BIG DATA AND MINERAL EXPLORATION, HOW TO BE AN ORGANISED DATA HOARDER

Suzanne Bergman

Kenex
Creating opportunities in the spatial world

INTRODUCTION

Explosion of new data sources:

- local and national government,
- research institutes
- and individual explorers.

Geoscientists have to use an **overload** of raw data to **drive progress, innovation and profit.**

Data is provided in **various formats**, with **uniquely attributed features** and at **different resolutions**

BUILDING A DATA EMPIRE

All investments are risks

Valuable asset

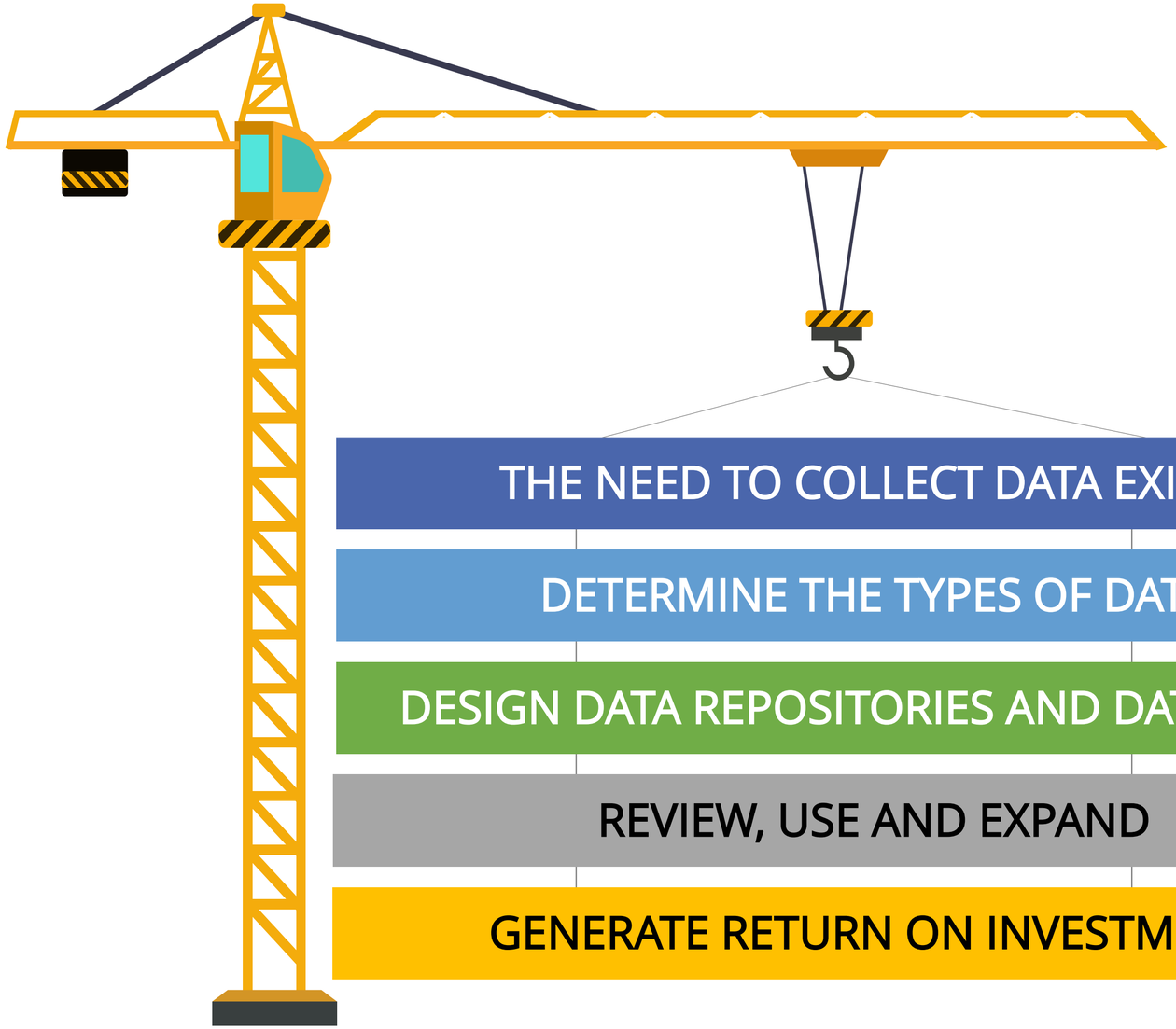
OR

Costly, unstructured hoarding of useless data

Minimise risk?

