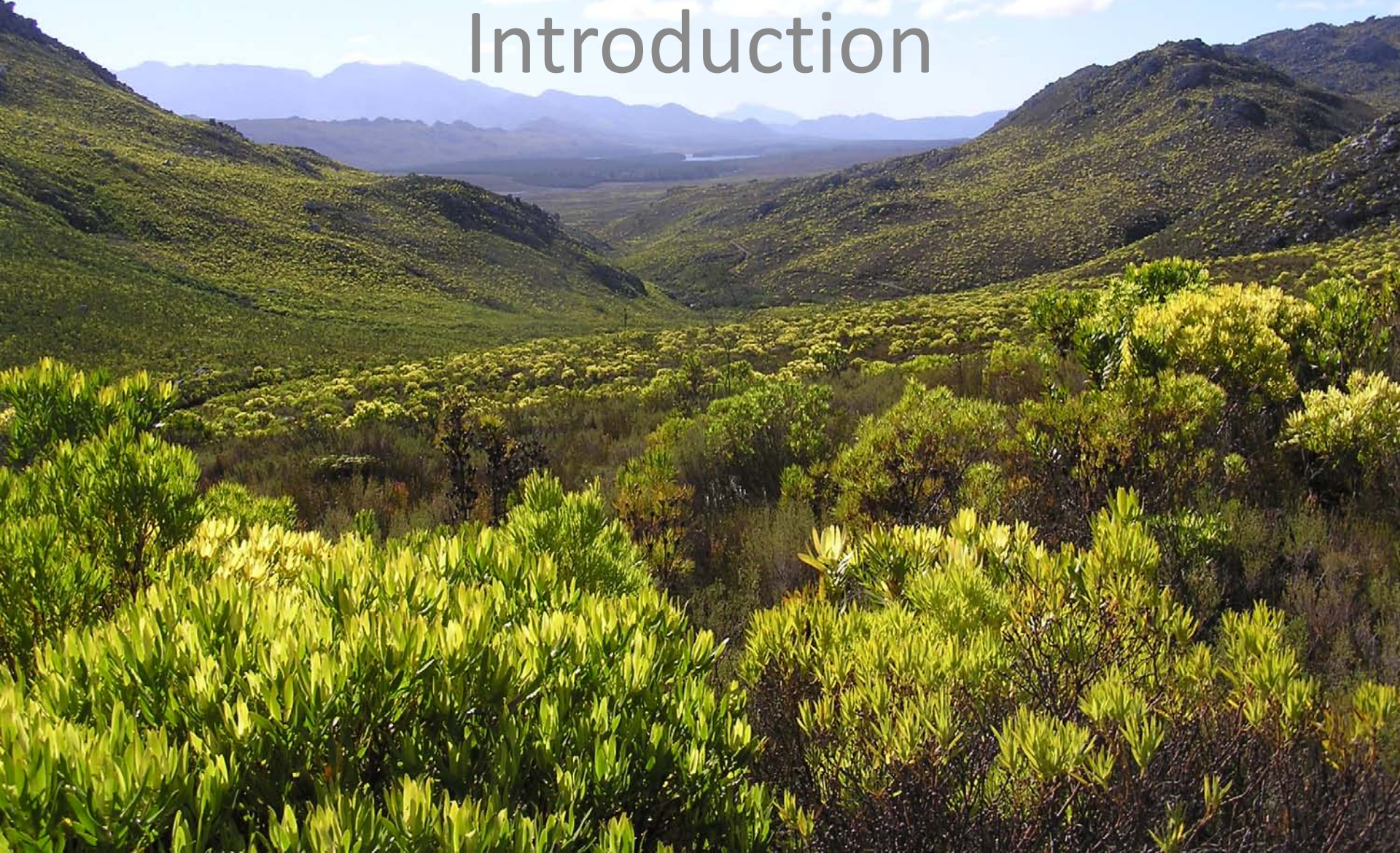


DATA MANAGEMENT

Introduction



DATA MANAGEMENT

A key focus area for SAEON

Victoria Goodall

Data Manager – SAEON Fynbos Node

GSN Indibano

March 2010 – Port Elizabeth



SAEON

**South African Environmental
Observation Network**

SAEON's VISION

A comprehensive, sustained, coordinated and responsive South African Earth observation network that delivers long-term reliable data for scientific research and informs decision-making for a knowledge society and improved quality of life.



Introduction

- Data Management is described as comprising all the disciplines related to managing data as a *valuable* resource
- This is a very broad statement – what does it mean in reality?
- It can include the development and execution of policies, practices and procedures that manage the full “data life-cycle”



Data Governance

- This broad topic covers aspects certain aspects of data management within an organization such as:
 - Data Quality
 - Business Processes
 - Risk Management



Data Architecture, Design & Analysis

- Data storage
- Data inspection
- Data cleaning
- Data transformation
- Data modeling
- Data mining – using data to uncover knowledge rather than using data for descriptive purposes only



Database Management & Security

- Adding, deleting, updating data
- Database administration
- Database Management System – systems / software / end users
- Data access
- Data privacy
- Data security
- Legal implications – numerous Acts dealing with data availability



Data Quality

- Data cleaning – do not underestimate the time that this process will take
- Data integrity
- Data quality assurance
- This is an incredible important step in the data management process – if you don't get this right, the rest will not work out!



Data Warehousing

- Data integration – for example merging data from different bioinformatics' repositories
- A secure environment to store data from multiple sources
- A lot of data has already been collected – there is often no need to collect more data at vast expense. Use what is already out there...
- Differentiate from the original raw data, and datasets created through analytical processes



Document & Content Management

- A system for tracking electronic documents such as reference papers, images etc.
- Efficient storage, naming and summaries of documents will allow you to easily access documents which are relevant for you at that time
- This can be another *valuable* resource for you



Metadata Management

- Creating metadata
- Metadata discovery
- Metadata publishing
- Metadata registry

– *We will discuss this in detail later today*



Why is Data Management Important?

- Conducting research without a solid data management process is like trying to build a house without foundations
- Good Data Management facilitates good Project Management
- It forces you to understand and document what is in your data, meaning that less time is wasted when you revisit data and don't remember what you did or how it was collected



So What???

*All talk and
nothing
concrete??*

