadsheet was labelled 'not georef' to indicate that no coordinates were available (see Table 2 in Appendix I). Another conflict occurred when coordinates were misplaced to a different location, or simply presented a very high degree of uncertainty on the map. To resolve this, the knowledge of the Chief Curator, Professor Herbert, was essential. Usually, the Chief Curator knew either the collector's collection events, or was aware of the geo-political changes in the country of collection. Also, the Chief Curator understood the interpretation of the symbols used on the specimen records, and was able to clarify the queries. Paterson et al (2016) state that in resolving erroneous and misleading label information, such as collecting localities and dates, the knowledge of the curator is crucial; the curator might know about the collector in question, or might have collected other specimens from the same locality, or at the same time. Good records information, such as locality descriptions, can lead to more accurate georeferences with smaller uncertainty values, and thus provide users with much more accurate and higher-quality data (Chapman and Wieczorek, 2006).

Data export to Microsoft Excel

The Bivalvia database was exported into a Microsoft Excel spreadsheet, retaining complete formatting and layout (Figure 1). In the MS Excel datasheet, columns (ID and ID1) were added on each side of the spreadsheet, containing the same sequential numbers in exact order. Adding these numbers minimises the chance of errors caused by mixing up records while filtering and sorting many rows in the Excel datasheet. It is highly recommended that the entire process of georeferencing follows guidelines that are designed to reduce errors and repeatability (Paterson et al, 2016). A copy of the sorted datasheet was made, in which subsequent queries were made. In the copied sheet, the 'locality' column was filtered by selecting 'unique record only', and a new copy of the datasheet was made. After filtering, 5,000 records were found to have unique localities, and these were used for the georeferencing exercise. The remaining 20,000 records were considered 'excluded records', because they had duplicate locality descriptions which were already represented in the 5,000 unique records. This is very important because in some instances, many specimens are collected in the same locality with similar descriptions. In those cases, it is imperative that 'unique locality only' are georeferenced in a batch mode. In this way, one needs only deal with a single record out of many with similar locality descriptions in the database, therefore saving

invaluable time. After the georeferencing process was completed, the georeferenced spreadsheet was exported back into the original database in MS Access format, and, through a series of queries, the georeferencing information from the 5,000 'unique records' was added to the corresponding 20,000 'excluded records' in the database.

Georeferencing of the Bivalvia database

Georeferencing of the Bivalvia database was primarily done through locality descriptions. The 5,000 unique records were sorted electronically and formatted in a CSV file before upload to the GEOLocate Web Application (http://www.museum.tulane.edu) (see Table 3 in Appendix I). GEOLocate is a platform for georeferencing natural history collection data, and is currently being developed as a web service through integration and development of BioGeomancer (BioGeomancer Working Group, 2005) (Figure 2). Tools such as BioGeomancer work better when georeferencing is done in batch mode. The locality description is submitted and the georeference reports back by providing further information on uncertainty, where several options exist from the locality information (Chapman and Wieczorek, 2006). After data were georeferenced and while the database was still online, I evaluated each record individually by marking the non-georeferenced records for further review, and also assessed and validated each record for uncertainty error (Figure 3).

In most instances where geographic information was given, uncertainty data were usually attached for each record georeferenced.

Locality descriptions of many records of the Bivalvia database are based on named places that might have changed in size over time. In some instance, this phenomenon renders the current extent of a named place greater than its historical range (Chapman and Wieczorek, 2006). For this reason, GEOLocate uses an uncertainty polygon by clipping a circle where it overlaps the ocean for terrestrial data, and thereby providing a much more accurate representation of the locality (Chapman and Wieczorek, 2006). This allowed me to either agree or modify the extents that might not reflect the uncertainty predictions from the several options that GEOLocate suggested. In order to accurately georeference the Bivalvia database, the knowledge of the Chief Curator and Google Earth were constantly referred to for verification of the current locality information during data sorting and validation (Figure 3).

Importing and merging of georeferenced data into the main database

Once the process of evaluation and assessment was completed, the next step was to import the georeferenced database back into the main database. This was executed by importing and converting the georeferenced CSV file into MS Access format and