- Databases and data repositories
- Collect only what you need

DEVELOPING A DATA WAREHOUSE



THE NEED TO COLLECT DATA EXISTS

DETERMINE THE TYPES OF DATA

DESIGN DATA REPOSITORIES AND DATABASES

STAGE 1

REVIEW, USE AND EXPAND

STAGE 2

GENERATE RETURN ON INVESTMENT

STAGE 3

DATA WAREHOUSE DESIGN

Extract

User-friendly sharing methods and platforms

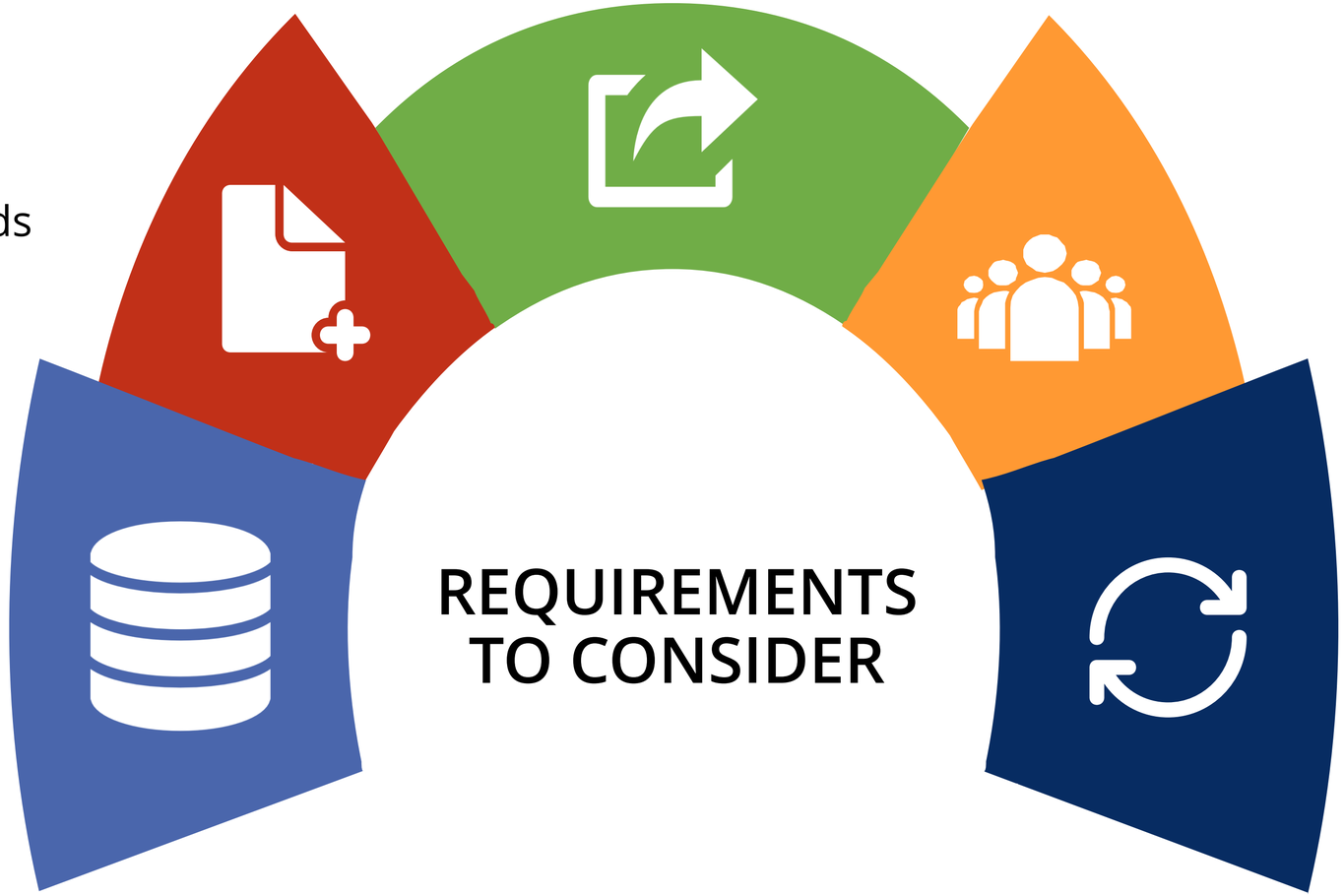
Access

Security

Collect

Data formats and relationships

Metadata standards



Use

User-friendly sharing methods and platforms

Access

Security

Store (Critical!)

Software and hardware

Access and security policies

Storage requirements

Ensure there is room for growth

REQUIREMENTS TO CONSIDER

Update

User-friendly sharing methods and platforms

Access

Security

CHOOSING THE RIGHT SOFTWARE

AVOID

- Expensive software that is not suitable for your needs
- Takes a lot of time to implement
- Presents a steep learning curve for staff

CHOOSE FIT FOR PURPOSE SOFTWARE

DON’T BE SEDUCED BY ALL THE IMPRESSIVE SOFTWARE AVAILABLE

YOU’RE NOT GOOGLE, YET!

CONSIDER THE FOLLOWING

- I can store the type and size data I need
- It is compatible with existing hardware, software and network capabilities
- It has efficient disaster recovery routines
- Leaves room to grow without compromising efficiency

