

https://dspace.org/

Putting fossils on the map: Applying a geographical information system to heritage resources

Van ver Walt, M.; Cooper, A.K.; Netterberg, I.; Rubidge, B.S.

2015-12
Academic of Science of South Africa (ASSAf).
http://hdl.handle.net/10539/21606
Van ver Walt, M. et al. 2015. Putting fossils on the map: Applying a geographical information system to heritage resources. South African Journal of Science 111(11/12): Article #2014-0371.

Downloaded from a DSpace repository. https://dspace.org/

Putting fossils on the map: Applying a geographical information system to heritage resources

AUTHORS:
Merrill van der Walt[1]
Antony K. Cooper[2]
Inge Netterberg[3]
Bruce S. Rubidge[1]

AFFILIATIONS:
[1]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Built Environment Unit, Council for Scientific and Industrial Research, Pretoria, South Africa
[3]Netterberg Consulting, Leuven, Belgium

CORRESPONDENCE TO:
Merrill van der Walt

EMAIL:
Merrill.VanDerWalt@wits.ac.za

POSTAL ADDRESS:
[1]Evolutionary Studies Institute, Private Bag 3, University of Witwatersrand 2050, Johannesburg, South Africa

DATES:
Received: 23 Oct. 2014
Revised: 26 Mar. 2015
Accepted: 20 Apr. 2015

KEYWORDS:
Spatial data; Permo-triassic; database; Karoo Supergroup

HOW TO CITE:
Van der Walt M, Cooper AK, Netterberg I, Rubidge BS. Putting fossils on the map: Applying a geographical information system to heritage resources. S Afr J Sci. 2015;111(11/12), Art. #2014-0371, 7 pages. http://dx.doi.org/10.17159/sajs.2015/20140371

© 2015. The Author(s). Published under a Creative Commons Attribution Licence.

A geographical information system (GIS) database was compiled of Permo-Triassic tetrapod fossils from the Karoo Supergoup in South African museum collections. This database is the first of its kind and has great time applicability for understanding tetrapod biodiversity change though time more than 200 million years ago. Because the museum catalogues all differed in recorded information and were not compliant with field capture requirements, this information had to be standardised to a format that could be utilised for archival and research application. Our paper focuses on the processes involved in building the GIS project, capturing metadata on fossil collections and formulating future best practices. The result is a multi-layered GIS database of the tetrapod fossil record of the Beaufort Group of South Africa for use as an accurate research tool in palaeo- and geoscience research with applications for ecology, ecosystems, stratigraphy and basin development.

Introduction

The fossil record of the Karoo Supergroup, which comprises a largely unbroken temporal record of tetrapod evolution from the Middle Permian to the Middle Jurassic,[1,2] provides a unique opportunity to set up a GIS database of fossil occurrences which can be utilised to answer questions relating to ecological and biodiversity change through time. The Karoo fossil record is the best preserved ecological assemblage of pre-mammalian terrestrial tetrapods documenting the stem lineages of both mammals and dinosaurs.[3,4]

We geocoded palaeontological data for use in a geographical information system (GIS) for palaeoscience research to explore issues relating to the biodiversity of Permian and Triassic tetrapod faunas. This was the first time a GIS had been applied to the fossil records of the Karoo Supergroup. With the cooperation of seven South African museums and institutes (Council for Geoscience, Pretoria; Ditsong Museum, Pretoria; Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg; National Museum, Bloemfontein; Albany Museum, Grahamstown; Rubidge Collection, Wellwood, Graaff-Reinet; Iziko South African Museum, Cape Town) that curate collections of Karoo tetrapod fossils, a GIS incorporating the South African databases of fossil records collected from the Beaufort Group, Karoo Supergroup has been compiled.

