

CHAPTER 32

NEW TECHNOLOGIES FOR LONG-TERM BIODIVERSITY MONITORING

Data quality & real-time availability

HUGUES LEFRANC¹, REGINO NÚÑEZ² & JUSTIN STEVENTON³

¹ *Natural Processes Monitoring Team, Doñana Biological Station (CSIC), Sevilla, Spain.*

² *Advanced Technology Systems (SATEC), Sevilla, Spain.*

³ *Steventon Consulting, Redmond, WA, 98052 USA*

Corresponding author email: hugo@ebd.csic.es

1. INTRODUCTION

Now, more than ever before, environmental data are needed to inform efficient environmental management. Information on biodiversity and ecosystems is essential, for governments as well as the scientific and educational communities. Much of this information, however, is not readily accessible, yet scientists need access to high quality data for their research; planners need readily available and reliable information to make appropriate decisions for the management of environmental resources; and we should not forget the need to satisfy the growing interest in biological information from the general public. Enabling access to this information is of vital importance in the face of the rapid decline in biological diversity. The purpose of this chapter is to describe the solution developed to automate the process of data collection and facilitate the use of long-term monitoring information by the Natural Processes Monitoring Team (Equipo de Seguimiento de Procesos Naturales, ESPN) of the Doñana Natural Park.

2. OBTAINING QUALITY DATA

Data are of high quality if they are suitable for use in operations, decision-making and planning (Juran 1964 in Chapman 2005). It is important for the organisations that plan to make their data available to others to have a policy with respect to the quality of data. Projects for the evaluation of biodiversity require large amounts of information on species in order to obtain significant results. These projects are very expensive to run, and data collection, particularly gathering information on species, is invariably the most time-consuming activity. However, data quality principles must be applied at all the stages of

C. Hurford, M. Schneider & I. Cowx (eds.), Conservation Monitoring in Freshwater Habitats

© 2009, Springer. Printed in the Netherlands

the process of data management (capture, digitalisation, storage, analysis, presentation and use). Any loss in quality at any of these stages reduces the applicability, usage and validity of the data.

Traditionally, during fieldwork, observers either follow a standard protocol and fill in the appropriate forms or make field notes recording the place, date, time, location, etc. The forms are then handed in to the office for their storage and subsequent analysis. This method, although widely used, has some problems both in the field and in the office.

Data collection with pen and paper is not an ideal method for guaranteeing the quality of the information. Handwritten material could, at the time of digitalisation, be badly interpreted due to the handwriting or even discarded for being illegible. The drifting attention of the observer due to the repetitive nature of the work, or due to bad weather, can lead to errors in form completion, or even the loss of forms. The absence of a structure or uniform format in the field notebooks makes the transfer of the information to a database even more complex and time-consuming, not to mention the tedious process of transcribing handwritten data into a digital format that also can lead to errors.

The key factors for improving the quality of the data are prevention and validation. The prevention of errors is closely related to the information gathering and database entry phases. The development of a simple and easy-to-use interface for data input can be a way of minimising errors, because there are fewer opportunities for them to occur. Although considerable effort should be directed towards the prevention of errors, it is also important to consider the validation of data (error detection plus correction). And one of the best ways to reveal errors is to allow data to be seen (Edwards 2004).

3. NEW TECHNOLOGIES

Biodiversity informatics (Bisby 2002) is an emerging discipline born from the need to find more efficient methods for capturing and processing environmental information. It draws on collaborations between biologists and software and network engineers. The objective is to acquire with precision, represent, communicate, integrate, analyse and apply the information extracted from natural systems. It is supported by the application of new technologies to guarantee the fulfilment of protocols, and the introduction of data integrity and quality controls. This system will facilitate the rapid presentation of reports and data in real or quasi-real time, with all of this based on a powerful device for database management.

New technologies have the potential to revolutionise the process of data gathering. Personal Digital Assistants (PDA) and handheld electronic diaries have the same functionality and capabilities as microcomputers and, owing to their portability, allow observations to be digitised in the field. For example, the Global Positioning System (GPS) allows users to accurately register the location of specific sites; over the Internet, via WiFi (wireless Internet) or using a mobile telephone network, users can send observations from a remote database in the field to database servers; and, finally, the data can easily be made available to analysts and managers through the Web.