TYPES OF DATA

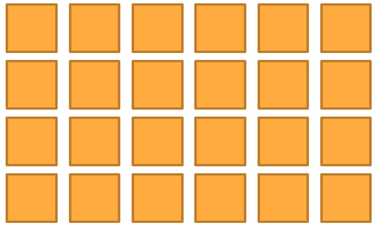
STRUCTURED DATA

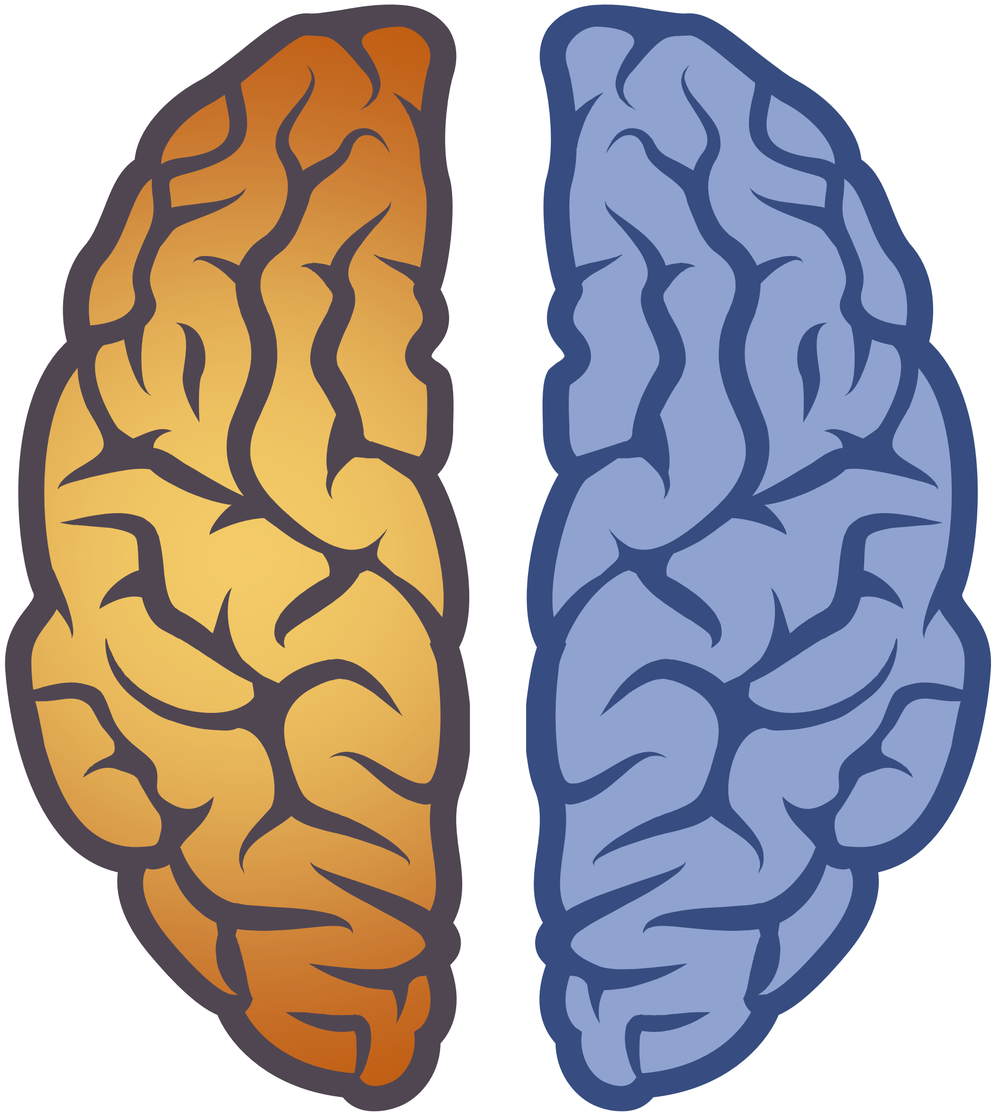
Highly organised

Stored in fields and rows

Simple to search, enter, store, query and analyse in a relational database

Example:
Spatial vector data (point, line or polygon) and a related table of attributes i.e. drill holes and drill logs, geochemistry and assays





UNSTRUCTURED DATA

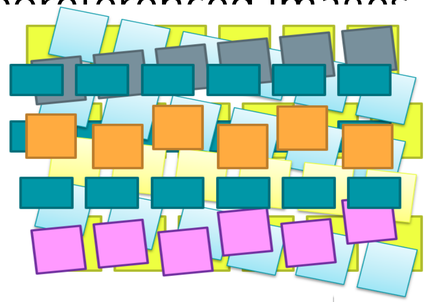
Unorganised internal structure doesn't fit into relational databases

Analysing can be laborious and time-consuming

More likely to get lost or corrupted left unmanaged.

Examples:
Non-spatial: documents, images, pdfs, e-mails and account files

Spatial data: Geophysical grids, aerial photographs, geology maps or georeferenced images



RELATIONAL DATABASES

STAR

Stores attributes in one ‘flat’ table e.g. a single shapefile of points and their attributes in the attribute table

Main issue:

Space they occupy
Uncontrolled growth
Empty fields
Slow to query and analyse

SNOWFLAKE

Normalises and reduces the dimensions of datasets

Moves attributes into separate tables that relate to a main table by using a foreign key