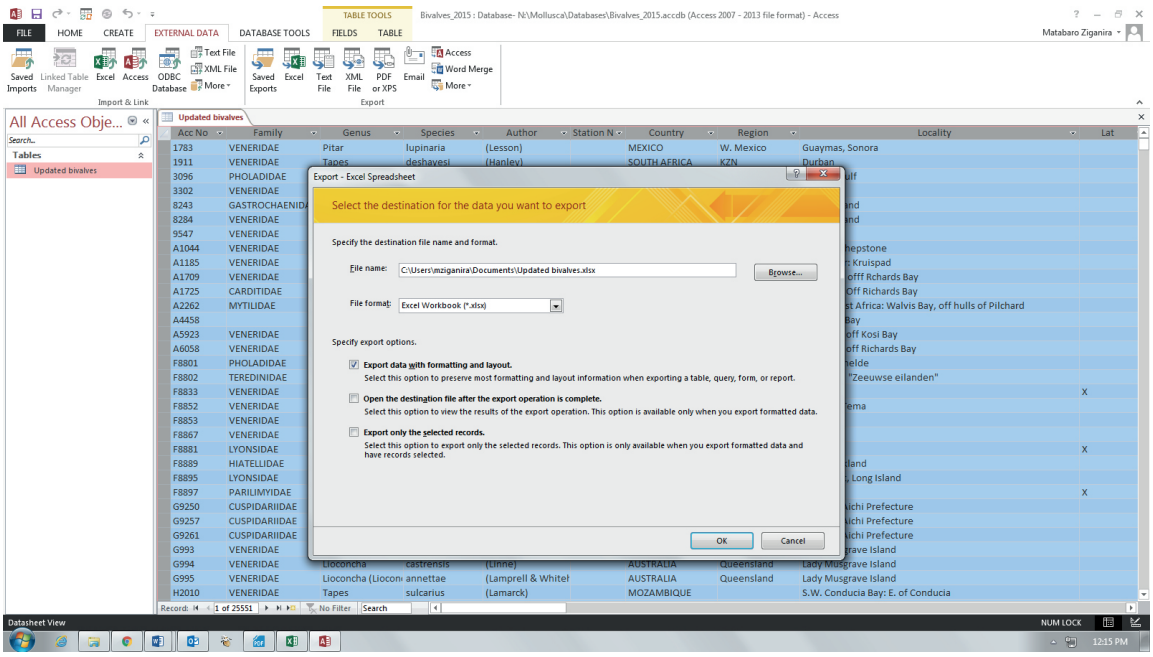

Figure 1. The import process of the Bivalvia database into Microsoft Excel spreadsheet from Microsoft Access.

merging the two spreadsheets as one database. It was expected that the 5,000 unique georeferenced records would influence the 20,000 non-georeferenced records in the main database by adding geographic information to records with similar locality descriptions. However, if the merging is not properly executed, errors and confusion might negatively affect the main database. In order to prevent errors during this process, two new columns were inserted into the georeferenced database, namely 'ID1' and 'locality backup'. The 'ID1' column had ascending numeric values of one to 5,000 and was inserted as column number one of the spreadsheet. The 'locality backup' was a duplicate of the locality column that was used during georeferencing, and was placed next to the 'ID' column as the last column of the spreadsheet. This strategy is imperative because it exposes errors where numeric values do not correspond to the associated locality description after the merging of the two spreadsheets. A copy of this database was made for reference. In the original database, fields entitled 'georeference comments', 'correction status', 'precision', 'error polygon', 'multi results', 'radius uncertainty', 'radius uncertainty1', 'radius uncertainty2', and 'locality fixed' were inserted in this table. Through creating and executing queries in MS Access, the information in the georeferenced database was combined with the original database. Columns labelled 'latDD' and 'longDD' in the two databases were interconnected based on the similarity of their locality descriptions. This allowed the georeferenced record to directly add geographic information to records in the main database with similar locality descriptions. This means that small numbers of unique records are able to influence the entire dataset, thus saving valuable time and money.

The geographic information derived from the process of georeferencing is usually in the format of degrees decimal. Some of the 20,000 'excluded records' in the main database had already been allocated geographic information in the format of degrees, minutes, and seconds. Because of the format differences, it was important that the 'excluded records' be converted into degrees decimal format so that consistency was maintained in the database. To do this, a new MS Excel spreadsheet was created from the 'excluded

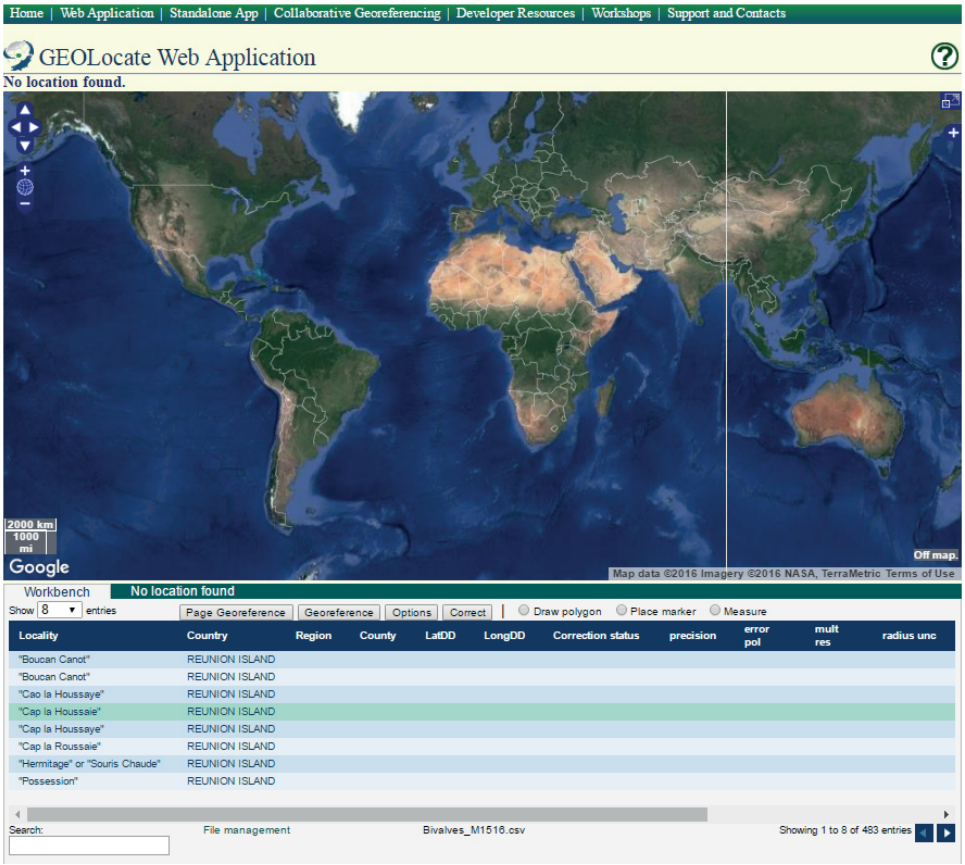

Figure 2. GEOLocate Web Application showing data being uploaded before the georeference process begins. The Georeference options allow changes before the application runs.

