WWF for a living planet®

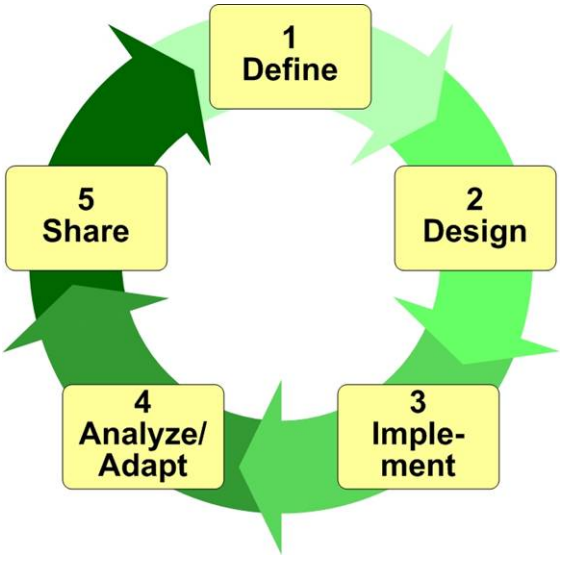
WWF Network Standards

Resources for Implementing the WWF Project & Programme Standards

Step 4.1

Manage Incoming Data

May 2006



Step 4.1 Manage Incoming Data

Contents

This document is intended as a guidance resource to support the implementation of the *WWF Standards of Conservation Project and Programme Management.* Although each step in these *Standards* must be completed, the level of detail depends on the circumstances of individual projects and programmes. Accordingly, each team will have to decide whether and to what level of detail they want to apply the guidance in this document.

This document may change over time; the most recent version can be accessed at: https://intranet.panda.org/documents/folder.cfm?uFolderID=60985. This document is based in part on materials developed by Foundations of Success.

Written by: Alexis Morgan, WWF-Canada **Edited by:** Foundations of Success

Please address any comments to Sheila O'Connor (soconnor@wwfint.org).

Manage Incoming Data

What Is Data Management?

As you design your project, and then especially once you have begun implementing your monitoring plan, your project team will begin generating many different kinds of data. As shown in the table below, these data can include everything from the initial project boundaries sketched out on a map to measurements of the population size of one of your biodiversity targets to subjective assessments of stakeholder buy-in to digital photos of project team members implementing a strategy.

Data management is a process to ensure that diverse data sets can be efficiently collected, integrated/processed, labelled/stored, and then easily retrieved through time by people who want to use them. In simple terms it could be taken to mean "a place for everything, and everything in its place."

Different Types of Data and Where They Come From and Are Stored

Type of Data	Examples of Sources	Examples of Databases
Quantitative – data that can be represented as numbers including both *continuous* data measured along a scale & *categorical* data recorded in intervals or by groups	Biological censuses or transects of species, counts of poaching incidents, household stakeholder opinions recorded on a 4-point scale, numbers of tourists visiting a site	Paper logbooks, simple spreadsheet tables *(Excel)*, relational databases on desktop computers *(Access)* or online servers
Qualitative – data that are not easily represented in numerical form	Stories from stakeholders or focal group interviews	Word processor documents, relational databases, folders of audio or visual clips
Spatial – data that are linked to specific geographic coordinates (typically quantitative, but could be qualitative)	Locations where animals have been poached recorded on a Global Positioning System (GPS) unit, boundaries of a national park	Paper maps in a file cabinet, Geographic Information (GIS) systems (*ArcGIS software products*)
Financial – a special form of quantitative data that contain financial information	Business records, project operations	Spreadsheet tables *(Excel)*, accounting software (*ACCPAC* or *QuickBooks* software*)*
Pictures & Images – photos, drawings, and other images	Before & after photos of a specific site, stakeholder drawings, conceptual models	Photo albums, slide files, computer file folders, album software
Video & Audio Clips – film, video, and audio materials	Recordings of stakeholder meetings, film clips of key project events	video library, computer file folders, archive software
Metadata – data about your other data; the documentation that accompanies any dataset	Lists of all your databases, descriptions about fields in a database, information about pictures in a photo album	Paper list, spreadsheet file

Why Is Data Management Important?

WWF project and programme teams typically invest considerable time and money into monitoring as part of their project work. As a result, they end up with lots of data from many different sources. Unfortunately, all too often these data remain in the project team's field notes or on survey forms. And even if these data are transcribed to a central location, these data can quickly become outdated or obsolete, be misinterpreted through poor or inadequate documentation, or can even, at times, go missing entirely, especially if files are not maintained and backed up on a regular basis. The quality of our conservation work is directly related to having reliable, credible, and current information upon which to base our work. Without a data management strategy, data can become lost or rendered useless, or just unable to provide its optimal usefulness (Higgins 2005).