Recommended for

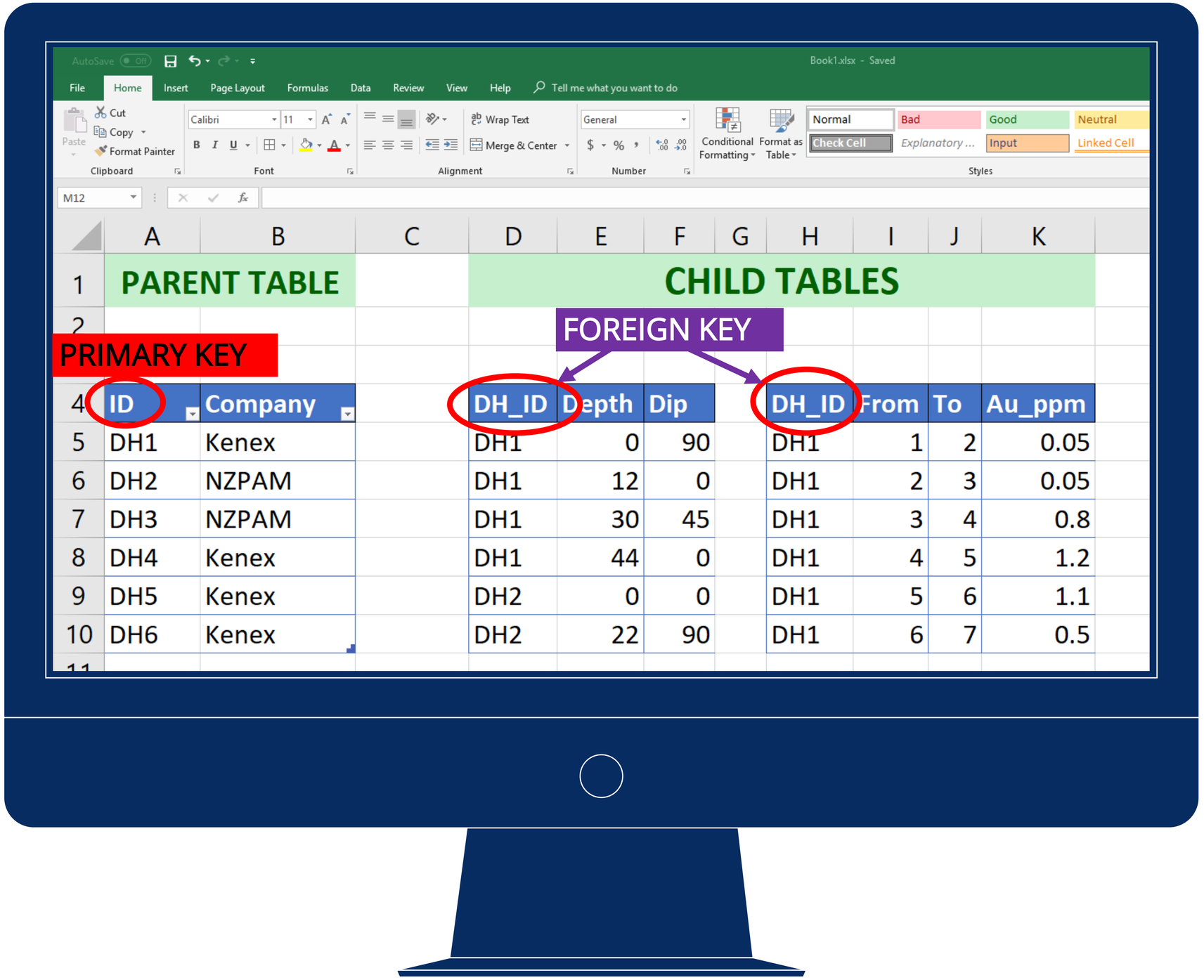
Tables with sparsely populated attributes, e.g drill hole tables and their related assay values, survey values and lithology logs.

