Introduction to managing spatial data

Diego Juffe Bignoli, Development Corridors Partnership 3 June 2019
Outline

1. Plan for data collection
2. Manage your data
3. Process your data – done this!
4. Share your data
1. Plan for data collection
The data management plan

- Explains how DCP will deal with data to comply with donor and legal requirements
- Describes the data workflow to manage DCP data
- Identifies types of data DCP will produce and defines data standards
- Provides guidance on how to manage different types of data
DCP approach to data management

Design

Process

Store

Share
Design

• Start by designing the data collection protocols: 2-3 page data management plan
• Make sure you comply with your project and funder privacy, ethical and quality control requirements
Process

Steps to make sure data collected is:
- Consistent: data is collected in the same way, properly labelled and documented, quality checked (clean) and not corrupt (it works!)
- Interoperable: data is compatible with your softwares and sharing platforms
• Ensure you store your data in a consistent, systematic manner (more later!)
• In DCP, data has to comply with the project data requirements and UNEP-WCMC securely stores all data relevant to the project. How about you?
Avoid shelves! How is data going to be disseminated internally and externally?
This is a fundamental part of disseminating results and influencing decision makers to make a change.
Make sure you know what data you need
What data do you need for your research?

- Before you start collecting data define:
  - Research questions
  - Methods and tools you are going to follow to respond to those questions
  - Data you need to implement your methods
  - Expected outputs from your research
  - If relevant: target audience and expected impact – what do you want to achieve?
Activity: what data do you need (research plan excel file)?

In groups and for 30 min use the excel file to think about and fill in (2 volunteers will report back):

- **Tab R_questions**: Your research questions, the methods you will use, the data you will need, and the outputs you expect
- **Tab data gaps**:
  - **Category of data**: species, ecosystems, water, employment, people’s views
  - **Format**: spatial data (rasters or vectors), tabular data, qualitative, video, audio, reports, pictures, paper maps, etc.
  - **Resolution**: 1 km², 20 m, not applicable
  - **Sources**: collect from the field, from existing databases, processed, etc.
2. Manage your data
3 Principles for managing spatial data

1. Identify the software and formats that are fit for your purpose: ArcGIS, QGIS and associated software, PostGIS, etc.

2. Be systematic and consistent on how you name and store your data: data standards

3. Define and stick to your workflow: the steps you will follow to collect, store and manage your data
The world database on protected areas (WDPA): 220,000 records collected from 500 sources and 193 countries

https://www.protectedplanet.net
Discover the world's protected areas
How is the WDPA managed?

1. Data standards
2. Clear data collection workflow
3. Metadata standards
## Data standards

Defines the database structure, accepted values and types of data in it.