Establishing and using a consistent data management framework will enable more effective use of the data by conservation projects and programmes. As a general rule, the more standardized the structure and content of the database, the more effectively it can be used by both humans and machines (FGDC 2006). Properly managed data:

- Enables the project team to explore the source, quality, and details of the underlying data behind decisions and therefore help to explain how and why results were achieved (or not achieved). In other words, it is central to performing adaptive management.
- Facilitates the engagement of stakeholders through easy understanding and sharing of data. When commonly understood information from a variety of sources is employed for decision-making, buy-in may, at times, be achieved more easily.
- Helps the team generate more comprehensive and attractive information products.

Establishing a consistent data management framework will also facilitate more efficient updates as the project evolves. In particular, it provides a clear rationale to project team members, donors, and partners on what data gaps exist and why these need to be addressed. It also helps project teams ingest standardised data from external sources and integrate them with internal data. It improves the transparency, accountability and learning of staff, partners and stakeholders by being able to retrace what data came from where. And finally, it is critical to ensure that data remain useable through time given the inevitable institutional memory loss and staff turnover.

When to Manage Data

Planning for data management should not be left until the moment when your project has already accumulated lots of data. Instead, you need to be doing this planning as you are designing your overall project actions and monitoring plan. For example, you should assign responsibility for data management as you are developing your project team in Step 1 and you should think about what types of data you will be generating as develop your monitoring plan in Step 2. Indeed, much of the data that you collect as you plan your overall project – for example information about your targets, threats, or key stakeholders – are data that you will need to manage even during this early phase. Likewise, data management itself needs to occur continuously over the life of your project. Data should be reviewed and transcribed as soon as possible after collection.

How to Manage Data

There is no one right way to manage the data for any given project. The specific tasks that you undertake will vary greatly depending on how large and complex your project is, how much data you expect to collect and use, and the technical capacity and resources of your project team. Some basic tasks to consider include:

1. Develop a table of the data sets you expect to have

The starting point is to develop a rough sense of what data sets you expect to have and how you will manage them. Much of this can be based on the information in your monitoring plan that you defined in Step 2.2. Key points to include in your table are:

- **What data the project will collect –** Develop a list of the different sets of data that you expect to collect over the life of your project. You should also note the form that each set of data will be in – for example text-based stories, GIS map layers, or digital photos. This can be done through a written document, a spreadsheet, or a database – you are in effect creating a database of your potential data sets.
- **Who will collect and manage the data** – For each data set, list who will be responsible for collecting the data, who will be responsible for managing them, and who will use them. This can be in text form, or if you want to be very explicit, you could even develop a flow chart showing the process of how each data set will flow through your project. In particular, you should designate a specific individual to be in charge of each data set. In many larger offices, the overall responsibility for data management or at least oversight of data management will typically be given to a part or full-time data manager (see next task).
- **How and where data will be stored –** Determine what type of database you will use for each type of data. For example, you may wish to put quantitative data in a computer-based database, spatial data in a GIS system, and digital photos in a photo archiving software program. You should also determine whether the master copies of these databases will be on a central office computer, on a web-accessible server, or some other location.

Example of Table of Data Sets for a Project

Data Set	Collectors & Managers	Storage	End Users
Results from marine transects: largely quantitative information about fish sightings, information about each survey	Collected by marine researchers on a monthly basis. Managed by project data manager.	Transcribed by researchers into Excel-based database at marine lab; should be backed up at project office	Used by project team members to measure health of coral reef target; this info conveyed to donors
Photos of damaged reef areas: photographs of coral reef areas linked to GPS coordinates	Collected by marine researchers once a year. Managed by project data manager.	Photos stored on marine lab computer in folder c://data/reefphotos. GPS coordinates and meta data in linked Excel spread sheet. Should be backed up at project office.	Used by project team members to measure health of coral reef target over time; this information conveyed to donors

- **Who will use the data** – Perhaps the most overlooked – and yet the most important – thing to consider in designing a data management is to think about who will ultimately be using the data. More than anything, your system should be designed so that the users can get the data they need when and where they need it. Otherwise, you are wasting your time.

2. Designate data managers

All programme and project teams can benefit from the support of a data manager, who as Higgens (2005) describes:

> …is able to collect information, perform complex analyses, produce quality maps, and administer all tabular and spatial information in an organized and efficient manner. The lead information manager/data manager would ideally be located in a field office and should be identified as early as possible [in the development of the project] to answer key information management questions and establish the data management structure.

Ideally data managers should be identified in Step 1.1 with the definition of the initial team and responsibilities. If this role has not yet been assigned to anyone, someone needs to be assigned responsibility. For many projects or programmes, lead responsibility for data management will probably be handled by staff with additional roles, but support in the set up of systems may need to be found from outside sources (e.g. consultants or other parts of the WWF Network).

3. Develop protocols and databases for different kinds of data

As noted above, most types of data are only really useful if they are collected and then stored in a standard fashion. For example, if you will be conducting transects to sample the number of certain species of fish on a coral reef over time, you will want to record standard information about each transect such as the number of each species of fish encountered, the time of day the transect was conducted, the degree of turbidity in the water, and who conducted the survey. Likewise, if you will be conducting a household survey to determine attitudes of local stakeholders about a national park, you will want to record standard information from each interview.

Each piece of data that you collect can then be defined and recorded in a standard fashion. You need to need to specify in advance how you will record each piece of data in a consistent and specific way. These ***protocols*** are typically laid out in an annotated table or document (sometimes called a codebook). For example, you might specify that people use specific names for each fish species, that they record time in 24 hour format, that turbidity be measured on a qualitative 1-5 scale, and that project team members use their initials to record who conducted the survey. These protocols should be developed by your data manager and monitoring teams and tested as you develop your data collection procedures. Wherever possible, as outline in Box 1, you should try to make use of existing data standards.

Once you have developed your data storage protocols, you should also figure out your protocols for transferring data from collection points to the project's databases and then to the ultimate users. For example, if the person conducting the fish transects records data on an underwater slate, then they may have to take that information and record it on a computer once they have completed the day's survey work. Likewise, the person conducting the survey may have to take answers from their field notebook and transcribe them.

As you develop your protocols, you also need to develop your long-term ***databases***. Most conservation projects will use some kind of electronic storage mechanism. There are many different types of databases that are available to accommodate different types of data. For example, quantitative tabular data can be stored in simple spreadsheet programs such as Microsoft Excel, in relational databases such as MS Access, or in custom web-based databases. Spatial data is typically stored in some sort of Geographic Information System (GIS). Note: Since nearly all WWF offices employ ESRI software as their main GIS tool, you should consider using ArcCatalog which provides a strong data management interface capable of establishing metadata, organizing the locations of file, and linking to ArcToolbox which in turn, allows for seamless import/export of files. Financial data are typically stored in a spreadsheet or in accounting software (many WWF Programme Offices currently use ACCPAC). Photos and other images can be stored in an album program that allows you to catalogue each entry with custom, searchable data tags that contain information about each photo. Failing these options, data can be stored in a computer directory, using a pre-established convention for file names. Here again, as outlined in Box 1, you should make use of common standards in developing your databases wherever possible. Finally, consider how WWF Connect might be able to help store and organize your data.

4. Review and transcribe data on a regular basis

The key step in managing incoming data is to implement the data review and transcription protocols that you have developed as soon as possible and as a part of normal operating routine. As a general rule, your project team members should try to transcribe their data as soon as possible after collecting it – all too often people allow data to pile up until it becomes overwhelming and then it never gets used. As you transcribe your data, it's also good practice to review and clean it up.

5. Clean and backup data

Cleaning data involves going through your data to catch any errors that were introduced during the collection, coding, or transcription processes. In particular, you should look for any gaps that may indicate missing data or for obvious outliers that signal some error. If possible, you should go back to the original data source to see if you can find the missing data or correct the errors.

Not backing up data on a regular basis is a mistake that almost everyone has to make for themselves before they truly appreciate the importance. Data managers should develop regular protocols for backing up data by making multiple copies and ideally, putting these copies in different physical locations.

6. Use and share data!

Data do no good if they just sit in a database. You thus need to make sure that your project analyses data and puts them to use, as outlined in Step 4.2 of the *Standards*. In addition, you should contribute your data to larger data sets as outlined in Box 1. WWF is a signatory to the Conservation Commons - meaning that we have pledged as a network to make our data available to all (where possible), so help magnify your conservation efforts by publishing your data for all to use.