SNOWFLAKE RELATIONAL DATABASE

01 One to One

02 One to Many

03 Many to Many

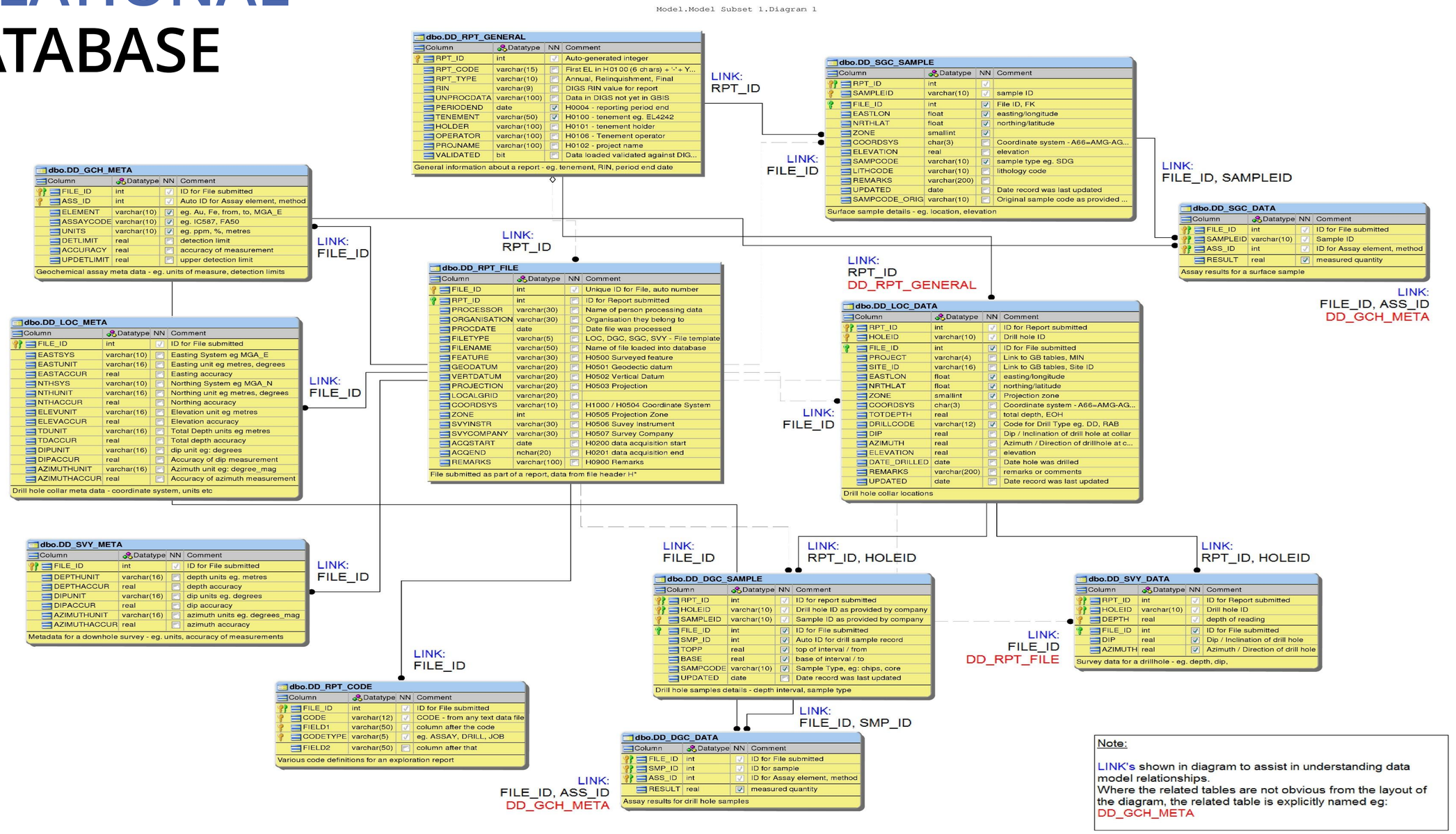


ID	Company
DH1	Kenex
DH2	NZPAM
DH3	NZPAM
DH4	Kenex
DH5	Kenex
DH6	Kenex

DH_ID	Depth	Dip
DH1	0	90
DH1	12	0
DH1	30	45
DH1	44	0
DH2	0	0
DH2	22	90

DH_ID	From	To	Au_ppm
DH1	1	2	0.05
DH1	2	3	0.05
DH1	3	4	0.8
DH1	4	5	1.2
DH1	5	6	1.1
DH1	6	7	0.5