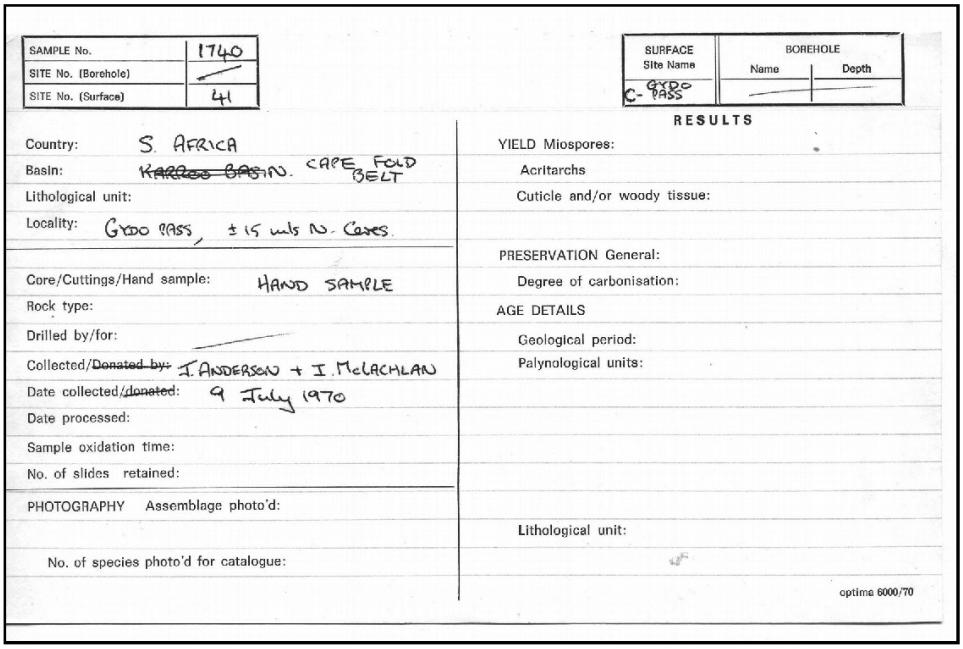
The hundreds of thousands of fossil artefacts stored and accessioned in museum collections are the foundation of our knowledge on past biodiversity. Great strides have been made in biodiversity informatics in providing digital access to extinct biodiversity data, for integration, interpretation, reconstruction and application objectives. Models for community data access are evident in abundant projects, such as:

- The Revealing Human Origins Initiative (RHOI)[5] Specimen Database, a collaboration of paleoanthropological and related projects studying Late Miocene (and Pliocene) hominins and other faunas in context, with the database including digital imagery and metadata that covers age, geology, collection elements and taxonomy;
- The digital@rchive of Fossil Hominoids[6], for which the primary mandate is to facilitate morphological investigations in the field of human evolution by providing digital data for the international scientific community;
- The Darwin Core metadata standard[7], a uniform standard designed to expedite the exchange of information about the geographic occurrence of species and specimen records in collections, with extensions for palaeontology.

These information systems, driven by distributed data retrieval, data capture and person-facilitated geospatial referencing, have enabled the investigation of novel research questions around ecological reconstruction, extinct biodiversity trends and predictive modelling.

Historically, details of fossils collected were recorded as hand-written descriptions on index cards or in catalogues (Figure 1). Such documentation included both data (e.g. species and location) and metadata (information about the record), such as who collected, prepared and/or identified the fossil, where the fossil is stored and who wrote up the index card.

There are a variety of standards available for metadata, such as the Dublin Core (ISO 15836:2009),[8] developed primarily for describing resources for discovery, and ISO 19115:2003,[9] for describing geographical data, of which the South African profile (subset) is SANS 1878-1:2005.[10] Dublin Core is primarily text-based, making it easy to enter information for its 15 metadata elements, while ISO 19115 makes extensive use of encoding, which facilitates automated processing and presenting the metadata in multiple languages. Metadata can be converted from one standard to another using an ontology or a cross-walk.[11] As ISO 19115 has encoded metadata and more detailed metadata elements, it is easy to convert its metadata to Dublin Core through a cross-walk (conversion table), but the reverse is difficult because of the need to subdivide metadata elements, text processing and, invariably, use human expertise. Hence, it would be better to use a metadata standard such as ISO 19115 for palaeontological records.



Source: *Evolutionary Studies Institute (ESI) Collections. Courtesy of Dr. Bernhard Zipfel.*

Figure 1: *Collected fossil recorded as a hand-written description on an index card.*

South African fossil-find field notes for the Beaufort Group (to be eventually reconfigured into museum index cards) were written up over the space of 150 years (since 1845[12,13]) and do not conform to any particular standard. The main disadvantage is that some records contain inadequate or ambiguous data, particularly relating to the precise location of the fossil provenance.

This paper focuses on the processes involved in establishing a GIS for tetrapod fossils from the Beaufort Group. It highlights the key challenges encountered during database establishment, as well as describing its main applications and future best practices for use as an accurate research tool in palaeontological research. This unique database is curated at the Evolutionary Studies Institute (ESI) at the University of the Witwatersrand and is available as a research tool to all bona-fide scientists

Creating a reliable product

Extensive fossil collections have been amassed from the rocks of the time-extensive Permo-Triassic Beaufort Group and curated in different museum collections in South Africa, providing a unique opportunity to incorporate these collections onto a GIS. Ultimately, this database will be expanded to include fossils from the Beaufort Group which are housed in overseas institutions such as the Natural History Museum, London; Smithsonian Institution, Washington DC and the Field Museum, Chicago. This GIS will enhance their utility in research relating to changing biodiversity patterns, both temporally and geographically, as well as stratigraphic and basin development modelling.

