

❖ **STANDARD 5: USE A CONSISTENT DATA MANAGEMENT FRAMEWORK IN ACCORDANCE WITH INTERNAL AND PARTNER ORGANIZATION DATA STANDARDS. [CONCEPTUALIZE]**

Rationale

Establishing and using a consistent data management framework will enable more effective analyses and updates within and across ecoregions, and support more effective implementation by the sponsoring organizations, partners and key stakeholders.

Recommended Products

- A populated database that includes all of the required data fields in the Ecoregional Assessment Data Standards¹ managed in a database structure consistent with those standards. Currently the Conservation Planning Tool [CPT] is the accepted standard for long-term data management and access.
- Spatial data representing the ecoregional portfolio, and source data on targets and threats used in the assessment.
- Documentation and metadata describing tabular and spatial data (products and source data) that includes projections, scale accuracy, data types, confidence levels, sources and contacts, etc.
- Data has documentation of, and is managed in accordance to, data sharing agreements, MOUs, restrictions on use, and disclaimers as negotiated with data sharing partners.

GUIDANCE

Data management is just as important as the data itself. Information learned and data created throughout the project may remain underutilized without good data management. If data is collected but not managed well or easily accessible, the “final plan” becomes the ONLY means of delivering the results to partners, stakeholders and the sponsoring organization. However, we have found that the underlying data supporting the assessment is perhaps the most important component of all. It underpins the ability of TNC/WWF and their partners to effectively use and implement the plan. A good data management program recognizes that: 1) data comes from a wide variety of sources, but should be managed in one place within a consistent framework; 2) datasets age, and should be accompanied by simple and effective recommendations for revisions and updates; 3) information and data will be shared with others and used in other phases of conservation action, and; 4) staff turnover will occur. Appropriate documentation allows knowledge to be passed on.

Neither The Nature Conservancy nor WWF have comprehensive data management strategies at present. However, the Conservancy has developed a number of tools, methodologies and

¹ http://conserveonline.org/docs/2004/09/ERA_Data_Standard_Version_1.0.zip

more recently standards, to aid in the management of tabular and spatial ecoregional information and is currently writing a conservation information systems strategy. This unit discusses data management strategies, data standards and tools, such as Geographic Information Systems (GIS) and TNC's Conservation Planning Tool (CPT).

Well-executed data management includes:

- A data management strategy,
- The support of a data manager,
- Adherence to standards for data and metadata,
- A readily accessible database design using appropriate tools such as GIS and database development software (i.e., CPT).

Data management strategy

Ecoregional assessment teams invest resources into data compilation to produce a set of products. This information can quickly become outdated, be misinterpreted because of lack of documentation, or even become lost. Conservation approaches depend on reliable, credible and current information. One of the great successes represented by TNC's ecoregional assessments is the collaboration between partner organizations and scientists who are gathering and interpreting the needed data. The data resource in hand during the assessment process is a unique and powerful resource, and can continue to be powerful only if well managed. Without a data management strategy for each ecoregional assessment, adequate assessment products may be produced in the short term, but in the long run we lose opportunities to leverage the wealth of data gathered. If poorly archived and documented, the data archive left behind once the project is complete may be of little use in following up on this work. For example, it becomes very difficult to assess key indicators necessary to monitor the progress toward our biodiversity conservation goals. Institutional data management strategies can enable the sharing of information between ecoregions and over large geographies. If adopted by the array of conservation organizations undertaking ecoregional assessments, the ability to utilize information between organizations can be realized.

We need to organize data systematically if we are to access, analyze, and share those data.

A data collection and management strategy should:

- fully understand the project's needed analyses, source data needed for analyses, and desired products for distribution.
- strategically gather the best available input data, and insure that adequate data sharing agreements are in place with data providers and users.
- insure appropriate spatial representation and scaling, as well as adequate cartographic standards, for maps and GIS products; using spatial data out of scale can lead to incorrect results.
- craft data standards, including standardized definitions, data dictionaries, and data models, by drawing on the experience of the organization, vetted with those who will use them, and shared with support provided. ERA Data Standard 1.0 is an example of

TNC using this approach. Standards should be formally adopted by the assessment team, and the project's data manager must support the team in using them.

- Use analytical software tools, such as spreadsheets, databases, or GIS, in ways that support data management by appropriately formatting and archiving resulting data sets.
- develop metadata associated with all spatial and tabular data sets archived.
- organize the output data and products for publication of reports, maps, or data sets.
- archive both the products and input data with documentation for access by others. Digital data archived be interpretable, documented, and designed to be transferable to other information systems in the future.
- recognize that the assessment produces an information product that is useful for awhile, while the data resource created may continue to be useful for additional purposes. In the end, the most important attribute of the data resulting from assessment is its ability to be interpreted by others and its ability to be transferred to other information systems when necessary.