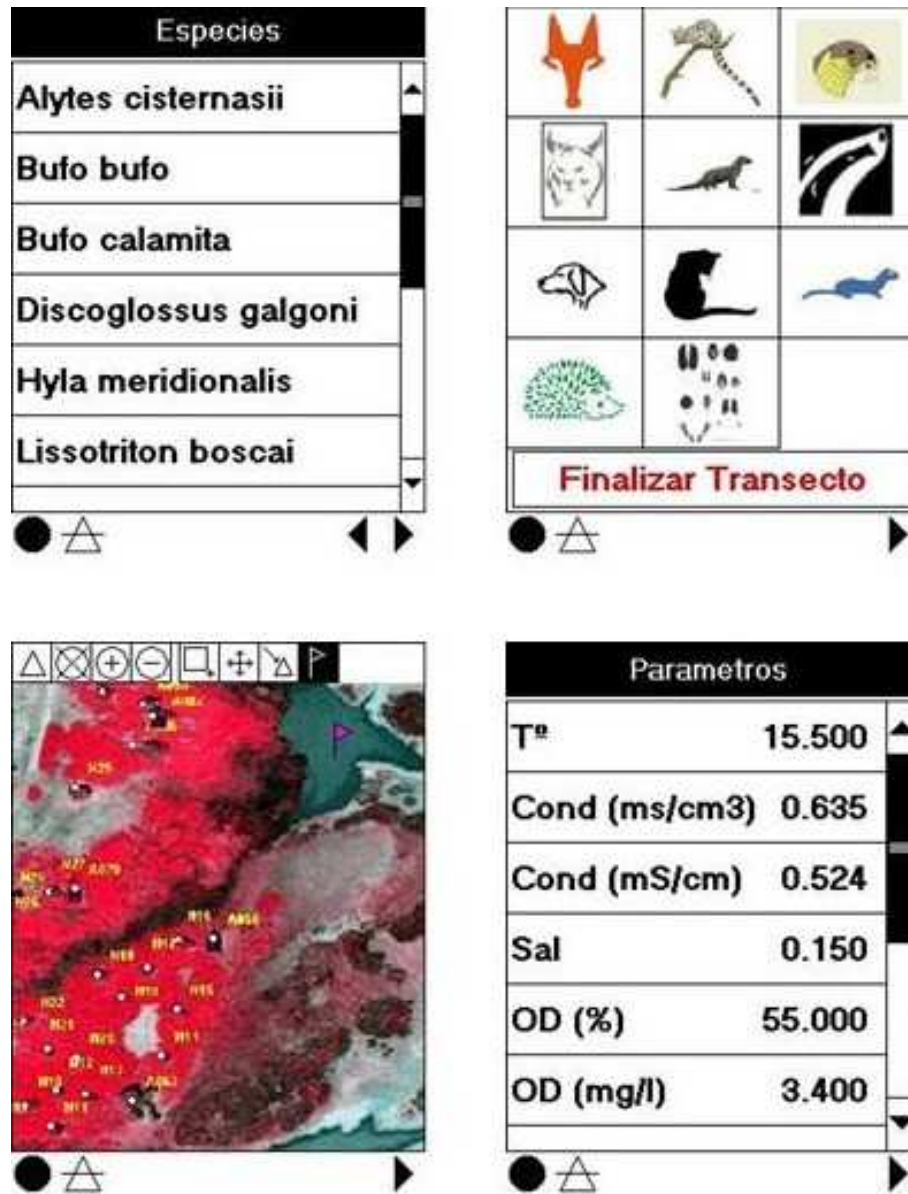


Figure 32.1 Examples of PDA screens used for collecting field data.