Problems that had to be overcome in setting up the GIS database related largely to a lack of consistency in the data, ambiguous locality data and outdated taxonomic records, requiring rigorous standardisation and updating.

While all the original data was provided in digital format, these were set up from manual records. This is the main drawback encountered when having to apply human interpretation verses the structured logic of the computer. The establishment of the GIS database highlighted the value of structuring data to suit GIS and other digital applications. The migration of paper records to useful electronic records could not simply be carried out verbatim as many of the data obtained from the contributing South African museums needed to be restructured to facilitate analysis through electronic means.

Table 1: Summary of processes for presenting data within a spatial system

Task order	Processes		
	Stage 1: Acquisition and processing of original data	**Stage 2: Establishing a GIS management system**	**Stage 3: Reconciliation**
1	Review and assimilation of original databases	Selection of map layers	Challenges encountered during mapping process
2	Creation of applicable data fields	Conversion of data	Future possible modes of action concerning standardising and fossil collection methods
3	Criteria established for the elimination of non-viable data with subsequent deletion	Establishing an alternative base data	Suggestions for improving the current success rate for digitising fossil localities
4	Division into automated and manual entry databases	Establishing a geodatabase	
5	Creation of map sheet shape file		

Table 2: Summary of specific processes for presenting digitised Beaufort Group data

Task order	Phases	
	Phase 1: Accessing and processing of data	**Phase 2: Development of spatial model**
1	Initial test case using the recorded fauna of the Evolutionary Studies Institute (ESI) database	Converting alphanumeric data to spatial data
2	Selection of datafields applicable for project objectives	Acquistion and application of selected map layers
3	Preliminary data cleaning	
4	Establishment of criteria list allowing for automatic validation	
5	Selection of data allowing for automatic insertion into GIS	
6	Establishment of manual insertion data	
7	Establishment of genera and species lists	
8	Acceptance or re-classification of taxonomic classification	
9	Establishment of final genera/species list for ESI database	
10	Assignment of fossil locality data to each specimen	
11	Repetition of tasks 2 – 10 for remaining six databases (ESI as template)	

Methodology

Mapping palaeontological specimens

As this database has been set up as a research tool to be used by palaeontologists, it is important to explain the methodology in detail so that users can fully understand why the GIS was created in this particular way.

The broad-spectrum processes were divided into three stages (Table 1): Stage 1: Acquisition and processing of original data; Stage 2: Establishing a GIS management system; Stage 3: Reconciliation.

More detailed processes involved in spatially mapping the fossils were subdivided into two phases (Table 2), Phase 1: accessing and processing of data and Phase 2: development of a spatial model.

Alphanumeric data was converted to spatial data because, for most of the records, the location was specified using geographical identifiers,[14] particularly farm names, rather than coordinates. Converting the data required rigorous 'cleaning', correction of spelling errors and standardisation of content to permit queries. Farm names with their corresponding farm numbers were aligned with the names registered with the Registrar of Deeds and the Surveyors General.

Once cleaning of data had been accomplished, selection of data fields applicable for spatial mapping was undertaken. Geospatial coordinates used for mapping species location and distribution are crucial for a reliable spatial system.[15] Access to geospatially referenced data from fossils provides a quantitative basis for biodiversity analyses over time and predictive niche modelling for determining sampling densities of various sites.

Providing locality coordinates proved a significant challenge. Most of the recorded specimens were associated with a georeference, but this reference was, in most instances, a worded description of the localities from where they were discovered with few records having geographic coordinates (Table 3).

To get to the point where data could be represented on a spatial map, two approaches were adopted. The first involved selecting records that qualified for automatic import into the system. The second approach involved records that could only be entered onto the system manually.

Table 3: Challenges posed by textual geo-referencing

Examples of textual geo-reference	Challenge posed
'Matjiesfontein' 'Bronkhorstspruit'	Same locality name for more than one record
'Ndanyana Hill on Hartebeespruit'	Topological nesting
'Tiparery (?Tipperary) Ndanayane? (Ndanyana) Hill'	Complex grammar
'0.25 miles NW of Mnweni River' 'Hantam Mts, Toren, 12 miles W of Calvinia'	Linear feature measurement
'On the road between Masite and Kolo'	Linear ambiguity
'Various sites in Bamboeshoek Valley' 'Possibly from Kiesbeen behind Platberg' 'Between the road and the Sundays River'	Vague localities
'Transvaal, Wiepe'	Change of political borders over time
''Fossil Bend' near Harrismith' 'Newcastle Admiralty Estate Group'	Historical and 'in-house' placename

Automated data entry procedure

Records with locality coordinates from a Global Positioning System (GPS) could be entered automatically. However, as the majority of records had only a farm name for the locality, a spatial database had to be created to allow records to be imported automatically to specific localities referenced as farm centroids. A farm centroid is the calculated gravitational centre of a polygon (farm boundaries are polygons). This centroid is calculated using the ArcMap field calculator which automatically sets a field value for a single record, or even all records.

Forcing such localities into a single point at the gravitational centre of the farm introduces error and inaccuracy into the data, but remains the best option to utilise locality data for the majority of fossils found prior to GPS usage. Current locality data were accurately captured by GPS.

Table 4: Methodology for automating spatial data

Phase		Task(s)			
1	**Obtain necessary data**	1. Landsat imagery	2. Geology (1:250 000)	3. Farms and farm boundaries	4. Fossil records
2	**Transform spatial data**	1. Data exploration	2. Check projections of spatial data for accuracy and relevance		
3	**Sort alphanumeric data**	1. Sort records per farm and per biozone			
4	**Digitise fossil record localities**	1. Digitise localities per farm using ArcGIS®	2. Align ID keys with alphanumeric database		
5	**Link alphanumeric database**	1. Link alphanumeric database to spatial database (compound key)			
6	**Output**	1. Spatial database of tetrapod fossil localities for the Beaufort Group of South Africa			

To allow data to be imported automatically, certain tasks had to be completed (Table 4). A geodatabase was created to house the spatial data of the farm, administrative, district and magisterial boundaries and local authorities' databases.[16] Various map layers (including Landsat 7 ETM+ Satellite imagery) were necessary as backdrop data to interpret the distribution patterns of fossil taxa.

Because most of the specimens in older collections lacked geographic coordinates for their place of discovery, the most accurate locality information in the majority of the databases was simply a farm and district name. To represent this locality information on the GIS, farm locality data was received in .FEA format from the Surveyor General and converted into shape file format. Alphanumeric data were exported as a point file and joined to the polygon data using a spatial join. The cadastre received from the Surveyor General contained farm boundaries and their farm numbers, but very few farm names. This lack of names posed a problem as localities for most of the specimens in the museum catalogues were given as locality names, which were assumed to correspond to the farm names. As such, the farm names were essential for the geocoding of the localities and thus the specimens.

To solve the number versus name problem, Environmental Potential Atlas (ENPAT 2004[17]) farm cadastre data was used as the new spatial layer to identify localities. For each farm, centroids were generated and used to geocode the specimens by linking the specimen locality names to farm names. Additional backdrop map layers included Surveyor General data for magisterial districts and provinces. These data were used to identify further localities, as farm names are not unique across the country. Digitised geological maps covering the extent of the Beaufort Group were included as additional backdrop data.

The Evolutionary Studies Institute (ESI) collection database was selected as the test case because of the high resolution of farm locality and map sheet data, to determine whether automated entry of palaeontological records was a feasible option. Unique localities were split into those with coordinates and those without, as the process for identifying the location of these two groups of localities was different[16].

Those localities with grid coordinates were extracted and all coordinate data converted to decimal degrees, imported into ArcGIS® as an event theme, converted to a shape file, and each specimen was located as a point in the spatial data file.

Localities that lacked coordinates were identified by districts, farm names and map sheet indices. As farm names are not unique and can be repeated for several districts, the map sheet index was used in addition to the farm names and districts as identifiers for the location of the farm localities. As an index shape file of the 1:50 000 map sheet series does not exist, a map sheet index shape file was created by digitising the sheets.

As a test run to determine how to automate the linkage of the locality name provided by the ESI data to farm records listed in the Surveyor General data, 13 distinct localities in the district of Beaufort West were selected (Table 5). According to the alphanumeric data, all these localities fall on the same map sheet except for the Winterberg (Gryskop) locality. Of the 13 localities, only seven were matched to the spatial data and of these only two localities fell on the correct map sheet,[16] indicating it would be difficult to automate the linking process.

Table 5: Unique localities in the Beaufort West District

Locality name	District
Beaufort West Commonage	Beaufort West
Biesjiespoort (on Nobelsfontein)	Beaufort West
Buffelsvlei	Beaufort West
Dunedin (formerly Quaggasfontein)	Beaufort West
Elandsfontein	Beaufort West
Heuninglaagte (Honinglaagte)	Beaufort West
Kuilspoort	Beaufort West
Leeu Kloof	Beaufort West
Leeufontein	Beaufort West
Leeukloof	Beaufort West
Rosary	Beaufort West
Uitspansfontein	Beaufort West
Winterberg (Gryskop)	Beaufort West

Another test was run to determine if 'selection by map sheet' could be used as a method to link alphanumeric data to spatial data. Map sheet 3123DD was randomly selected and alphanumeric records of the ESI collection located on this map sheet were selected, returning 41 records. These records were then queried such that only distinct localities would be returned, and resulted in 38 localities. As locality name should correspond to farm name, it follows that there should be 38 farms which intersect with this map sheet. A query was performed to select all the farms which lie wholly or partly on this map sheet and resulted in 16 farms - less than half the number of distinct localities. As the method of using locality name, farm name and farm centroid was not effective (because neither locality name nor farm name matched the government farm name), an alternative linkage solution needed to be created.

The database of Iziko South African Museum, which contains both locality names and formal government farm names, was used as the linkage mechanism. For each collection, a query for distinct records of specimens was run. Results from this query were input into a second query where the centroids with their government farm names were linked to the specimens with their locality names via the joining table derived from the Iziko South African Museum data. The results from these queries were imported into the GIS and merged to form one data file.

The automatic import of records yielded a poor success rate. Of the 19 718 specimens selected for automatic import, 8512 records were imported automatically. Against the total number of records (20 968) that required import into the GIS, this number indicates a success rate of 43.2%. The only solution was for the remaining data to be manually imported into the existing digital-palaeo system.

Manual data entry procedure and collation of Beaufort Group data

Data tables containing specimens that had failed to import automatically were established for each museum and imported into the spatial map under separate map layers. Each fossil entry was then systematically added in point format to the spatial map.

The final phase involved collating the point data of the seven collections into a single map layer to decrease the complexity of future queries. The original intention was to append all the data sets onto a single data set, but because of differing alphanumeric table structures and schemes of the contributing museums, a single data set was not directly possible. Accordingly, the alphanumeric table structures of the various data files were manipulated to conform to a designed standard structure (Table 6).[16]

Table 6: Alphanumeric table structure

Field name	Data type
Id	Short Integer (6)
Specimen_I	Text (50)
Current_Tax	Text (50)
Taxon_Assi	Text (50)
Biozone	Text (50)
Province	Text (50)
District	Text (50)
Locality_N	Text (50)
Mapsheet	Text (50)
Old_ID	Text (50)
Census_Dis	Text (50)
Farm_Name	Text (50)
Co_ord_X	Double (8.7)
Co_ord_Y	Double (8.7)

Shape files that had a multipoint geometry type, rather than a point geometry type, were converted to a point geometry type, and appended, resulting in a single point shape file. This file contains approximately 30 000 points and indicates where fossil vertebrates from the Beaufort Group were discovered in the field.[17]

Applying GIS to palaeontological research

Establishment of the Beaufort Group GIS has resulted in a research aid which can be utilised to answer questions relating to Permo-Triassic continental vertebrate biodiversity. The system displays taxonomic diversity of fossil tetrapods from the Beaufort Group, with specific reference to the housing of particular genera, including numbers of specimens of each genus and their respective locality and/or biozone data. This unique data set provides a record of fossil tetrapod biodiversity in the continental realm from the Middle Permian to the Middle Triassic, and shows accurate numbers of specimens of the various taxa which have been collected (e.g. Smith et al.[18]).

The spatial map allows for queries to be performed relating to geospatial distribution, which may assist in understanding the pattern of basin infill as well as significant biodiversity patterns. Because of the paucity of reliable lithological markers which extend over a large area as well as datable volcanic rocks, index fossils have been used as biozone markers to develop basin development models.[19-22] Now that the database of Karoo fossils is available on a GIS, biogeographic distribution patterns of index taxa can be determined and utilised in the development of basin development models. In addition, because the tropic level of each species is recorded in the database, changes in terrestrial vertebrate ecological relationships from the Middle Permian to the Mid Triassic have been explored.[23]

The database elucidates which genera occur in particular biozones, as well as numbers of individuals of each genus in successive biozones. Application of the GIS can highlight biodiversity changes across successive biozones, making it possible to calculate the extent of extinction of taxa within successive time slices, which allows for the determination of trends in biodiversity changes through time.

Through the application of the GIS database, the stratigraphic ranges of fossil specimens were calculated and a refined biostratigraphic subdivision of the *Eodicynodon*, *Tapinocephalus* and *Pristerognathus* assemblage zones (Middle Permian Beaufort Group) proposed.[24]

Previously, biostratigraphic maps of the Beaufort Group were compiled based on a rough estimate of the distribution of biozone-defining signature fossil genera. Through the utilisation of the GIS database, a far more precise biozone map of the Beaufort Group has been produced (Figure 2), with the capability of being continuously updated as new information is received. This method introduces an entirely new way of representing geographical fossil distribution data that can be used in basin development and tetrapod biogeographic studies.[25]

Future spatial map enhancement and maintenance

Given the scope of both the specimen locality data and the necessity for this data to be available in a readily usable form, efficiency and accuracy is of prime importance for the task of geospatial referencing. Individual institutions housing palaeo-collections typically lack the resources or informatics expertise to meet the challenges of georeferencing alone.[15] Designing a collaborative geospatial referencing methodology using the combined expertise of all palaeontologists in South Africa has yielded an accurate and reliable spatial map.

