

# The Kruger National Park Data Repository

## Judith Kruger

Programme Integrator: Science Support

E-mail: [judithk@sanparks.org](mailto:judithk@sanparks.org)

---

The Kruger National Park (KNP) data repository is an innovative and user-friendly way for people to access the vast sets of data that have accumulated from the many years of research that has been done in the Kruger National Park. The data repository is still in the development phase and as new datasets are added, the system will be upgraded to suit user and supplier needs.

Scientific Services has been involved with research on various topics in the Kruger Park, since the 1950's. With each completed research project has come the data collected in the field or generated in the lab, on subjects ranging from geology to tree population dynamics. This collection of data, particularly the long term census datasets for elephants and plains game, are infinitely valuable for researchers and managers. Making this information available to people was the driving motivation in developing the KNP data repository.

The compilation of this useful data access system has been developed over the last 5 years with the accent being on the ease of access to various data sets. The data repository has been structured to work on a uni-

versal system such as Yahoo! Or Google search engines, so a user can simply type in the keyword of the dataset they are looking for and, if it is available, they will find it.

The system that was developed especially for ecological data by NCEAS (National Centre for Ecological Analysis and Synthesis) in Santa Barbara, USA, was selected as the platform to use for the KNP data repository. The scientists at NCEAS have considerable experience in managing, developing and storing long term data sets in the USA and this expertise was essential in constructing a similar access system for the Kruger Park. The NCEAS staff have been working with Scientific Services to fine tune the system, over the last three years, so that is can accommodate all the SANParks and Kruger data requirements.

This system has been adopted by the Long Term Ecological Research Network in the USA and other Ecological bodies. Members of the International Long Term Ecological Research network (ILTER), like China and Taiwan, have also started using this system. This means is that a person will be able to do inter-continental



Photo: Richard Reynolds

searches by searching only one of these data catalogues and obtaining all the data from the different continents. This system is going to be expanded in South Africa to also include data from the SAEON (South African Environmental Observatory Network) sites.

The data repository has been constructed to protect sensitive data by using a login/ password system. This means that data can be restricted to a few users who are working with that particular dataset. Data that has not been published will be lead-time protected with only the metadata made available until the data is clear for public circulation. This data access system has helped to reduce the workload of staff who are being asked for data, as the requestors can now go directly to the website and obtain the data and the metadata themselves.

The advantages of this repository does not stop with data use and sharing but will be broadened to automation of analyses to determine whether certain ecological thresholds have been exceeded. A scientific workflow

program called Kepler will access the data directly from this data repository, perform the necessary analyses and provide output to managers via a web based system, indicating whether the ecological threshold of interest has been exceeded or not.

The data repository can be accessed at <http://dataknp.sanparks.org>. The data is freely available and can be downloaded.

The data is accompanied by metadata, which is the information that describes the data and is needed to be able to interpret and use the data. This provides the user with more background information on how, when and where the data was collected. This includes:

- Abstract for the dataset. This describes the project that collected the data.
- Geographic coverage. Area of the study e.g: Entire KNP or where transects were laid, with the beginning and end point GPS co-ordinates. If points are used then a GPS point for each will be given.