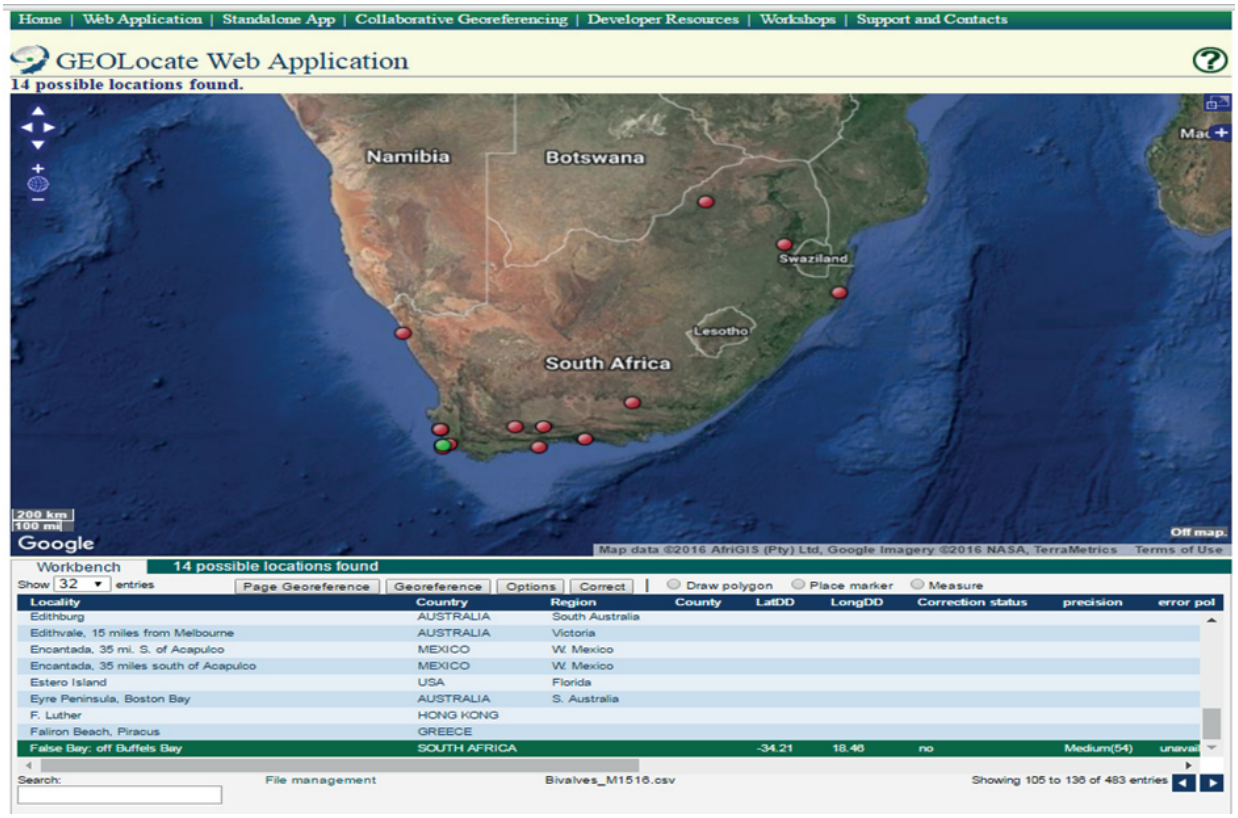

Figure 3. Example of georeference outcome for False Bay: off Buffels Bay in the Western Cape, South Africa, showing 14 possible locations found. 3a. The web application suggests that the green dot on the map is the correct location.