One of the most revolutionary of these new technologies is the free software CyberTracker (www.cybertracker.co.za). This is a program for data capture widely used in environment and wildlife conservation. It is a very efficient interface for collecting large quantities of geo-referenced data. CyberTracker can incorporate positional data from a GPS - if available (some e-diaries have built-in GPS, and others allow connection to small GPS receptors via Bluetooth). The original concept behind its development was to

create an interface that was very user-friendly and could be used by fieldworkers. The tool is based on the principle of using the touch screen of the PDA or electronic diary to store information. Computerising data *in situ* is considerably more efficient than digitising the data manually.

The application is very simple and allows for personalisation in its use. A user can design and edit a database, developing screen sequences with different interfaces (Fig. 32.1). For each methodological protocol for data acquisition, a sequence is created in which, step-by-step, all the possible options and elements needed for describing field observations should appear on the screen. The sequence, which is the same for all the observers, thereby enables a standardisation of the methodology, leaving little leeway for personal interpretation of protocols by fieldworkers.

With CyberTracker, the design of a database with sequential screens does not require knowledge of programming. The creation of elements for each screen allows the automatic generation of a structured database.

Once the data are downloaded onto the PC, they are transferred into a database for permanent storage, analysis, visualisation and export. Data are viewed in tables but can also be projected over maps and aerial photographs, allowing the verification of the entries before their final validation. The information can also be rapidly exported into other programs, such as spreadsheets or GIS (Geographical Information System) software, for advanced analysis. In this way, CyberTracker integrates the traditional gathering of environmental data and the disciplines of spatial information analysis.

In the context of a set of manual procedures with little or no automation, the ESPN of Doñana Natural Park began to consider the idea of incorporating this new technology for the process of capture, classification and exploitation of data deriving from its own work on long-term monitoring programmes. Due to the diversity of protocols, the vast number of observers and the annual deadlines for submitting reports, the use of this technology has meant a radical change in the way of acquiring data in the field, leaving aside the traditional field book, in exchange for the modern PDA.

4. THE STARTING POINT

In October of 2007 a new work area was developed in the ESPN, the area of data quality. The objective was to introduce IT tools to improve data collection, preserve the integrity of the information and offer end-users the information more rapidly. From March 2007, we started to use the CyberTracker software for the acquisition of data in more than 50 protocols for long-term ecological monitoring in Doñana Natural Park.

The ESPN was born with the objective of generating basic, reliable and verifiable information about the state of the physical and biological environment of Doñana. This information is used by the managers of Doñana Natural Park, as well as the scientific community, to detect changes in the fauna, flora and environmental quality of the protected area.

Table 32.1. Wetland monitoring protocols used in Doñana.

Habitat	Objectives
Aquatic fauna	Quantitative estimation of fish, amphibians, reptiles and macroinvertebrates in temporary and semi-temporary ecosystems. Specific monitoring of <i>Procambarus clarkii</i> (alien species). 38 locations: twice a year
	Quantitative estimation in shallow tidal waters of the marshes 3 locations: every other month
	<i>Trachemys scripta elegans</i> (exotic species) 4 locations: monthly
	Population dynamics of native turtles 5 locations: every four years
	Survey of amphibians in temporary ponds 30 locations: monthly checks when there is water
Aquatic vegetation	General description of coverage Specific description of the coverage, phenology and location Estimation of biomass 38 locations: twice a year
	Non-systematic monitoring of protected and alien species
Water	Estimation of water quality (T°, Cond., DO, pH, NO ₃ , NH ₄ , Chlorophyll a and turbidity) in temporary and semi-temporary ecosystems. 38 locations: twice a year.
	Estimation of water quality in ephemeral ponds Estimation of the area submerged and duration 30 locations: monthly checks when there is water
	System for early alert of algal blooms At least four locations: at least twice a week in high risk periods

The main aim of the ESPN consists of the implementation of an open-ended monitoring programme, to evaluate the spatial and temporal tendencies of the biological diversity and the ecosystems of Doñana. Its work is supervised by a scientific committee composed of researchers both within the Biological Station and from elsewhere. Every year since 2002, the ESPN and its 18 technical staff have applied more than 50 methodological protocols for capturing information in fields as diverse as water quality, hydrological dynamics and geomorphological processes; land use; dynamics of the aquatic and terrestrial vegetation communities; herbaceous plants and bush productivity; changes in the surface coverage of plant matter; distribution and population dynamics of invertebrates, fish, amphibians, reptiles, mammals, and of threatened, key and invading species. Some of the data collection process was already computerised, with data from PDA's downloaded onto non-shared databases and/or spreadsheets that had to be subsequently homogenised and filtered. Finally, after technical validation, the information was exported to permit its publication on the web, with this process being carried out by hand.

