german cooperation
DEUTSCHE ZUSAMMENARBEIT

Published by:

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

BIODIVERSITY INFORMATION MANAGEMENT AND REPORTING GUIDELINES FOR SOUTH-EAST EUROPE

In cooperation with GBIF IUCN

BIODIVERSITY INFORMATION MANAGEMENT AND REPORTING GUIDELINES FOR SOUTH-EAST EUROPE

Published by the
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn, Germany

Open Regional Funds for South-East Europe – Biodiversity (ORF-BD)
GIZ Country Office in Bosnia and Herzegovina
Zmaja od Bosne 7-7a, Importanne Centar 03/VI
71 000 Sarajevo, Bosnia and Herzegovina
T +387 33 957 500
F +387 33 957 501
info@giz.de
www.giz.de/regional-funds-southeasteurope

As at
October 2017

Printed by
Agencija ALIGO o.r.

Design and layout
Vanesa Prodanovic

Prepared by
Exatto d.o.o. za informacijske tehnologije

GIZ ORF-BD team in charge
BIMR Project Manager / Coordinator for Bosnia and Herzegovina
Azra Velagic-Hajrudinovic (azra.velagic-hajrudinovic@giz.de)
BIMR Management Advisor / Coordinator for Montenegro
Jelena Perunicic (jelena.perunicic@giz.de)

Text
Marin Grgurev, PhD
Petra Strbenac
Gabor Mesaros, PhD
Robertina Brajanoska
Halil Ibrahimi, PhD
Sanela Shkrijelj
in cooperation with Tim Hirsch (Global Biodiversity Information Facility – GBIF)
Maarten Hofman (International Union for Conservation of Nature Regional Office for Eastern Europe and Central Asia - IUCN ECARO)

Reviewed and endorsed by
BIMR Regional Platform South-East Europe

GIZ is responsible for the content of this publication.

On behalf of the
German Federal Ministry for Economic Cooperation and Development (BMZ)

ABBREVIATIONS

ABCD schema - Access to Biological Collections Data

ADBC - Advancing Digitization of Biodiversity Collections

API - Application Program Interface

BioCASE - Biological Collection Access Service

BIS - Biodiversity Information System

CAEN - Croatian Agency for Environment and Nature

CBD - Convention on Biological Diversity

CBI - Centre for Biodiversity Informatics, Faculty of Biology, University of Belgrade

CDDA - Common Database on Designated Areas

CITES - Convention on International Trade in Endangered Species of Wild Fauna and Flora

CoL - Catalogue of Life

DiGIR - Distributed Generic Information Retrieval

DINA - DIgital Information system for NAtural history data

DOI - Digital Object Identifiers

DwC - Darwin Core

DwC-A - Darwin Core Archive

EC - European Commission

EEA - European Environment Agency

EIS - Environment Information System

EML - Ecological Metadata Language

ETC/BD - European Topic Centre on Biological Diversity

EU BON - European Biodiversity Observation Network

EU IPA - EU Instrument for Pre-accession Assistance

GBIF - Global Biodiversity Information Facility

GEF - Global Environment Facility

GUID - General Unique Identifier

IBA - Important Bird Areas (BirdLife International)

iDigBio - Integrated Digitized Biocollections

IPA - Important Plant Areas

IPT - Integrated Publishing Toolkit

IS - Information System

ITIS - Integrated Taxonomic Information System

LSID - Life Science Identifiers

IUCN - International Union for Conservation of Nature

MGRS - Military Grid Reference System

NBIS - National Biodiversity Information System

NBN - United Kingdom National Biodiversity Network

NCD - Natural Collections Descriptions

NPIS - Nature Protection Information System

NUTS - Classification of Territorial Units for Statistics

PESI - Pan-European Species Directories Infrastructure

SDF - Standard Data Form

TAPIR - TDWG Access Protocol for Information Retrieval

TDWG - Biodiversity Information Standards

UTM - Universal Transverse Mercator (coordinate system)

UUID - Universal Unique Identifier

CONTENTS

PREFACE

South-East Europe (SEE) is one of the richest parts of Europe in terms of biodiversity. In order to conserve and sustainably use these biodiversity assets and valuable natural resources under a concerted regional approach, a regional consensus on principles and key elements of a biodiversity information management and reporting (BIMR) mechanism in line with the Convention on Biological Diversity (CBD) and European Union (EU) requirements is required. It will enable regional exchange of data and information for collaborative monitoring, reporting and management of (shared) biodiversity resources. Accession to the EU constitutes a common goal for economies of SEE, where an important prerequisite is the transposition and full implementation of the environmental acquis communautaire, especially the Birds Directive (2009/147/EC), the Habitat Directive (92/43/EEC) and the EU Biodiversity Strategy 2020. Therefore, BIMR is a crucial component for all economies in the SEE region and improvements are needed.

The first ORF-BD publication, *BIMR Regional Assessment Baseline for South-East Europe*, focused on the current stakeholder situation, as well as on the policy, legal and institutional frameworks and information system set-ups in the field of biodiversity. It highlighted various gaps regarding the availability and quality of data as well as technical and human capacities and skills, particularly in terms of: a) design and complexity of biodiversity information systems; b) standardized data forms for data collection; c) data formats and availability; d) stakeholder cooperation, data usage and authorship rights; e) capacities and skills for BIMR related tasks; f) legislation related to BIMR; and g) data backup, insurance and data sharing.