This data management effort will enable the assessment team to:

- perform analyses needed for the assessment, and have access to these data for additional uses as well.
- insure data quality and accuracy, and perform confidence assessments of output data and products (e.g. viability assessments).
- define data gaps and needs for further assessment.
- develop data distribution plan that considers data licensing, sensitivity, data sharing agreements and MOU's.
- contribute the results of the assessment, including its archive of interpreted biodiversity information, to a larger conservation information effort that – if our conservation goals are to succeed – must be a partnership between TNC, WWF, Natural Heritage programs, Conservation Data Centers, and the many other organizations with whom we work.

Data management support

All ecoregional assessment teams need the support of a data manager to collect, standardize and manage information. This individual will perform, or will work with others to perform analyses and produce maps and other information products. The lead information/data manager should be identified as early as possible to develop the data management strategy and establish the data management structure for an ecoregion. This person should coordinate the management of the data during active planning and between editions of the plan. Data managers are also responsible for coordination among GIS analysts, technicians, Operating Units, project leads and team members. Organizations or field offices with limited staff capacity may need to contract with a US State Heritage Program, Conservation Data Center, or other partners to manage ecoregional information, but it will be important to

communicate expectations and establish clear data sharing agreements with these programs.

The data manager must be an integrated part of the assessment team. Their roles in defining data management strategy, identifying priority data to be managed, facilitating team members in interacting with the data, setting and enforcing standards, and systematizing data is essential to efficient assessment project work. The data manager's role is also key to insuring the project leaves behind a useful data archive that is documented and transferable for use in other future projects or information systems.

A one-size-fits all, standardized data management strategy will not be suitable for all assessments. However, data managers should be able to cross-walk and integrate their data management framework with other data management systems through the use of data, mapping, and cartographic standards. Existing data management tools will enable data managers to administer information, but in special cases they will need to create innovative solutions for data storage, analysis and distribution. Documenting this innovation is the key to providing learning opportunities. Documenting data gaps and known updates to key data not used in current iteration, but that should be used in the next iteration, is also crucial to enabling efficient future assessment iterations and revisions. Developing a "next steps" section in relevant reports will serve to document the identified areas where more data are needed to create a more complete decision support tool.

Ecoregional Assessment Data Standards and Metadata

In large organizations, dozens of assessment teams are collecting and managing information in parallel. Data standards guide the process of collecting, managing and storing data to keep data organized and useful for assessment teams, partners and for further analyses. Data standards identify the minimum contents of a data set including conceptual attributes and values for tabular and spatial data sets. Metadata is the documentation that accompanies any data set whether tabular or spatial. Metadata records information such as the source, reliability and scale of the data, the citation, appropriate uses of the data, and a contact person or agency. Metadata also documents the accuracy, projection and derivation of spatial data.

An organization-wide standard that defines the manner in which data are documented saves time and resources in developing data and preparing it for use. More importantly, it enables an organization to use data stemming from each individual effort to the benefit of the whole. Several organizations have developed and implemented data standards (US National Park Service, US Geological Survey, etc). The Nature Conservancy, along with input from partner organizations, has developed its first iteration of data standards to formalize data definitions and a data model for archiving ecoregional assessment data sets. The Ecoregional

Assessment Data Standard version 1.0 was released in August, 2004 and is available on ConserveOnline at:

http://conserveonline.org/docs/2004/09/ERA_Data_Standard_Version_1.0.zip

The standards are intended to guide ecoregional data managers and planners on the minimum data attributes that should be compiled in an ecoregional assessment, and how to format these for compatibility with other data sets and future iterations of CPT. They will enable us to roll up and share ecoregional datasets – an increasingly important task within and between TNC operating units and with our conservation data partners.

Version 1.0 of the standard addresses descriptive data for ecoregions, ecoregional assessment targets, viability, conservation goals, target occurrences, and conservation areas, but does not yet cover other data important to assessments such as threats, strategies, or spatial data associated with these. Additional effort will be made to expand on these standards to better incorporate topics including spatial data, viability and integrity, threats, and status measures. Assessment teams are encouraged to use this data standard to effectively manage conservation information in ecoregional assessments, and to contribute their experience to future standards development.