The introduction of these new technologies in the ESPN has involved a major investment in training to extend their use across the workforce. Interestingly, some of the members of the ESPN, who were initially reluctant to use the new IT, were later among the most convinced and adept users of the PDAs and CyberTracker. This updating process has also meant an improvement in protocols, as it entailed a profound revision of the data

collection procedures. After the initial implementation, it can be stated that the most important benefit has been the improvement in data quality, reflected in the complete compliance with protocol requirements, the disappearance of omitted data and semantic homogeneity. In addition, the experience has allowed the maximum automation of the chain of data capture, reducing the minimum number of users necessary, for the processes of acquisition and digitalisation, to just one. Thus, CyberTracker has shown itself to be an optimum program to comply with most of the criteria for data quality.

This process has returned naturalists to a leading role on the staff in place of the technical experts. It is a very democratic program in the sense that with knowledge of nature, and a little practice, anybody can collect data with this system with no previous expertise. Thanks to CyberTracker, we have made the most of the time spent on the fieldwork. For example, in the case of the monitoring of the aquatic ecosystems of Doñana, the limnological sequence allows the user to perform 18 different protocols with the same sequence (Table 32.1). Once the data have been downloaded, all the information can be interrogated through the same database.

Another very important benefit of the software is the considerable saving of time dedicated to the digitising process. This gain allows the observer more continuity in their fieldwork, while at the same time, affording more time for error detection. The possibility of seeing the data directly, the same day, in a table and projected over a map, makes it easier to detect errors in numbers and in geographic positioning or failed data.

However, CyberTracker, despite all the improvements that it has given to the ESPN, has not provided, to date, a reduction in the time between data acquisition and delivery to the final users, i.e. the scientific personnel, managers and the general public. As the program does not amass all of the consolidated data in a single platform or unified scheme that would allow independent access to the process, it has not solved the problem of distributing the information. If the data cannot be downloaded centrally onto a single PC, the information manager will have the laborious task of compiling the data, and the more PDA users there are collecting the same sequence of data, the more dispersed the information will be. Updating CyberTracker was also difficult: if we wanted to change a data sequence, we would have to gather in the PDAs and do this manually. Furthermore, the software did not, originally, allow us to make the information gathered by the technical staff of the ESPN available in real time. Therefore, we proposed a new software development that would benefit both CyberTracker and the ESPN. This was necessary to centralise data and to deal with incorporating the data from all the monitoring protocols (more than 50) and from the technical workforce (18 people).

With the financial help of the Spanish government programme for the recognition of Unique Scientific and Technical Infrastructures (ICTS), and in collaboration with the consulting company SATEC (specialists in communication systems), the ESPN asked the CyberTracker programmer, Justin Steventon, to carry out some modifications and improvements to the software. Basically, the technical request consisted of making it possible to use CyberTracker in a network, with all the sequences and databases migrating to a single server. The synchronisation, the transfer of sequences, the downloading of data and the visualisation of data were to be performed remotely from the server using an internet connection. As a consequence of these improvements, it would be possible to

download data collected with the PDA to the server, from any location with wireless coverage or by using the mobile telephone network.

5. PLANNING FOR DATA TRANSFER TO A CENTRAL SERVER

The first step for optimising the process of data capture and storage is to develop a structured diagram showing how the processes of capture, consolidation and validation, and finally exploitation of data can be separated on different levels, based on a classical three-level diagram with a client-server structure.