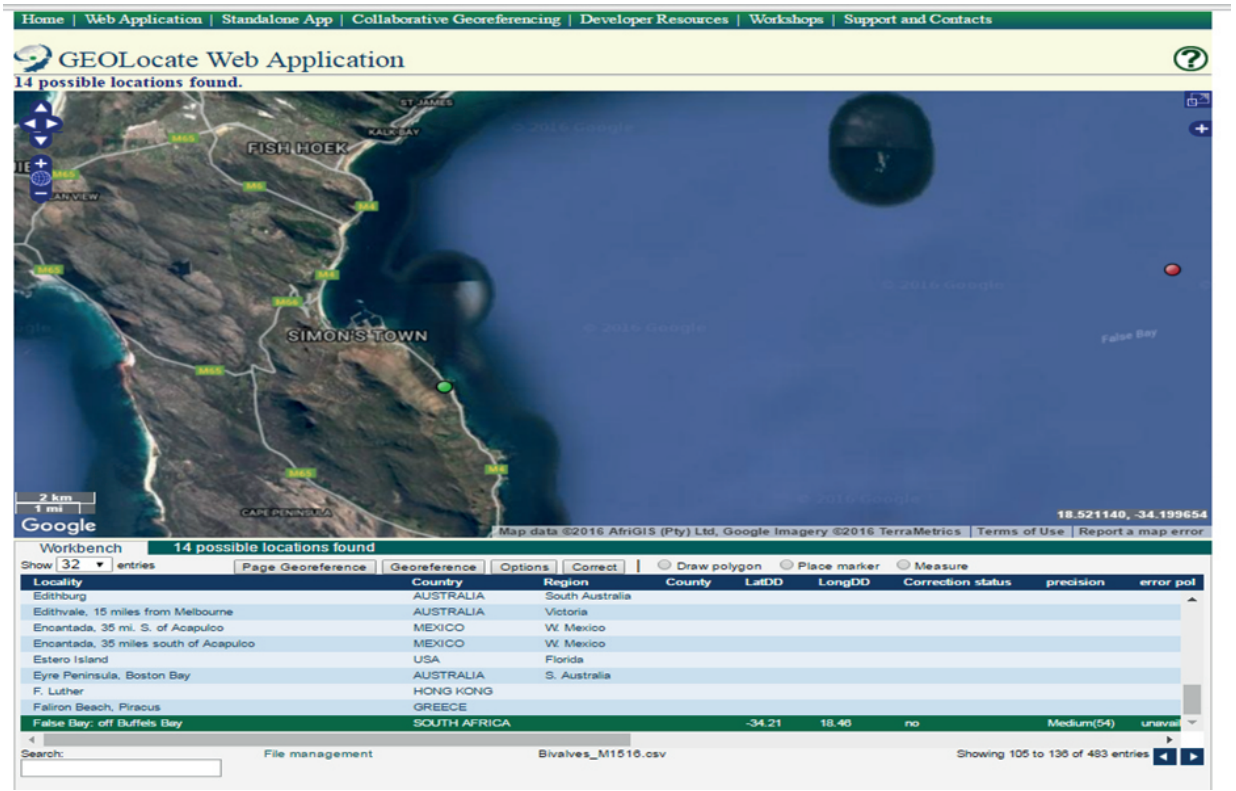

Figure 3b. By magnifying the map, it is clear that the green dot is off False Bay but located inland, in Murdock Valley.

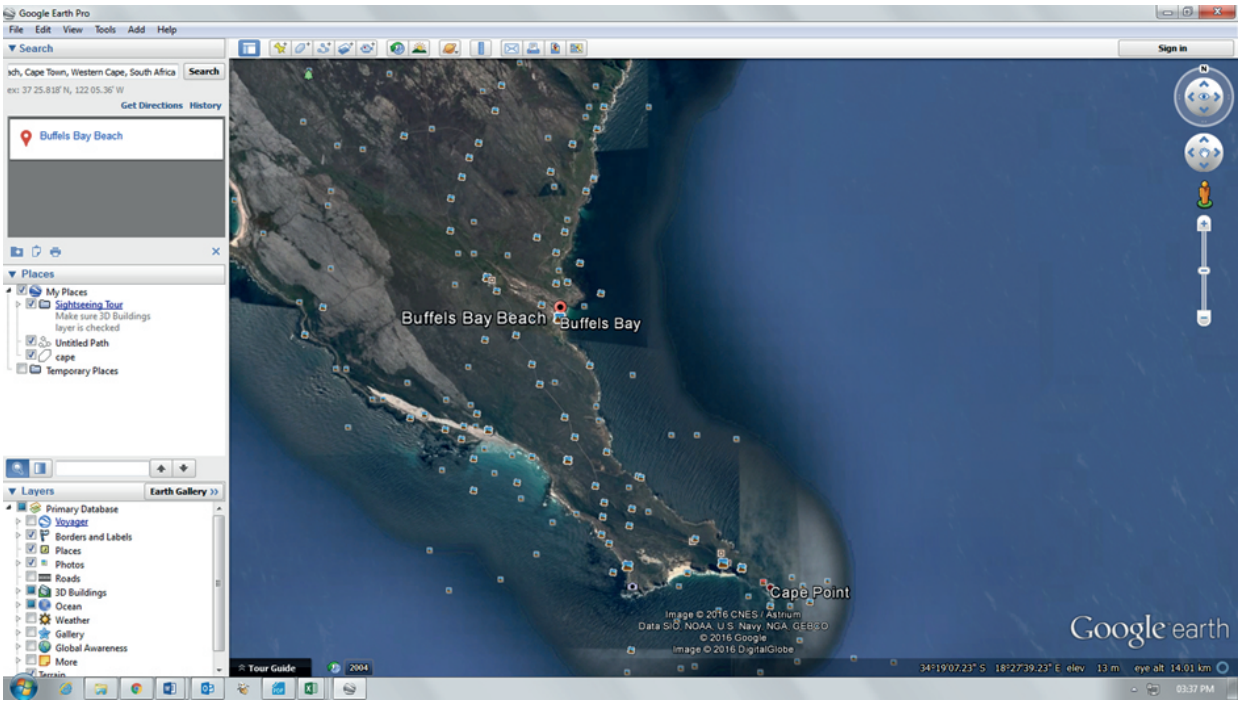

Figure 3c. Google Earth search for 'Buffels Bay' suggests that the correct location is not that suggested by GEOLocate in (a); it is the 'red dot' below the 'green dot'.