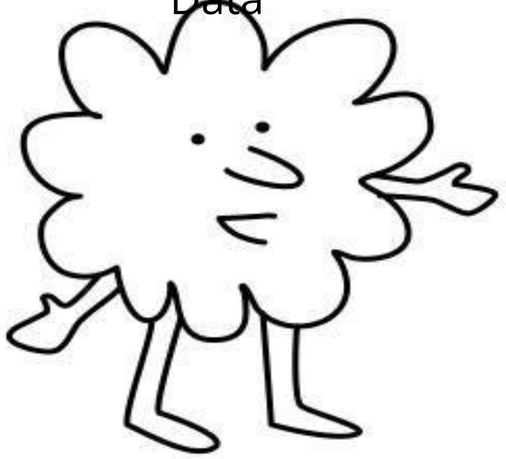
RELATIONAL DATABASE



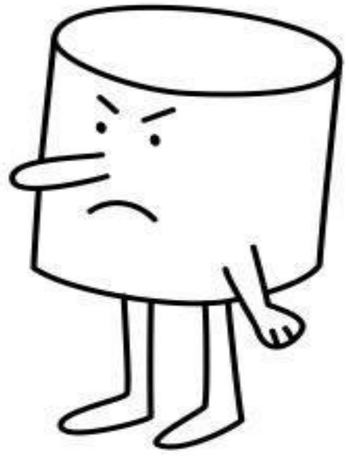
An example of a relational database from the New South Wales Digital Imaging of Geological System (DIGS) database.

