

# Relational Databases

## Part 1: Introduction

# Data Models

- Flat Files, e.g. Excel, Word – mature and widely used
- Relational Databases – mature and widely used
- Other: Hierarchical, Network, Object-Oriented, Object-Relational
- Statistical, graphing, GIS programs

# Flat files (e.g. spreadsheet)

Collection Table

<b>Genus</b>	<b>species</b>	<b>Seamount</b>	<b>Date</b>
Beryx	splendens	Great Meteor	12/22/2005
Lucifer	typus	Jumeau West	3/15/1972
Madrepora	oculata	Hjort	5/12/1995

Seamount Table

<b>Seamount</b>	<b>Depth</b>	<b>Latitude</b>	<b>Longitude</b>
Great Meteor	140	30	28.5
Jumeau West	255	23.6	168
Hjort	1500	59.31	158.48

# Formats

- Proprietary: Excel, Statistix
- Text: CSV (comma delimited), tab-delimited
- Binary non-proprietary: NetCDF, HDF, SB2
  - Q: what is binary?
  - Q: what do you do when you get a binary file?

# Spreadsheet pros and cons

## – Good

- Widely used, easy to learn for simple graphical and statistical analyses
- Commonly already installed on most computers

## – Bad

- Can encourage “bad practices” – create data that can’t easily be used
- Lack of auditability – hard to “back track” how data were manipulated
- Hard to interrelated data from different files (i.e. conduct sophisticated analyses) - lose information on the structure of data

# Relational Databases - another option

- A relational database is essentially a series of tables with relationships defined between the tables based on shared elements
- Generally has a query language built in

# Relational Database

Collection Table

Genus	species	Seamount	Date
Beryx	splendens	Great Meteor	12/22/2005
Madrepora	arancione	Jumeau West	3/15/1972
Madrepora	oculata	Hjort	5/12/1995

Seamount Table

Seamount	Depth	Latitude	Longitude
Great Meteor	2256	30	28.5
Jumeau West	125	23.6	168
Hjort	15	59.31	158.48

*“Show me all the seamounts within scuba-diver depth that have Madrepora corals”*

# Database Management Systems

- Commercial Products
  - Microsoft ACCESS (part of Microsoft Office)
  - Microsoft SQLserver
  - Oracle
- Freeware
  - MySQL
  - PostgreSQL
  - MiniSQL
  - OpenOffice Base

# Advantages and Disadvantages of using a DBMS

## Advantages

- additional capabilities
  - sorting
  - query
  - integrity checking
- more error control
- more efficient data entry (sometimes)
- easier to maintain (sometimes)

## Disadvantages

- few graphical or statistical capabilities
- proprietary formats may limit archival quality of data
- require different expertise

# Why Learn Relational Databases?

- Every resource management agency uses them: NMFS, EPA, UN...
- Every major conservation organization uses them: WWF, CI, IUCN red list, WCMC (World Conservation Monitoring Centre)...

# Tour of a Relational Database

Using the NaGISA data

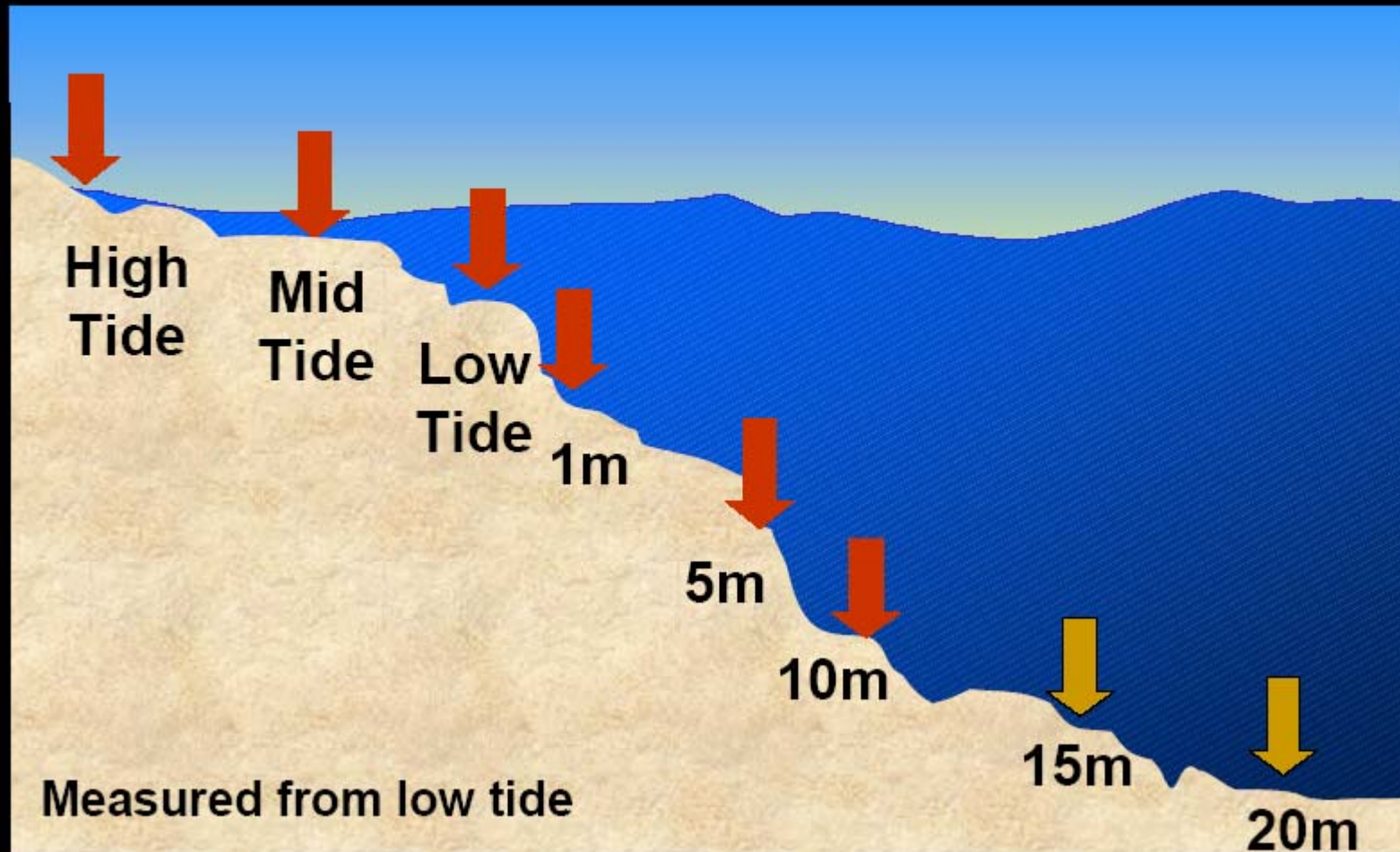
# NaGISA

using standard methods to compare nearshore  
biodiversity on large scales

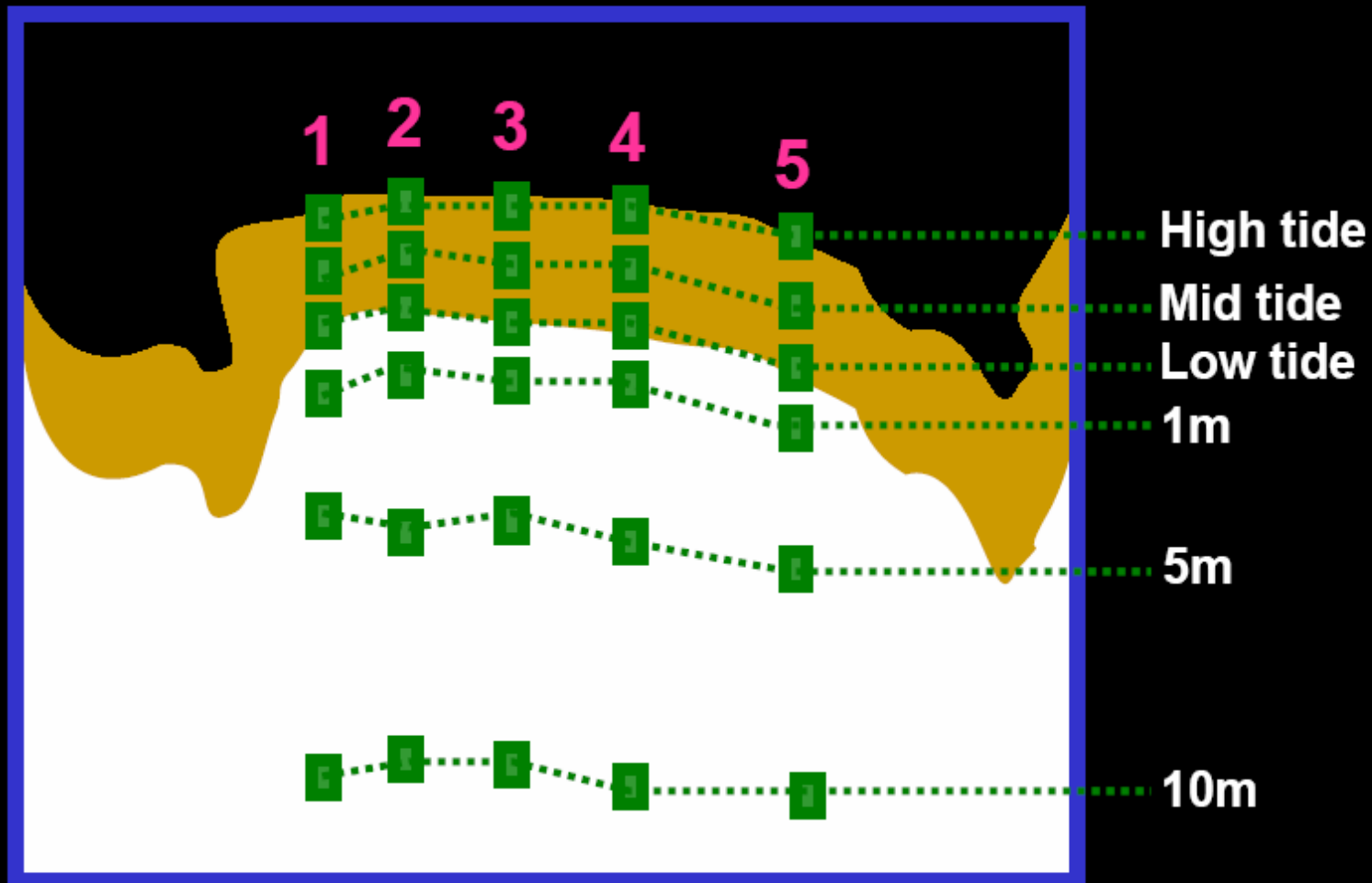
(<http://www.nagisa.coml.org/>)



There are **6 mandatory** and **2 optional** heights



**5 replicates** of each sampling height are made



# Macroalgae Quadrature Set Instructions



Digitally photograph each quadrature

**Collect all\* macroalgae**



**Collect everything**

# (Demonstration of Nagisa Database)

# Access elements

- Tables: to store data
- Queries: to ask questions from database
- Forms: to create a user interface
- Macros: to automate repetitive tasks
- Modules: create non-standard functionality, using Visual Basic

# Tables

- Must be defined before they can be used
- Note data editing, sorting, adding a new element, scrolling around.
- Vaguely resemble Excel sheets but:
  - Unit of information is a ‘record’ (line)
    - Saved automatically as soon as one moves to another record
  - Consists of ‘fields’
    - All fields of a record remain together
  - No restriction on the number of lines
    - Excel: 64,000 lines
  - Order of records is not fixed!

# Queries

- Select and sort records of a table
- Pull together information from related tables
- Simple statistics
  
- In principle: SQL
  - But access interface hides SQL from the user
- In practice: query grid
  - Define which tables will be used, fields that have to be shown, criteria for selection and sorting...