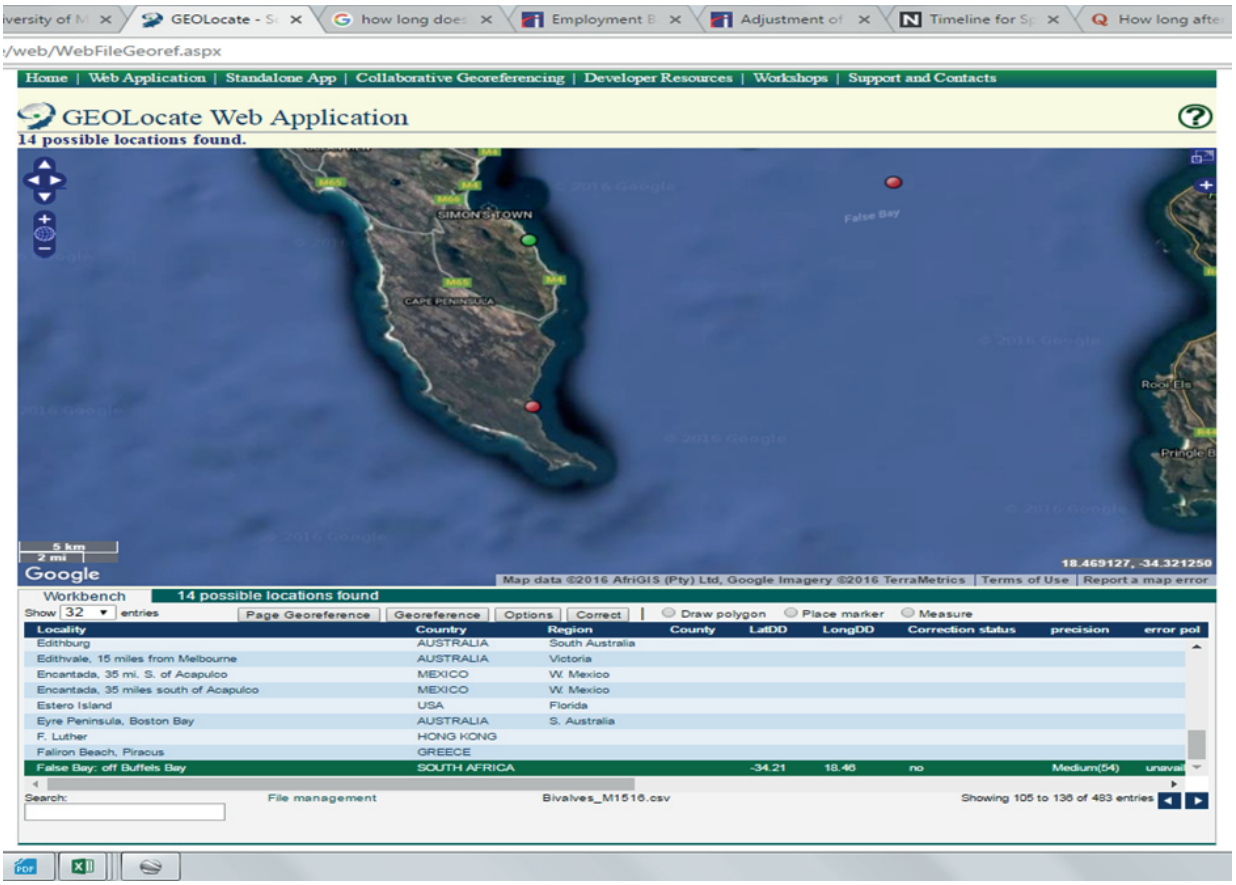

Figure 3d. The arrow pointing to the 'red dot' is the correct locality, and requires adjustment.

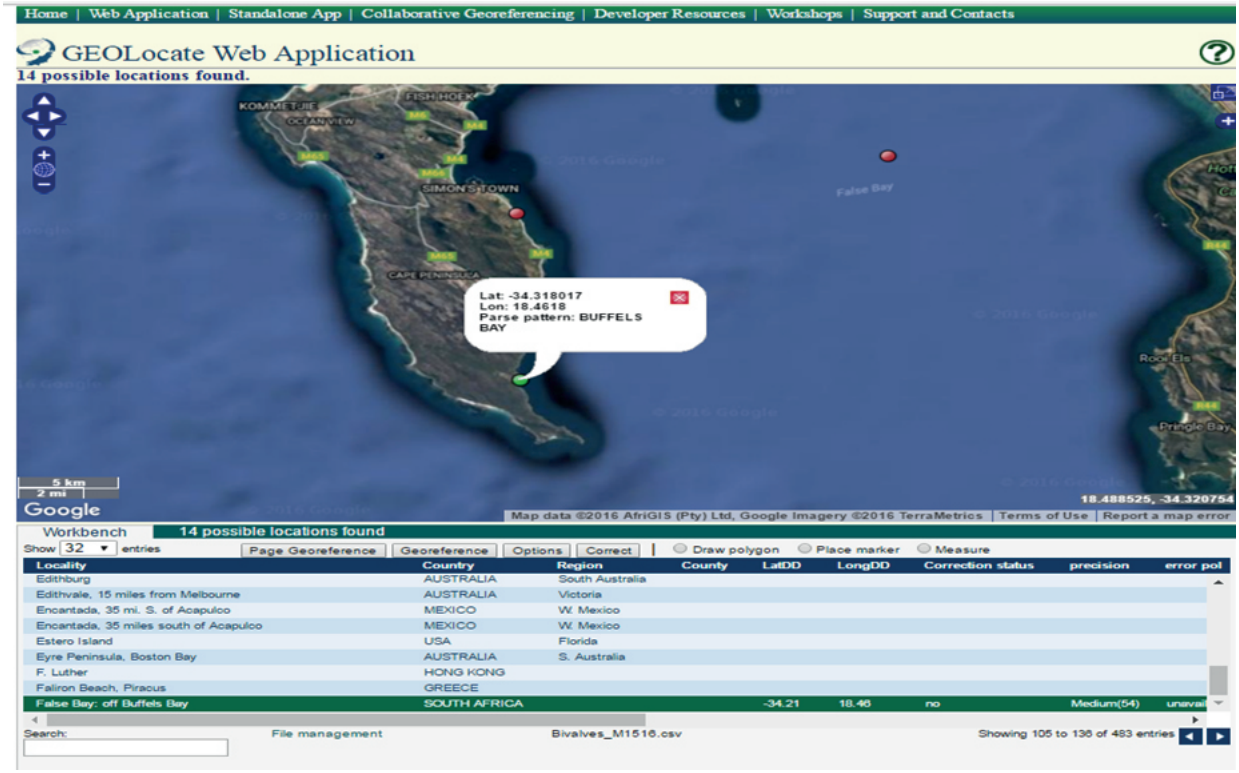

Figure 3e. The new 'green dot' shows the new corrected locality location.