Data management tools

While no tool manages information for us, several software tools are in use in TNC, and several conservation information systems are in development. A Conservation Information Systems Strategy is currently being written to address our priorities for developing new systems.

As one chooses or develops the tools that will aid their data resource management activities, remember that information systems and software technologies come and go. In the end, the most important attribute of the data resulting from assessment is its transferability, while the most important attribute of the information system software used is its ability to provide for transfer of the data to another system to continue its utility in the future.

The information collected in an ecoregional assessment, ranges from tabular and spatial data of conservation targets (species, communities and ecological systems) and their attributes, to names of participants and experts consulted in the process. Below is a brief description of some of the primary tools used for spatial and tabular data management.

The Nature Conservancy developed the **Conservation Planning Tool (CPT)** to manage the common types of tabular data collected during the Ecoregional Assessment process. CPT is a relational database developed for Microsoft Access and is the Conservancy's primary tool for tabular data management designed to standardize data storage. With CPT, one can catalogue data related to portfolio sites, targets, ownership/administration, threats, strategies, stratification units, participants and experts, and general ecoregion description.

data. The tool helps organize and archive data, identify data gaps, standardized data for larger scale roll-ups and analyses and ready data for import/export for other purposes such as GIS and Conservation Area Planning. For a useful description of one team's experience with CPT see Populating the CPT: the Southeast Division's Experience in the Resources section of this document.

While the most-recently released version (CPT 1.5) does not comply with the new data standard, the soon-to-be released CPT 1.7 will incorporate all of the tables and fields from Ecoregional Data Standard 1.0 making it the tool of choice for assessment teams to meet the standard.

A Geographic Information Systems (GIS) provide one of the most comprehensive, standardized and flexible data management tools. Widely available, with ongoing support and innovations, GIS will provide a user with the basic data management and analysis tools to work with all spatial data, in both vector and raster formats. A GIS will also manage associated tabular information (the attributes) of spatial data and enable the user to update this information. Finally, the GIS provides a platform for map creation for communication products. There are many GIS software products with different approaches, costs, learning curves, sophistication, and each has its individual strengths and weaknesses.

ESRI is a software developer company that has developed dozens of specific tools for a variety of applications for the past 25 years. ESRI provides the standard GIS packages used at TNC due to its ease of use, reliability and scalability. Data development, analysis and storage are the more common applications of the ESRI tools. There are a variety of tools and applications..

ArcView 3.x is a simple, robust and programmable application that enables users to use data layers, connect them to databases, perform simple analyses and create good quality maps. Biodiversity analysis extensions for ArcView have been developed by a variety of authors.

ArcGIS is the more comprehensive software. ArcGIS offers extra tools that enable additional operability in analyses, databases, complex data models and metadata management. ArcGIS provides such a comprehensive GIS, that it is now becoming the standard recommendation for GIS, analysis and spatial data management. The preferred data management format is the Geodatabase.

Geodatabase is a data management software. It has the ability to store many types of spatial and tabular data. Since all related data are managed in one database, updates and distribution of ecoregional information product can be made much easier. The tabular data stored are available through Microsoft Access (version 2000) or ArcGIS (version 8.3 and above) software, and the spatial data stored are available through ArcGIS. The personal geodatabase allows you to create complex queries in Access and view spatial entities their relationship in ArcMap. In addition, the geodatabase can export spatial features as shapefiles or coverages. The Relational Database Management System (RDBMS) of the geodatabase data model explicitly allows users to link tabular data and targets, goals,

geography, and MARXAN input files with the spatial data representing them in one repository.

Also, ESRI offers a yearly training event for conservation applications that is accessible for non-profit organizations for discounted or free access. Read up on the Society for Conservation GIS sponsored by ESRI at <http://scgis.org/>.

There are several useful tools currently under development including a new version of the CTP, a core ecoregional assessment data model, and the Ecoregional Data Management Tools Software 1.0 (EDMT). These tools are innovations in the field of conservation data management.

The *Conservation Planning Tool version 1.7* is currently being developed. CTP 1.7 will have increased functionality and updated content including:

- an update of content and tables to more closely reflect requirements of ERA Data Standard 1.0;
- lookup tables for Threats and Strategies will reflect the latest TNC taxonomies; and
- values for Ecoregions and Assessment Areas will be populated with the most current lists and tools.

The *Core Ecoregional Assessment Data Standard and Model* is being developed to help implement the ecoregional data standards in new information systems. The standard will include fewer attributes, contain an explicit data model and will incorporate spatial data. The Core Data Standard and Model is meant to form the “core” of any data management system, allowing the flexibility to add additional components to suit regional approaches. The Core ERA Data Model has been designed in a UML (Universal Modeling Language) environment (Visio) that is exported and transferred in XML (Extensible Markup Language) format. This format is compatible with ESRI geodatabase creation and allows flexibility in sharing and implementing the model as new software.

In 2005, the Ecoregional Data Management Tools (EDMT) pilot project (see TOOLS section below) spawned a formal project to develop *Ecoregional Data Management Tools Software 1.0*, on a shared TNC server. This project is currently in development and results will be formally evaluated by TNC in April 2006 to determine how these tools may serve the organization. A summary of this project can be accessed [here](#).

Working in collaboration with the development of data standards and the development of a shared core data model, EDMT extends the current shared model and standards to provide functionality that extends the core model and existing CPT application. These include:

- 1) Integrating spatial data into the CPT model. EDMT 1.0 is built on an SDE/Geodatabase model, and features a personal geodatabase version for use on laptop computers.
- 2) Systematizing data resulting from ecoregional portfolios and also data extending beyond portfolio sites and across the ecoregional landscape.

- 3) Compatibility with site optimization tools such as Marxan, SITES, and SPOT, without requiring the use of these tools,
- 4) Freshwater and marine assessment data are supported and can be used in conjunction with terrestrial data for analysis and developing ecoregional-scale conservation vision.

OPPORTUNITIES FOR INNOVATION

Under development

CASE STUDIES

Under development

TOOLS

CPT version 1.5 is available from Marjorie Bennett (mbennett@tnc.org) or Kurt Eckerstrom (keckerstrom@tnc.org). View a summary of the new version [here](#).

[*Ecoregional Assessment Data Standard, Version 1.0*](#) is a standard that outlines the minimum required data attributes that must be compiled during an ecoregional assessment. The standard provides “dictionaries” of data attributes necessary for all stages of the EA process and definitions of these attributes.

EDMT GIS Tools. The Ecoregional Data Management Tool (EDMT) was a pilot project (2003–2005) tasked with designing a suite of products that would 1) meet the need of Northwest chapters for efficient data assembly and management and readily accessible products of ecoregional assessments (ERA), and 2) design and test a system for addressing these needs TNC-wide. A data model and custom GIS software have been completed and are useful for ERA data rollup (limited currently to the rollup of SITES input files), development of biodiversity status measures, development of custom products relevant to external partners, and support for analyses at the scale of ecoregional portfolio sites, multi-ecoregions, and larger (e.g., global habitats). (*note: this is all done within the assigned assessment units chosen across ecoregions, and therefore limited to that scale of analysis*) A comprehensive summary of the development and use of these tools can be accessed [here](#). For more information contact Steve Farone (sfarone@tnc.org).

[*GIS Tutorial Dataset*](#) provides data and a tutorial to teach basic GIS functions useful to conservation planning. This tutorial can be found at <http://gis.tnc.org/knowledge/training/CSDintroAV.php>

An example of a Metadata report for the Southern Rocky Mountains Ecoregion Boundary can be viewed at http://gis.tnc.org/community/projects/FWI/metadata/srock_ecoreg.htm

RESOURCES

Websites

ArcView tools are available at the GIS.TNC.ORG website. See http://gis.tnc.org/systems/sw_links.php?display=sw_esri

The authoritative source for extensions for ArcView can be found at <http://arcscripts.esri.com/>

To obtain ESRI GIS products for Conservancy offices, please contact Susan Miller (smiller@tnc.org) or Demian Ryebock (dryebock@tnc.org) at the Conservation Systems Office.

Publications

BCIS has developed an eight volume framework for data management and sharing available at www.biodiversity.org

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

Executive Overview

- Volume 1: Principles
- Volume 2: Procedures Manual
- Volume 3: Custodianship
- Volume 4: Data Access
- Volume 5: Metadata
- Volume 6: Standards & Quality Assurance
- Volume 7: Core Datasets
- Volume 8: Tools & Technologies

Groves, C. (2003). Drafting a conservation blueprint: A practitioner's guide to planning for biodiversity. Washington, The Nature Conservancy. Island Press.

Minor, Brit (2005). Populating the CPT: the Southeast Division's Experience. Southern Resource Office, The Nature Conservancy.

<<http://conserveonline.org/2005/03/m/en/PopulatingSEDivCPT.doc>>

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

Sayre, R., E. Roca, et al. (2000). Nature in Focus: Rapid Ecological Assessment, Island Press.