Box 1. Standards That Enable You to Contribute Your Data to Larger Data Sets

One of the most important places you can contribute your data are to the growing number of databases that are developing at national, regional, and global levels around the world. If you contribute your data, then other practitioners can make use of your findings and learn from your experiences. In order to make your data accessible to outside parties, they need to conform to international data standards that provide the basis for open sharing of data. These standards need to occur on several levels:

Standard Software Formats – Your data need to be in an electronic format that other users can either directly read or at least import. For example, because of their dominant market position, most people can read files generated by Microsoft programs that use the *.doc, *.xls and *.mdb formats for documents, spreadsheets and databases respectively. Non-proprietary formats include HyperText Markup Language (HTML) used for web-pages or, in more recent years, various flavours of eXtensible Mark-up Language (XML). For example, in the world of spatial data, ESRI's Geography Markup Language (GML) is becoming the standard XML encoding for geospatial information.

Standard Data – Your data also need to fit the structure of the databases that you will be contributing to. As a simple example, if you are reporting numbers of birds of different species in a census and your data are in the form of nesting pairs whereas the database wants to know individual adult birds, your data will not be compatible unless they are converted. Most databases will outline the format that data need to be in.

Standard Terms – A particularly important aspect of the need for standard data is the need to have standard terminology. If you have recorded bird names in the local language, chances are they won't be useful at a global level. To this end, it's important to use scientific (latin) names. Similarly, if you call a threat "cattle grazing" and another project terms it "livestock" then there will be no way to compare results. Specific resources that you may wish to consult for terminology include:

- **Geographic Place Names** – GeoNet Names Server (GNS): http://gnswww.nga.mil/geonames/GNS/index.jsp
- **Biological Species and Other Taxonomic Information** – Integrated Taxonomic Information System (ITIS): http://www.itis.usda.gov/
- **Habitats, Threats, and Conservation Actions** – IUCN/CMP Classifications & Authority Files: http://www.redlist.org/info/authority_files
- **General Terms:** California Environmental Resources Evaluation System (CERES): http://gis.ca.gov/catalog/thesaurus.epl?mode=browse or GEneral Multilingual Environmental Thesaurus (GEMET): http://www.eionet.eu.int/gemet

Standard Metadata – In addition to contributing your data themselves, you should also contribute meta information about your dataset. Typical metadata include identification of the data being described, the source of the data and a contact person/organization, the quality of the data, entity ,or attribute information (if in a database or spreadsheet), its publication date, distribution information (including rights/liabilities), and the name of the individual completing the metadata. Metadata are typically recorded in a separate file (often using HTML or XML) that accompanies the main data file. The US Federal Geographic Data Committee has created a metadata standard that is fairly widely accepted throughout the world and is suggested to all WWF staff as a good metadata reference to employ. For more information on this standard, please go to: http://www.fgdc.gov/metadata/geospatial-metadata-standards.

Examples

Information Management Plan for the Thousand Islands - St. Lawrence Islands National Park (to be uploaded to WWF Connect).

WWF-Canada document: On WWF CONNECT (to be uploaded to WWF Connect).

References

Biodiversity Conservation Information System, 2000. Framework for Information Sharing: Principles. Busby, J.R. (Series Editor).

Federal Geographic Data Committee (FGDC). 2006. *Geospatial Metadata Standards.* Website available at http://www.fgdc.gov/metadata/geospatial-metadata-standards.

Higgins, J., 2005. Ecoregional Toolbox - Standard 3: Use a consistent data management framework in accordance with internal and partner organization data standards. The Nature Conservancy.

Margoluis, R. and Salafsky, N. 1998. *Measures of Success: Designing, Managing, and Monitoring Conservation and Development Projects.* Island Press

NatureServe, 2006: http://www.natureserve.org/prodServices/biotics.jsp Accessed March 23, 2006.

NRC, 1997. Bits of Power: Issues in Global Access to Scientific Data, National Research Council, USA, 1997.

Open Geospatial Consortium. http://www.opengeospatial.org/resources/?page=faq Accessed: March 23, 2006.

Reichl, O., 1998. An Information Management Plan for the Thousand Islands Ecosystem. St. Lawrence Islands National Park, Mallorytown, Ontario.

Salafsky, N. and R. Margoluis, 1999. Greater Than the Sum of Their Parts: Designing Conservation and Development Programs to Maximize Results and Learning. BSP: Washington DC http://www.fosonline.org/images/Documents/greater_than_layout.pdf

World Conservation Monitoring Centre, 1998. WCMC Handbooks on Biodiversity Information Management (8 volumes). Reynolds, J.H. (Series Editor). Commonwealth Secretariat, London.