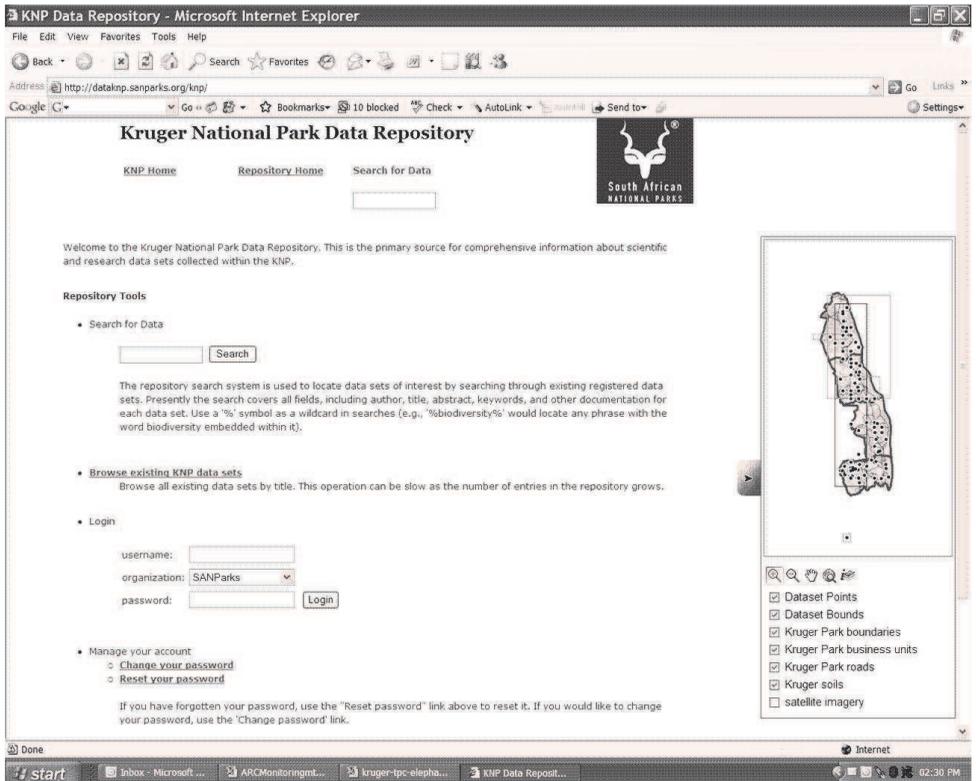


Figure1: Example to webpage for the KNP Data repository. Users will be able to search for data or browse through existing datasets. The interactive KNP map will also allow users to search for data linked to specific areas in the park.

- Temporal coverage. The dates that the data was collected.
- Keywords. words that characterize the database
- Taxonomic coverage of the dataset. Information on the genus and species name of the individuals that were sampled in the dataset. This is often provided in a table format.
- Data Usage rights. A paragraph that describes the intended usage rights of the data. Specifically including any restrictions (scientific, technical, ethical) to sharing data within the public scientific domain.
- Access control. Restricted datasets that have certain people that are able to access this data.
- Methods. The methods that were used in the study are listed here.
- People and organizations. Contact details of people associated

with the dataset and also the role that they played e.g: meta-data provider, principal investigator

Figure 1 is what users can expect to find once they log on to the webpage.

Figure 2 provides an example of the webpage following a search for a specific data set in the data repository. Once the list has been generated of available datasets, these can be opened and the page will display

as in figure 2.

The KNP data repository will continue to provide data to the people who need it. Managing and archiving data correctly as it accumulates is essential to maintaining the integrity and validity of the data. Making it easily accessible assists researchers in streamlining their work, while also providing scope for the development of comparative work and related datasets and will also avoid potential duplication of research efforts.

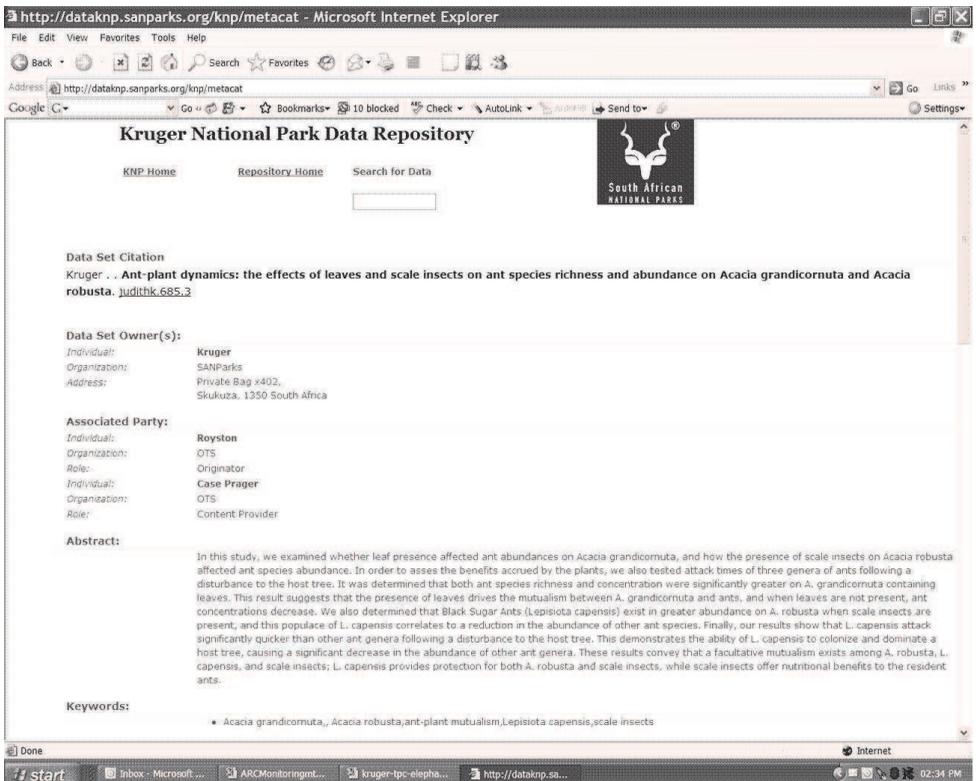


Figure 2: An example of a data set citation following a search of the data repository, showing the associated metadata details including the abstract.