In order to improve SEE biodiversity data management and reporting towards the CBD and the EU, a second ORF-BD publication, *BIMR Regional Guidelines*, was developed by regional experts and reviewed by EU accession economies in the period from June – October 2017. It serves to assist all governmental and non-governmental stakeholders in effective data management, thereby systematically improving both the quality and usability of data and compliance with EU and CBD standards and requirements.

The German Federal Ministry for Economic Cooperation and Development (BMZ) supports the deveopment of BIMR Regional Guidelines and pilots implementation through the Regional Network for Biodiversity Information Management and Reporting project as part of the GIZ Open Regional Funds for South-East Europe-Biodiversity (ORF-BD), in close dialogue and coordination with the relevant stakeholders and partners.

We hope that these comprehensive *BIMR Regional Guidelines* will be used widely and contribute to enhanced monitoring and reporting across the SEE region and beyond.

Gabriele Wagner
GIZ Sector Fund Manager – ORF-BD

ACKNOWLEDGMENT

This publication is a result of joint effort of ministries, competent authorities, research institutions, NGOs and experts from EU accession economies in the SEE region to develop regionally adapted guidelines for planning, design and development of biodiversity information systems. This endeavour, which involved pooling of expertise from the SEE region as well as from Croatia, was pursued with determination and in the spirit of excellent cooperation at all levels: political, technical and administrative. All parties and persons involved are greatly acknowledged for their contribution to this work.

INTRODUCTION

Biodiversity is a broad term that is a part of our daily life and livelihood; in general, it refers to variation at the ecosystem, species and genetic level. Collection of scientific data on biodiversity is a demanding process and securing adequate data storage, maintenance and exchange on a national level as well as globally is still a challenge.

To adequately assess biodiversity status for the purposes of robust decision-making and management of biodiversity and ecosystems, we need digitized, structured and verified biodiversity data. Additionally, there is a need to establish mechanisms for exchange of data, standards and experiences. This can be achieved through a review and implementation of common technical and biodiversity standards for data exchange, species and habitats lists, as well as through continuous dialog, coordination and communication among all relevant stakeholders in the region.

For this reason, these Guidelines focus on topics/aspects that are important to tackle and consider in the course of planning, designing and developing biodiversity information systems, and also in the process of biodiversity data management and reporting.

Clean and structured data is a fundamental prerequisite for all analyses and reporting, and represents the core value of any system. The Guidelines therefore place special focus on the means of producing and ensuring good quality data that can easily be managed, analysed and exchanged between different stakeholder and systems.

BIMR Regional Guidelines are intended for stakeholders from the environment and nature protection sector that handle biodiversity data management and reporting tasks on the local and regional level.

Targeted stakeholders are expected to have a basic understanding and knowledge of the biodiversity data collection process and the fundamental terms used in the technical aspects of biodiversity data management.

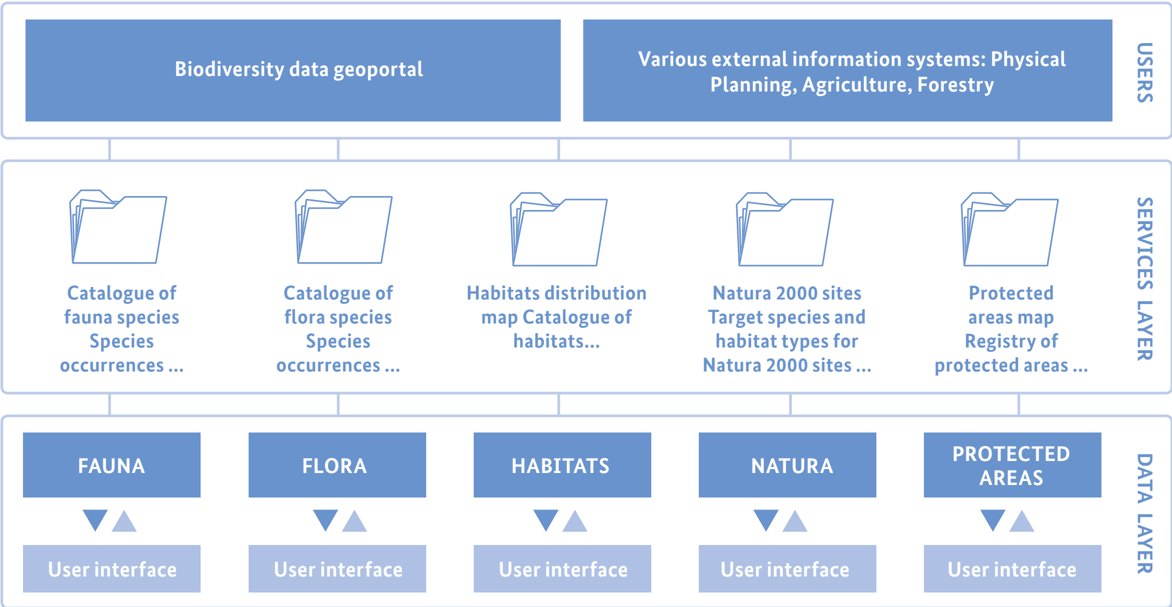
Each chapter in the Guidelines contains basic information and guidance for application, including practical examples where appropriate. