

Poster B-49

BISI: an integrative, extensible and minable database of South African wildlife and indigenous livestock biomaterial information



Authors:

Athlee Maclear (*National Bioinformatics Network*)
Darren Otgaar (*National Bioinformatics Network*)
Dr. Junaid Gamielien (*National Bioinformatics Network*)
Dr. Paul Bartels (*National Zoological Gardens/BioBankSA*)
Dan Jacobson (*National Bioinformatics Network*)
Dr. Antoinette Kotze (*National Zoological Gardens, South Africa*)

Short Abstract: BISI is a collaborative project employing the graph storage and mining capabilities of DigaBase, to establish an integrative, extensible and minable database of South African wildlife and indigenous livestock biomaterial information, that will facilitate the identification of specific priorities for biodiversity conservation, research and biotechnology development in South Africa.

Long Abstract:

DigaBase (DGB), developed by the National Bioinformatics Network (NBN) of South Africa, is a graph-relational database with a digraph data storage architecture. Individual datum are stored as vertices, and the relations between them as directed edges. The digraph data structure is superior to conventional relational databases in that the employment of long-established graph mining algorithms is made possible. Thus far, transitive closure has been implemented and uses a custom graph query language (GQL) that is an extension of SQL, to retrieve all objects of a given type reachable by directed edges from a logical union of user-specified entry points.

The DGB architecture is powerful and unique as the integration of data from semantically rich knowledge domains, of which the biological sciences are prime examples, is greatly simplified. Biological queries are often prohibitively complex, and routinely require complicated scripts or purpose-built relational databases to integrate data. Moreover, such solutions are limited to answering a single or small selection of queries, while the DGB through the employment of the GQL allows such queries to be easily constructed and executed across the data in the graph.

The NBN envisages the application this technology to build a vastly integrated graph database enabling cross-disciplinary queries to be performed over a range of data and edge types, creating a potent hypothesis-driven data mining engine facilitating the discovery of yet unknown correlations between biological entities and events.

Currently, the construction of three initial data models is underway, a molecular mouse/human model, a plant model and a South African wildlife and indigenous livestock sample-tracking model. The latter is a collaborative project between the NBN and BioBankSA and has been dubbed BISI, BioBank Information System Initiative.

BioBankSA, whose operations fall under the National Zoological Gardens (NZG) of the South African National Research Foundation (NRF), is a cooperative of key stakeholders, driving the collection, value enhancing, banking and distribution of wildlife and indigenous livestock biomaterials viz. blood, serum, hair, sperm, egg-cells, embryos, and tissue samples, for biodiversity conservation, research and biotechnology development purposes in South Africa.

The objectives of BioBankSA include: ensuring that a representative collection of samples from key wildlife species are banked and effectively managed and utilised to maximize the benefit to society; facilitating biomaterial value enhancement through collaborative research and development both locally and internationally; functioning as an integrated resource facility linking partner collections; providing a comprehensive biomaterials information system serving a range of research and management groups spanning multiple disciplines locally, regionally and globally; establishing a single port of entry from which African biomaterials may be accessed for research, biodiversity conservation and biotechnology development purposes; developing and coordinating processes relating to Access and Benefit Sharing (ABS), Intellectual Property Rights (IPR) and Indigenous Knowledge that relate specifically to the exploitation of wildlife and indigenous biomaterials, to ensure the benefits emanating from the distribution and use South African wildlife genetic resources are equally shared; and to advance appropriate education and training so as to secure a sustainable biomaterial resource facility that enables stakeholders to make informed choices while conserving and maintaining genetic resources for future generations.

The completion of the BISI project is integral to the achievement of many, if not all, of the objectives of BioBankSA. More specifically, the initiative aims to establish an efficient, integrative sample tracking management system built on top of the innovative DGB data storage architecture developed by the NBN. The choice to use the DGB data storage structure will ensure that when the generation of high-throughput molecular data from wildlife samples becomes feasible, the necessary storage, integration and subsequent mining facilities can be readily employed. This system will create a research-oriented, minable database of wildlife and indigenous livestock biomaterial information enabling the identification of specific research priorities. BISI expects to extend the network to incorporate other such biological resource centres in the Southern African region thereby facilitating the integration and effective management of wildlife and indigenous livestock biomaterials and ensuring maximum benefit is captured for the people of South Africa.

It is hoped that BISI will form the core component of the planned integration system for biodiversity informatics by linking to other biodiversity information sinks, specifically those concerning transmission of disease between wildlife and livestock populations in the region, thus enabling the prediction of Long Term Environmental impacts and epidemiological risk management with a far greater efficacy than is currently possible.

Finally, the ultimate end goal is that BISI will facilitate and expedite BioBankSA's mandate to efficiently store and manage current wildlife and indigenous livestock genetic variation as an insurance policy against future biodiversity loss.