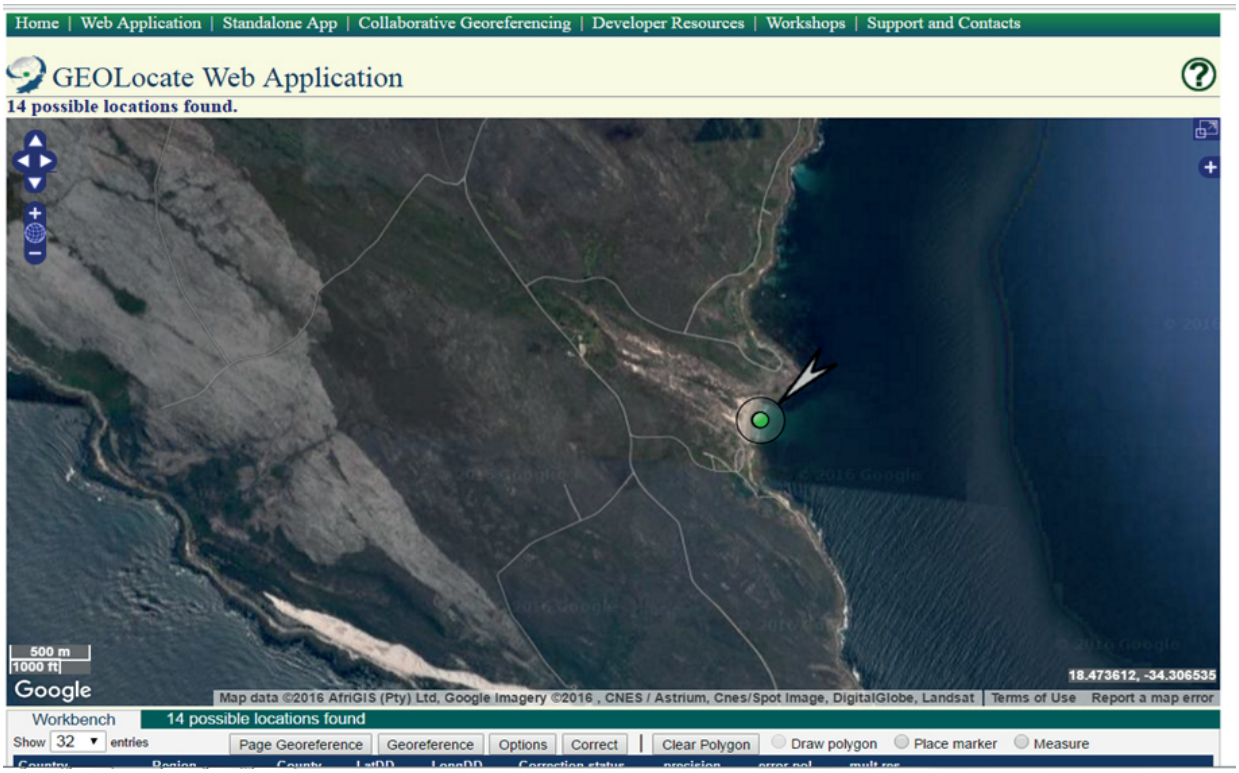

Figure 3f. The locality is marked with a circle around it, showing uncertainty in the locality description.

records', filtered to select records with latitude and longitude information only. All fields were deleted except 'accession number', 'latitude', 'longitude', 'lat.DD', and 'long.DD'. Before the column 'accession number', a new column labelled 'ID1' was again used as reference in case errors occurred during importing and merging of the two databases. In order to convert degrees, minutes, and seconds into decimal, five new columns were inserted next to the columns for latitude and longitude. Two more new columns named 'LatDD' and 'LongDD' were inserted and designated for the formula. After conversion to degrees decimal, the column of latitude South (S) was sorted first, followed by the column of longitude West (W) so that coordinates of South-West were marked negative. After this, the Excel file was exported into MS Access through a query which linked fields 'lat.DD' and 'long.DD to latitude and longitude of the main database, and precisely replaced all blank spaces.

Discussion and Conclusions

The issue of collection records not being 'fit-for-use' is huge and vital, but major concerns have focused on certain aspects of the problem (accuracy, management) without saying much about readability, tone, and interest. Perhaps one of the most exciting research directions for the use of database collections is to focus on how success through implementation of either digitisation or GRAP 103 projects is evaluated. Given the current biodiversity initiatives in South Africa, an immediate benefit to fully and effectively leverage these collections for research should not be overlooked. Even in their current state, the KwaZulu-Natal Museum collection databases have informed biodiversity projects nationally and internationally, and georeferencing the Bivalvia database has thoroughly added value to records that were poorly sampled. Data from this database have been used extensively in Professor Herbert's research publications, and supplied to the following national projects:

1. Distribution data on alien terrestrial molluscs for The National Status Report on Biological Invasions and their Management in South Africa in 2017. See: van Wilgen, B.W. and Wilson, J.R.U. (eds), in prep.
2. Distribution data on Karoo endemic snails used in the impact assessment for the proposed shale gas fracking in the Karoo. See: CSIR, 2016.
3. Marine mollusc data, including bivalves, will also be included in Atkinson, L. and Sink, K. (eds), in prep.
4. AfrOBIS: a marine biogeographic information system for sub-Saharan Africa. See: Grundligh et al, 2007.