An important aspect of data management that will improve reliability of the data set is the accurate assignment of taxonomic data. This accuracy is crucial for trend analysis. The updating and verification of both biozone assessment and taxonomic assignment must be accomplished through ongoing collaboration of palaeontologists, each selected for their expertise in specific taxa of animals and/or geological expertise. Specimens in collections need to be checked and identifications must be updated using current taxonomic diagnoses.

The application of GIS technologies could have significant impact as it could open further avenues of GIS-based research in palaeontology. The data could be used for four-dimensional (4D) spatio-temporal modelling. These dimensions can be distorted by geological or other processes and hence there are different spatio-temporal dimensions that are relevant for the fossils, including the various 4D environments through which the fossil has been taken, changed, moved, etc., such as weathering, erosion, re-deposition, lithification, metamorphism, diagenesis, faulting, folding, etc.

Ongoing contribution to the GIS project involves the further development and refinement of the Spatial Map of Beaufort Group fossil specimens (e.g. Van der Walt et al.[25]). Continual refinement and upgrading will ensure a reliable product that may confidently be used as an analytical and research tool.

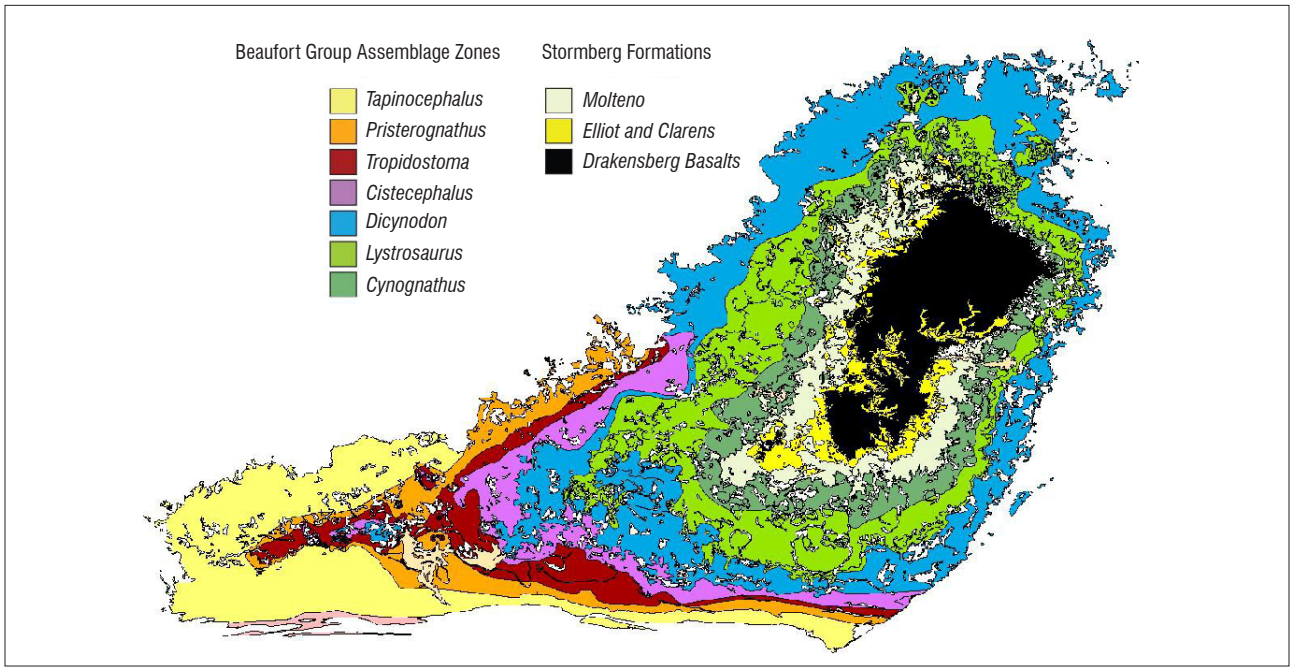


Figure 2: Defined biostratigraphic map of the Beaufort Group, Karoo Supergroup of South Africa – produced through the application of GIS technology.