DATA REPOSITORY FOR UNSTRUCTURED DATA

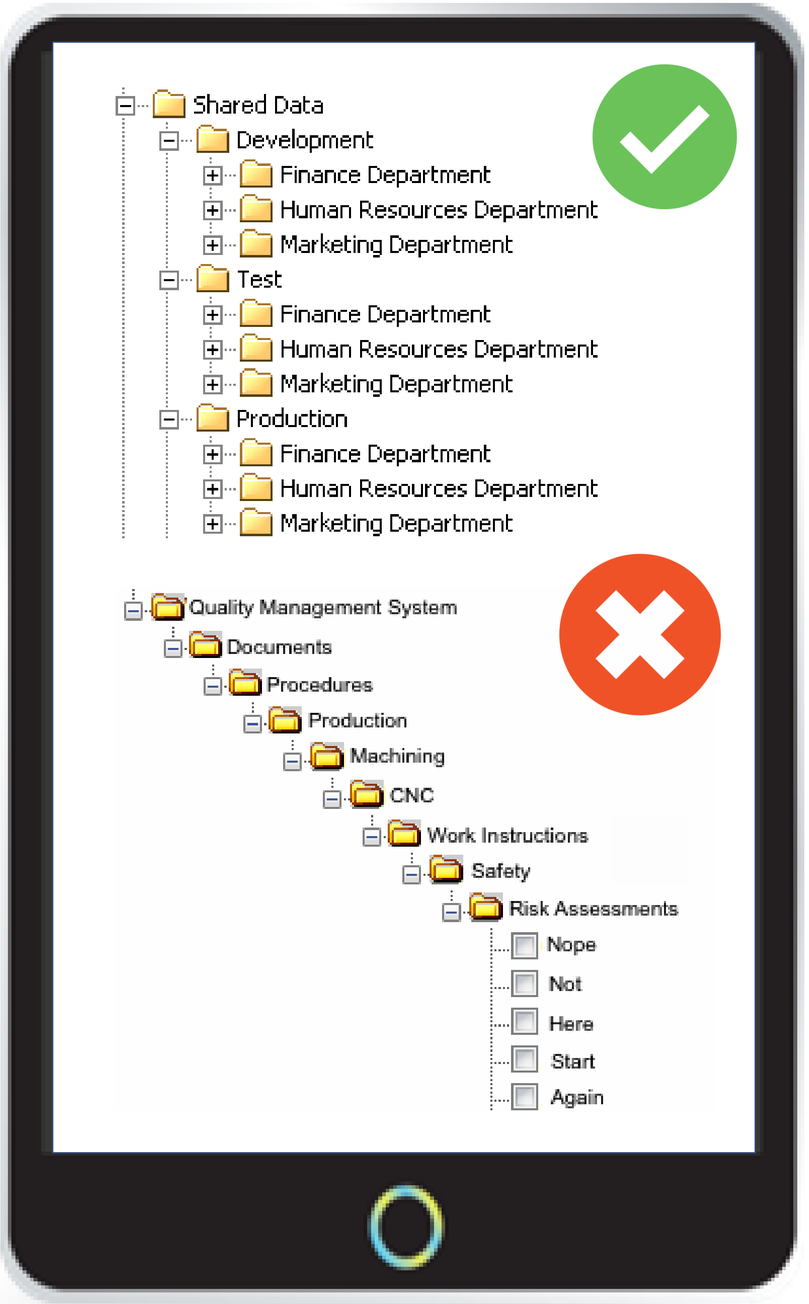
Unstructured Data



Relational Database



"I'm sorry, I'm just not that into you..."



- Sensible naming conventions
- Folder tree structures, simple nesting
- Don't store duplicate files
- Regular backups

WHAT'S NEXT ?

01

REVIEW

Test and ensure data is retrieved properly:

Add, delete and search for data

Append, select and update queries

Verified in different software packages, i.e. mining and GIS software

Document the design

02

USE

Start using your data for projects

Train users

The users' feedback will lead to additional changes to the design as it grows

03

EXPAND

Start growing your data empire

QAQC datasets to ensure quality and standardized before adding them

Discard poor quality data (e.g. hard to handle formats or not useful to projects)

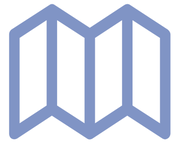
Ongoing performance monitoring