As a research institution, it is important that our databases are correctly cleaned and accurately georeferenced with the fewest possible errors. Goodwin et al (2015) argue that data quality is an important consideration in herbarium digitisation, which is essential if the potential of herbaria for enhancing our understanding of key questions in systematics, biogeography, and environmental studies are to be realised (Penn et al, 2018). However, it is still recommended that the end-users of these datasets assess the quality and the accuracy of the data, in order to inform land-use planning and decision making. Robertson et al (2016) developed an R package, *biogeo*, tool for the detection and correction of errors (data cleaning) and for assessment of data quality of collections datasets consisting of occurrence records. This R package, *biogeo*, could transform museum collection databases, especially during data cleaning and quality assessment before or after data are georeferenced.

Georeferencing provides many advantages for data use in various capacities. For instance, the ability to identify geographical data gaps and to define priorities for collection. This is particularly important when aiming to link different data types and sources, such as floristic and trait data (Spehn and Korner, 2010). The other important value of georeferencing is the ability it provides to link the database's original content with other georeferenced data contained in other databases (Spehn and Korner, 2010). Although these were not implemented, it is anticipated that the digitised records (through georeferencing) of the Bivalvia collection database will ultimately be linked to other databases, and used to update coordinates to these other datasets. This strategy will allow coordinates from the Bivalvia database to be transferred to records of other databases with similar locality descriptions without undertaking the full exercise of georeferencing. In this way, valuable time and money will be invested effectively. Also, efforts to produce better and sustainable database collection management applications that maximise effective sharing of biodiversity information should be encouraged.

Acknowledgements

The author is grateful to the KwaZulu-Natal Museum, Prof. D.G. Herbert for reading the manuscript, providing support during the project and comments to improve the manuscript, and B. Muller for providing the georeferencing training.

References

Atkinson, L. and Sink, K. (eds), in prep. *Field Guide to Offshore Marine Invertebrates of South Africa*. Pretoria: Malachite Marketing.

Berent, P., Hamer, M., and Chavan, V., 2010. Towards demand-driven publishing: Approaches to the prioritization of digitization of natural history collection data. *Biodiversity Informatics*, 7, pp.47-52.

BioGeomancer Working Group, 2005. *BioGeomancer*. [online] Available at: < https://sites.google.com/site/biogeomancerworkbench/home>.

Chapman, A.D., 2005. Uses of Primary Species Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

Chapman, A.D., and Wieczorek, J. (eds), 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility.

CSIR, 2016. *Shale gas development in the Central Karoo: A scientific assessment of the opportunities and risks*. [online] Available at: <http://seasgd.csir.co.za/scientific-assessment-chapters/>.

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R.I., and Scotland, R.W., 2015. Widespread mistaken identity in tropical plant collections. *Current Biology*, 25, R1066–7.

Grundlingh, M.L., von St. Ange, U.B., Bolton, J.J., Bursey, M., Compagno, L., Cooper, R., Drapeau, L., Griffiths, C.L., bel Hassen, M., Herbert, D.G., Kirkman, S., Ohland, D., Robertson, H.G., Trinder-Smith, T., van der Westhuysen, J., Verheye, H.M., Coetzer, W., and Wilke, C., 2007. AfrOBIS: A marine biogeographic information system for sub-Saharan Africa. *South African Journal of Science*, 103, pp.91-93.

Kilburn, R.N. and Herbert, D.G., 1994. 'Then a-dredging we will go, wise boys' – an outline of the Natal Museum Dredging Programme. *South African Journal of Science*, 90, pp.446-448.

Paterson, G., Albuquerque, S., Blagoderov, V., Brooks, S., Cafferty, S., Cane, E., Carter, V., Chainey, J., Crowther, R., Douglas, L., Durant, J., Duffell, L., Hine, A., Honey, M., Huertas, B., Howard, T., Huxley, R., Kitching, I., Ledger, S., McLaughlin, C., Martin, G., Mazzetta, G., Penn, M., Perera, J., Sadka, M., Scialabba, E., Self, A., Siebert, D., Sleep, C., Toloni, F., and Wing, P., 2016. iCollections – Digitising the British and Irish Butterflies in the Natural History Museum, London. *Biodiversity Data Journal*, 4, e9559.

Penn, M.G., Cafferty, S., and Carine, M., 2018. Mapping the history of botanical collectors: spatial patterns, diversity, and uniqueness through time, *Systematics and Biodiversity*, 16, pp.1-13.

Rios, N.E., and Bart, H.L., n.d. *GEOLocate - Software for Georeferencing Natural History Data*. [online] Available at: <www.museum.tulane.edu/geolocate/web/WebfileGeoref.aspx>.

Robertson, M. P., Visser, V. and Hui, C. 2016. Biogeo: an R package for assessing and improving data quality of occurrence record datasets. – *Ecography* 39: 394–401.

Spehn, E.M. and Korner, C. (eds.), 2010. *Data Mining for Global Trends in Mountain Biodiversity*. Boca Raton, Florida: CRC Press, Inc.

van Wilgen, B.W. and Wilson, J.R.U. (eds), in prep. *The status of biological invasions and their management in South Africa in 2017*.

Appendix I

Table 1. Major fields of the original Bivalvia database and their descriptions.