Conclusion

Digital archiving and spatial cataloguing of artefact collections data are central to biodiversity research. The way forward for the digitised spatial map of the Beaufort Group of the Karoo Basin is to create easier access for researchers, where currently access to collections is limited by their distributed nature on museum shelves.

Combining all the South African databases of vertebrate fossils from the Beaufort Group onto a geographical information system is an important tool which could be utilised to address numerous different issues relating to Permo-Triassic continental tetrapod biodiversity in space and time. In the process of setting up this GIS, it became apparent that the course of action is not simply a matter of combining all the databases and coming up with answers.

Many metadata have been collected for recorded fossils from the seven large collections of vertebrate fossils in South Africa, but, particularly for specimens collected long ago, the quality of the metadata was not initially compliant with the requirements for a reliable GIS. Lack of complaince is largely because of a lack of consistency of data in the different collections, and also because precise GPS coordinates of localities of many of the fossils were not recorded, and only a farm locality was given. Certain specimens were listed more than once, despite the selection of distinct records. This is because of capture errors and lack of standardisation in the original data. In a number of the spreadsheets, the supposedly unique identifiers of some specimens were not actually unique, which made removing redundant data much more difficult.

Since overcoming these obstacles, the outcome is a GIS database that is reliable and applicable. The result is the establishment, for the first time, of a Geographical Information System incorporating all vertebrate fossils from the rocks of the Beaufort Group which are curated in South African Museums, have already been utilised to address questions relating to Permo-Triassic biodiversity and palaeoecological change.[18,23,26-28] As the Karoo geological succession preserves the most complete record of Middle Permian to Middle Triassic continental tetrapod biodiversity, a study of this nature is of great importance for an understanding of early continental tetrapod biodiversity changes as well as for Gondwanan basin modeling.

The objective of the established GIS database is for its utilisation as both a research tool and digital archive and as such, it requires continuous improvement, updating and refinement. Future work will include standardising the records for the fossils, for both data and the metadata.

Acknowledgements

The authors are indebted to the Council for Scientific and Industrial Research, the Department of Science and Technology, the National Research Foundation, the Centre of Excellence in Palaeoscience, the Palaeontological Scientific Trust and the University of the Witwatersrand for financial assistance. We thank and have benefitted from numerous discussions with many researchers: Fernando Abdala, Rose Prevec, Mike Raath, Roger Smith and Mike Day. Special thanks to Billy De Klerk (Albany Museum), Bernhard Zipfel (Evolutionary Studies Institute), Johann Neveling (Council for Geoscience), Elise Butler (National Museum), Richard, Robert and Marion Rubidge (Rubidge Collection), Sheena Kaal and Roger Smith (Iziko Museum) and Stephanie Potze (Ditsong) for allowing generous access to their museum records.

Authors' contributions

M.v.d.W. was the project leader and wrote the manuscript. A.C. made conceptual contributions and added to/edited the manuscript. I.N. performed the GIS queries; automated import and made experimental and conceptual contributions towards establishing the GIS database. B.R made conceptual contributions and added to/edited the manuscript.

References

1. Hancox PJ, Rubidge BS. Breakthroughs in the biodiversity, biogeography, biostratigraphy, and basin analysis of the Beaufort Group. J Afr Earth Sci. 2001;33:563–577. http://dx.doi.org/10.1016/S0899-5362(01)00081-1
2. Rubidge BS. 27th Du Toit Memorial Lecture Re-uniting lost continents–Fossil reptiles from the ancient Karoo and their wanderlust. S. Afr J Geol. 2005;108(1):135–172. http://dx.doi.org/10.2113/108.1.135
3. Broom R. The mammal-like reptiles of South Africa and the origin of mammals. London: HF & G. Witherby; 1932.
4. Nicolas MV, Rubidge BS. Assessing content and bias in South African Permo-Triassic Karoo tetrapod fossil collections. Pal Afr. 2009;44:13–20
5. The Revealing Human Origins Initiative (RHOI) [document on the Internet]. c2008 [cited: 2008 Oct 19]. Available from: http://rhoi.berkeley.edu

6. Virtual Anthropology. The digital@rchive of Fossil Hominoid [A-1909] [document on the Internet]. c2009 [2009 Jan 02]. Available from: http://www.virtual-anthropology.com/3d_data/3d-archive

7. Biodiversity Information Standards (TGWG). The Darwin Core [page on the Internet]. c2015 [2015 June 05]. Available from: http://wiki.tdwg.org/darwincore

8. International Organization for Standardization (ISO). ISO 15836:2009, Information and documentation — The Dublin Core metadata element set. Geneva:ISO; 2009.

9. International Organization for Standardization (ISO). ISO 19115:2003, Geographic information — Metadata. Geneva: ISO; 2013.