5.1 Data capture

At this level we needed to generate the sequences necessary for allowing data capture by PDA users. The data can then be transferred either in real time (if there is wireless or mobile phone coverage) or delayed, and then transferred en masse when the mobile devices are connected to PC workstations. The basic tools for this process are:

- PDA with Wireless connection;
- An updated version of CyberTracker, for data transmission through the web; and
- Observation sequence (methodological protocol).

5.2 Data consolidation

This is facilitated using a network relational database, MySQL. This database (DB) will store data in double format: on the one hand, it is a repository of the CyberTracker database itself, and, on the other hand, due to the proprietary way in which CyberTracker stores its data (encrypted), it also stores certain types of data in a non-encrypted format, so that they are accessible directly by third party applications. The basic components of this level are:

- FTP Server;
- CyberTracker; and
- MySQL DB Server.

5.3 Data exploitation

This has two methods of consulting and/or exploitation, due to the double format in which data are stored. The first one is for the CyberTracker application itself, to enable researchers to perform complex tasks such as making advanced filters for consultations; and the second method is to enable the general public to access these data en masse, through the ICTS of the Doñana website. The components of this third level are:

- CyberTracker; and
- ICTS Web (<http://icts.ebd.csic.es>)

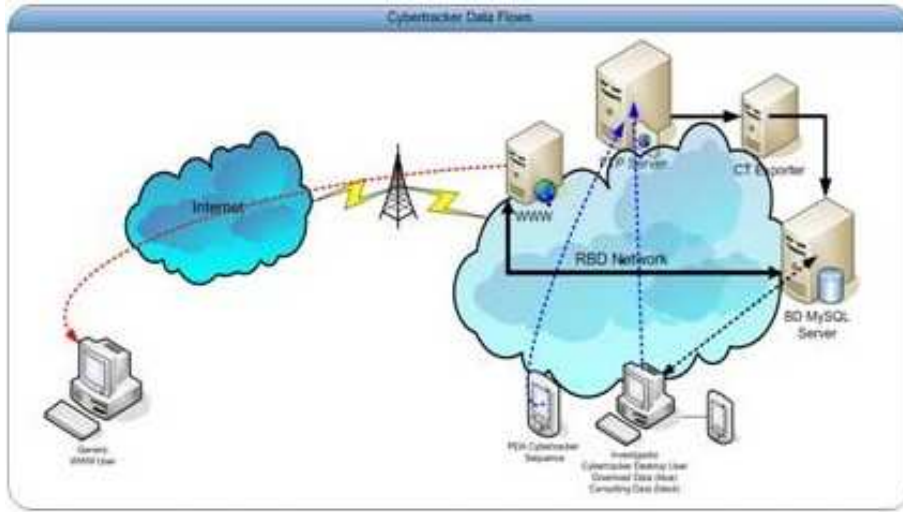


Figure 32.2. A diagram illustrating the network infrastructure in Doñana Natural Park.