Field	Description
Accession no.	The catalogue number assigned to the specimen in the database.
Family	Family of the specimen by taxonomical classification.
Genus	Genus of the specimen by taxonomical classification.
Species	Species of the specimen by taxonomical classification.
Author	Person who first described the species.
Station no.	Number assigned to describe where the specimen was found. This number is rarely used for land snails but commonly used during collection of marine molluscs. The number is assigned to label the material, while information associated with the number is kept in the field book. This method simplifies re-writing of information on each label and makes it easier to query information on the label and in the field book.
Country	Country in which the specimen was collected.
Region	State or province within the country where the specimen was collected.
Locality	a) The position of a feature in space; b) The verbal representation of this position (i.e., the locality description) (Chapman and Wieczorek, 2006).
Latitude	Describes the angular distance that a location is north or south of the equator (degrees, min., sec.), measured along a line of longitude (*q.v*) (Chapman and Wieczorek, 2006).
Longitude	Describes the angular distance that a location is east or west of the prime meridian (*q.v*) (degrees, min., and sec.) on the earth's surface along a line of latitude (*q.v*) (Chapman and Wieczorek, 2006).
Depth/Altitude	How deep in the sea or height from the ground the specimen was found.
Day	Calendar day the specimen was collected (very important for database query).
Month	Calendar month the specimen was collected (very important for database query).
Year	Calendar year the specimen was collected (very important for database query).
Collector	Person(s) who collected the specimen.
Habitat	Brief description of the ecological place of collection.
Source	Information on whether the specimen was donated/purchased, etc.
Notes	Additional description of the locality and how the specimen was collected. e.g. Dived, dredged.
Determiner	Person who identified the specimen.
Other	Additional description of the locality and how the specimen was collected. e.g. Dived, dredged.
Cupboard	Place where the specimen is kept or stored in the collection room.
Institution	Organisation in charge of keeping the specimen eg. KZN-Museum.
Lat. DD	The latitude coordinate (in decimal degrees) at the centre of a circle encompassing the whole of a specific locality. Convention holds that decimal latitudes north of the equator are positive numbers less than or equal to 90, while those south are negative numbers greater or equal to -90. Eg. -42.5100° is roughly the same as 42°30'36" S (Chapman and Wieczorek, 2006). This is very important for mapping purposes.
Long. DD	The longitude coordinate (in decimal degrees) at the centre of a circle encompassing the whole of a specific locality. Decimal longitudes east of the Greenwich Meridian are considered positive and less than or equal to 180, while western longitudes are negative and greater than or equal to -180. Eg. -122.4900° is roughly the same as 122°29'24" W (Chapman and Wieczorek, 2006). This is very important for mapping purposes.
Entry date	Exact date the specimen was databased.
L/D	Live or Dead. If Live, it is usually followed by LPT (was found live, is Preserved in alcohol and Tissue was taken for DNA analysis).
Habitat type	Ecological niche description where the specimen was found.
Accuracy	How accurate are the GPS coordinates? Are the coordinates for the exact place where the specimen was found? Or for the whole region or game reserve etc.?
Collection date	Primary collection date of the specimen (in full format).
Databased by	The person who captured the record in the database.

Table 2. Examples of ambiguous and poor locality descriptions that did not provide geographic information during georeferencing of the Bivalvia database.

Locality	Country	Region	County	Habitat description
20 mí. East of San Juan, Bahia de San Juan	PUERTO RICO	San Juan	not georef.	Among strangled seaweed
Alexandra Junction	SOUTH AFRICA	KwaZulu-Natal	not georef.	
Anchor Reef, off Inhagonda area	MOZAMBIQUE		not georef.	
Labronico Sea	ITALY		not georef.	
Mainland	TANZANIA		not georef.	
Malaya	THAILAND	Penang	not georef.	
North Sea: Near Dogger	UK		not georef.	
Off Somali Republic	SOMALIA		not georef.	
Okhotsik Sea: Tauyskaya Guba, Nagaeva Bay	JAPAN		not georef.	
Persian Gulf: As Shaam	KUWAIT		not georef.	Sand among coral rubble

Table 3. Descriptions of fields included in the CSV spreadsheet for GEOLocate Web application tool.

** information should be filled through Google Earth search to identify the country/region they are currently associated with.*

*** information expected, otherwise geographic information will not be provided.*

Field	Description
Locality**	a) the position of a feature in space; b) The verbal representation of this position (i.e., the locality description).
Country*	State of entity.
Region*	State or province within the country where the specimen was collected.
County	If the locality cannot be found or is confusing, it was annotated 'not georef' and later checked for review. This is most convenient and can occur in the database itself. Attempt was made to correct the spelling (if applicable) or verify the locality description on Google Earth (Chapman and Wieczorek, 2006).
Lat. DD	See Table 1. If the locality description matches the spatial representation, geographic information will be added in Degrees decimal.
Long. DD	See Table 1. If the locality description matches the spatial representation, geographic information will be added in Degrees decimal.
Correction status	Labelled 'yes' if correction was made during evaluation and assessment of a record and 'no' if no georeferencing took place.
Precision	With measurements and values, it describes the finest unit of measurement used to express that value (Chapman and Wieczorek, 2006).
Error polygon	Geographic information will be added at the end of georeferencing.
Multiple results	Geographic information will be added at the end of georeferencing.
Radius uncertainty	The unit in length in which the uncertainty is recorded (eg., mi, km, m and ft).
Radius uncertainty (circular polygon)	The upper limit of the distance from the given latitude and longitude describing a circle within which the whole of the described locality must lie (Chapman and Wieczorek, 2006).
Habitat description*	Describe the ecological sphere of the habitat. e.g. Fine sandy and muddy.
ID	Assigned number to confirm and facilitate the export of georeferenced records into the main database. This number is assigned to both the accession number and the locality description during filtering and sorting of the main database so that it complies with the field requirements of the GEOLocate tool.