10. South African Bureau of Standards (SABS). SANS 1878-1:2005, South African spatial metadata standard, Part 1 — Core metadata profile. Pretoria: SABS; 2005.

11. Nogueras-Iso J, Zarazaga-Soria FJ, Lacasta J, Béjar R, Muro-Medrano PR. Metadata standard interoperability: Application in the geographic information domain. Comput Environ Urban. 2004;28(6):611–634. http://dx.doi.org/10.1016/j.compenvurbsys.2003.12.004

12. Bain AG. On the discovery of the fossil remains of bidental and other reptiles in South Africa. Transactions of the Geological Society of London. 1845:53–59. http://dx.doi.org/10.1144/transgslb.7.53

13. Cohen A. Mr. Bain and Dr. Atherstone: South Africa's Pioneer Fossil Hunters. Earth Sci Hist. 2000;19(2):175–191.

14. International Organization for Standardization (ISO). ISO 19112:2003, Geographic information — Spatial referencing by geographic identifiers. Geneva: ISO; 2003.

15. Beaman R, Wieczorek J, Blum S. Determining space from place for natural history collections. D-Lib Magazine. 2004;10(5):9. Available from: http://www.dlib.org/dlib/may04/beaman/05beaman.html

16. Cooper A, Netterberg I. Mapping palaeontological locations and specimens. icomtek document number: 0451-0001-701-A1. Pretoria: CSIR; 2004

17. Nicolas MV. Tetrapod biodiversity through the Permo-Triassic Beaufort Group (Karoo Supergroup) of South Africa. Johannesburg: University of the Witwatersrand; 2008.

18. Smith R, Rubige BS, Van der Walt, M. Therapsid biodiversity patterns and palaeoenvironments of the Karoo Basin, South Africa. In: Chinsamy-Turan A, editor. Forerunners of Mammals: Radiation Histology Biology. Indiana, IN: Indiana University Press; 2011. p. 223–246.

19. Catuneanu O, Hancox PJ, Rubidge BS. Reciprocal flexural behavior and contrasting stratigraphics: A new basin development model for the Karoo rhetoric foreland system, South Africa. Basin Res. 1998;10:417–439. http://dx.doi.org/10.1046/j.1365-2117.1998.00078.x

20. Rubidge BS. Biostratigraphy of the Beaufort Group. Pretoria: South African Committee for Stratigraphy, Biostratigraphic series No 1. Pretoria: Council for Geoscience; 1995; p. 1–2.

21. Hancox PJ. A stratigraphic, sedimentological and palaeonvironmental synthesis of the Beaufort-Molteno contact in the Karoo Basin. Johannesburg: University of the Witwatersrand; 1998.

22. Neveling J. Biostratigraphic and sedimentological investigation of the contact between the Lystrosaurus and Cynognathus Assemblage Zones (Beaufort Group: Karoo Supergroup). Johannesburg: University of the Witwatersrand; 2002.

23. Nicolas MV, Rubidge BS. Changes in Permo-Triassic terrestrial tetrapod ecological representation in the Beaufort Group (Karoo Supergroup) of South Africa. Lethaia. 2010;43(1):45–59. http://dx.doi.org/10.1111/j.1502-3931.2009.00171.x

24. Day MO. Middle Permian continental biodiversity changes as reflected in the Beaufort Group of South Africa: A bio- and lithostratigraphic review of the Eodicynodon, Tapinocephalus and Pristerognathus assemblage zones. Johannesburg: University of the Witwatersrand; 2013.

25. Van der Walt M, Day M, Rubidge B, Cooper AK, Netterberg I. A new GIS-based biozone map of the Beaufort Group (Karoo Supergroup), South Africa. Pal Afr. 2010;45:1–5.

26. Irmis RB, Whiteside JH, Kammerer CF. Non-biotic controls of observed diversity in the paleontologic record: An example from the Permo-Triassic Karoo Basin of South Africa. Palaeogeogr Palaeoclimatol Palaeoecol. 2013;372:62–77. http://dx.doi.org/10.1016/j.palaeo.2012.07.014

27. Fröbisch J. Vertebrate diversity across the end-Permian mass extinction – separating biological and geological signals. Palaeogeogr Palaeoclimatol Palaeoecol. 2013;372:50–61. http://dx.doi.org/10.1016/j.palaeo.2012.10.036

28. Abdala F, Rubidge BS, Van Den Heever JURI. The oldest therocephalians (Therapsida, Eutheriodontia) and the early diversification of Therapsida. Palaeontology. 2008;51(4):1011–1024. http://dx.doi.org/10.1111/j.1475-4983.2008.00784.x