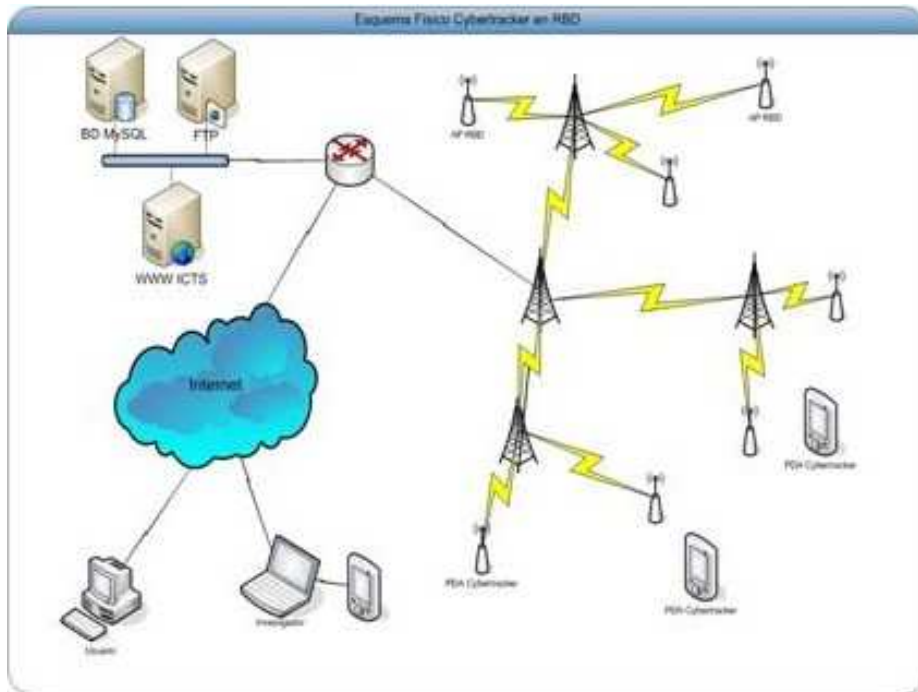


Figure 32.3. A diagram illustrating the flow of monitoring data in Doñana Natural Park.

Figure 32.2 shows the physical diagram outlining the implementation of the new processes for monitoring with CyberTracker in the Doñana Biological Reserve (DBR).

Observations made can be sent to a FTP/HTTP server on the DBR network from the PDAs, making use of the WiFi Access Points to the Reserve's data network or via an Internet connection directly from the PDA, and from a PC, connected to the Internet and synchronised with the PDA. The data on the FTP/HTTP server are transferred onto the MySQL DB using scheduled processes of CyberTracker and stored in both non-encrypted and CyberTracker proprietary formats. Both generic users and researchers can access the data, the former via the Internet, by consulting the website of the ICTS, and the latter, by access through the CyberTracker application that is installed on their personal equipment and making use of the centralised database. The stages in the new monitoring scheme are as follows:

The observer loads the CyberTracker sequence corresponding to the appropriate monitoring protocol onto the PDA. Data collection is initiated following the chosen sequence step-by-step. The CyberTracker configuration allows the instantaneous transmission of the data when desired if there is wireless or GSM coverage, or subsequently, when the aforementioned connections become available or on instigation by the user. The data are sent to an FTP server in the CyberTracker proprietary exchange format. A process launched by a Windows scheduler downloads the data from the FTP server to the MySQL DB in CyberTracker format, while, in parallel, another independent process exports the CyberTracker data to tables previously designed for each type of sequence/protocol, to ensure that these data can be used by applications other than CyberTracker. This flow of data between users and the servers is illustrated in Fig. 32.3. Researchers can use the CyberTracker software from their own equipment through the central DB. Generic users can view the data in a standard format through a web interface using the application for data exploitation on the ICTS website.

6. THE APPLICATION OF OUR DESIGN IN INTERNATIONAL MONITORING NETWORKS

One of the major difficulties in forming global strategies for evaluating biodiversity is data comparability. Reliable measurement of trends and changes in biodiversity requires standardised data collection and analysis in order to be comparable. Therefore, if we want to evaluate data both at a local scale and global scale, we will need a coherent system for data collection and communication.

The technology incorporated in the long-term ecological monitoring programme of Doñana Natural Park has recently been put forward as a technological reference for the regional network LTER-Europe (European Long-Term Ecosystem Research Network, www.lter-europe.net). The aim is to extend the use of the CyberTracker software as a tool for the normalisation and harmonisation of data acquisition protocols. This would improve the quality, and quasi-real time availability, of data for end-users.

7. REFERENCES

- Bisby F.A. (2002) The quiet revolution: biodiversity informatics and the internet. *Science* **289**, 2309-2312.
- Chapman A.D. (2005) *Principles of Data Quality*. Report for the Global Biodiversity Information Facility 2004. Copenhagen: GBIF.
- Cotter G., Frame M., & Sepic R. (2004) Integrated science for environmental decision making: the challenge for biodiversity and ecosystem informatics. *Data Science Journal* **3**, 38-59.
- Edwards J.L. (2004) Research and Societal Benefits of the Global Biodiversity Information Facility. *BioScience* **54**(6), 485-486.