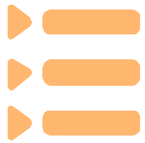
Refresher training

QA/QC

Incorrect unit conversions

Spatial projection issues



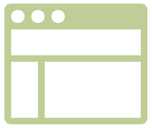


Attributes assigned to the wrong column

Misspelled words

Drag and drop errors

Violating field types (i.e. text in numeric fields)



METADATA

CONCLUSION

Big initial investment of time and resources

Plan, test and review data warehouse design

Always follow standards set up in the data warehouse design

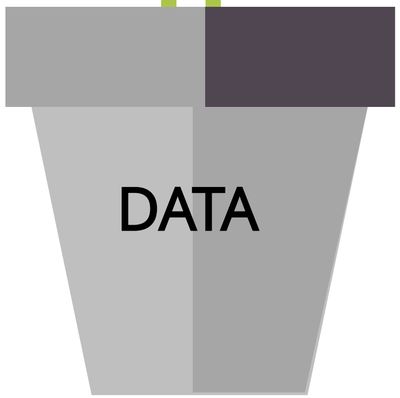
Only collect good quality, relevant data

Harness the inherent energy of your data

Initial design is only a modest version of the eventual product

Use your data as the energy that fuels informed decisions

An efficient system that will increase the value of the data and overall productivity



THANK YOU

Want to get the most out of your data?
Let's talk...

Kenex

Creating opportunities in the spatial world