### Table

<table>
<thead>
<tr>
<th>No</th>
<th>Requirement</th>
<th>Provided by</th>
<th>Field Name</th>
<th>Type</th>
<th>Length</th>
<th>Accepted values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minimum</td>
<td>UNEP-WCMC</td>
<td>WDPAID</td>
<td>Number (Double)</td>
<td>N/A</td>
<td>Assigned by UNEP-WCMC. Unique identifier for a protected area.</td>
</tr>
<tr>
<td>2</td>
<td>Minimum</td>
<td>UNEP-WCMC</td>
<td>WDPA_PID</td>
<td>Text (String)</td>
<td>52</td>
<td>Assigned by UNEP-WCMC. Unique identifier for parcels or zones within a protected area.</td>
</tr>
<tr>
<td>3</td>
<td>Minimum</td>
<td>Data provider</td>
<td>PA_DEF</td>
<td>Text (String)</td>
<td>20</td>
<td>Allowed values: 1 (meets IUCN and/or CBD PA definition); 0 (does not meet IUCN and/or CBD PA definition; currently stored outside WDPA).</td>
</tr>
<tr>
<td>4</td>
<td>Minimum</td>
<td>Data provider</td>
<td>NAME</td>
<td>Text (String)</td>
<td>254</td>
<td>Name of the protected area (PA) as provided by the data provider.</td>
</tr>
<tr>
<td>5</td>
<td>Minimum</td>
<td>Data provider</td>
<td>ORIG_NAME</td>
<td>Text (String)</td>
<td>254</td>
<td>Name of the protected area in original language.</td>
</tr>
<tr>
<td>6</td>
<td>Minimum</td>
<td>Data provider</td>
<td>DESIGN</td>
<td>Text (String)</td>
<td>254</td>
<td>Name of designation.</td>
</tr>
<tr>
<td>7</td>
<td>Complete</td>
<td>Data provider</td>
<td>DESIGN_ENG</td>
<td>Text (String)</td>
<td>254</td>
<td>Designation in English. Allowed values for international-level designations: Ramsar Site, Wetland of International Importance; UNESCO-MAB Biosphere Reserve; World Heritage Site. Allowed values for regional-level designations: Baltic Sea Protected Area (HELCOM); Specially Protected Area (Cartagena Convention); Marine Protected Area (CCAMLR); Marine Protected Area (OSPAR); Site of Community Importance (Habitats Directive); Special Protection Area (Birds Directive); Specially Protected Areas of Mediterranean importance (Barcelona Convention). No fixed values for protected areas designated at a national level.</td>
</tr>
<tr>
<td>8</td>
<td>Minimum</td>
<td>Data provider</td>
<td>IUCN_TYPE</td>
<td>Text (String)</td>
<td>20</td>
<td>Allowed values: National, Regional, International, Not Applicable</td>
</tr>
<tr>
<td>9</td>
<td>Complete</td>
<td>Data provider</td>
<td>IUCN_CAT</td>
<td>Text (String)</td>
<td>20</td>
<td>Allowed values: Ia, Ib, II, III, IV, V, VI, Not Applicable, Not Assigned, Not Reported</td>
</tr>
<tr>
<td>10</td>
<td>Minimum</td>
<td>UNEP-WCMC</td>
<td>INT_CRIT</td>
<td>Text (String)</td>
<td>100</td>
<td>Assigned by UNEP-WCMC. For World Heritage and Ramsar sites only.</td>
</tr>
<tr>
<td>11</td>
<td>Minimum</td>
<td>Data provider</td>
<td>MARINE</td>
<td>Text (String)</td>
<td>20</td>
<td>Allowed values: 0 (100% Terrestrial PA), 1 (Coastal: marine and terrestrial PA), and 2 (100% marine PA).</td>
</tr>
<tr>
<td>12</td>
<td>Minimum</td>
<td>Data provider</td>
<td>REP_M_AREA</td>
<td>Number (Double)</td>
<td>N/A</td>
<td>Marine area in square kilometers.</td>
</tr>
<tr>
<td>13</td>
<td>Minimum</td>
<td>UNEP-WCMC</td>
<td>GIS_M_AREA</td>
<td>Number (Double)</td>
<td>N/A</td>
<td>Assigned by UNEP-WCMC.</td>
</tr>
</tbody>
</table>
Explains the steps through which the data is collected, processed and stored.
Metadata is key for publishing any database. It provides information about the data:

- How it is called
- how it has to be cited
- who has collected it and how
- Which is the coordinate system and projection
- how can you use it and how has the copyright

### Description:

The World Database on Protected Areas (WDPA) is a joint product of UN Environment and IUCN, managed by UNEP-WCMC and the IUCN working with governments, communities and collaborating partners.

The dataset described here shows the global distribution of terrestrial and marine protected areas.

The dataset contains protected areas designated at the national level and under regional and international conventions and agreements. International designations include those under the Ramsar Convention, the World Heritage Convention (United Nations Educational, Scientific and Cultural Organization, UNESCO), and sites under the UNESCO’s Man and the Biosphere Programme (MAB). Regional agreements include sites under the Nature 2000 network (European), as well as Marine Protected Areas designated under regional conventions such as the Convention for the Protection of the marine Environment of the North-East Atlantic (OSPAR) and many others. It also contains data on protected areas established by other means.

### Citation(s):

UNEP-WCMC and IUCN (year). Protected Planet: [Insert name of component database]. The World Database on Protected Areas (WDPA)/The Global Database on Protected Areas Management Effectiveness (GD-PAME) [Online]. [Insert monthly/year of the version downloaded]. Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

Other Cited References


Working with geopackages and QGIS database manager

Follow the PDF Geopackages_tutorial_2019
Issues with shapefiles

- Shapefiles cannot hold more than 2 GB of data
- It is not an open source format
- They can easily get corrupted (stop working)
- Only contain one geometry type per shapefile
- They are formed of 7 individual files
- Attribute names limited to 10 characters
What are GeoPackages?

An Open Format for Geospatial Information
GeoPackage is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information.

www.geopackage.org
Benefits of GeoPackages (GP)

1. Open source
2. Compatible with all GIS software packages
3. More stable and efficient in geoprocessing
4. No space issues (limit of 140 TB for each GP)
5. Light and easy to share as it stores all data in one file
How do they look like?

With a geopackage you can manage your data more efficiently and find your files quickly.

You can store in the same place as many tables, vectors and rasters as you like (limit 140 TB).
Let’s create a GP layer and GP database
1. Click on the GP Icon OR go to Layer/create layer and choose GP layer

2. A window will open, fill in the fields (see next steps)

3. Choose a location folder and name for your GP database. If you already have a GP database choose its location.

4. Choose name of the first file you will create in the GP database

5. Select type of data for that new file. If you don’t want to create any new spatial data choose (no geometry)

6. Select projection
7. Create fields for the new file. If you want no fields go to next step.

You will have to define name of the field, type of information that field will store and, for text fields, number of characters that will field will allow.

The fields will show in the field list.

**Important:** in this step you are not creating any geographic features. You are just defining the structure of the dataset. The dataset you will create will be empty.

8. Click OK
You have created your first GP layer which is also a GP database where you can add other data you have.
Let’s add existing files to the GP database you just created
1. Choose one spatial file you would like to include into a GP

2. Add it to the GP database you created

Option 1 Using an existing file in the layers list: choose export/save as and fill in the table choosing GP database location. A new file will show in your layers list

Option 2 – Using the QGIS Browser: choose the data in the QGIS browser and drag it into your GP database location in the browser

This step does not move the file it creates a new one so don’t forget to delete the old one!
Database manager: your GIS database

The database manager allows you to manage all your GIS databases in your computer but also:

- Import new or existing data into or from one database to another
- Export data to other places
- Use SQL query language
Database manager: The interface

- Space to manage your databases. Organised by type of database.
- Space for visualising details about each feature in your databases.
**Database manager: What each icon does**

- **Refresh DB manager**: Click here to start the DB manager.
- **SQL query tool**: Import: Use this to exchange data between places.
- **Import**: Use this to exchange data between places.
- **Export**: Export a file from the DB manager to another location.
- **Export**: Use these 3 tabs to explore each file. In this case, we are exploring PAs clip polygon.
- **Info**: Tell you about the parameters of the file.
- **Table**: Shows you the attribute table for that dataset.
- **Preview**: Shows a preview of the spatial data.
How to link an existing GP database you the database manager?
1. Click here to start the DB manager

2. Click here to link the GP you created to the DB manager
2. Choose your GP location

3. You can now see your GP is in the DB manager (if not press the refresh icon)
Now you can explore your GP database

Refresh DB manager

SQL query tool

- **Import**: Use this to exchange data between places
- **Export**: Export a file from the DB manager to another location
- Use these 3 tabs to explore each file. In this case, we are exploring PAs clip polygon
- **Info**: Tell you about the parameters of the file
- **Table**: Shows you the attribute table for that dataset
- **Preview**: Shows a preview of the spatial data
2. Process your data
Started with it in the morning. This is all about practice. Just remember...
4. Share your data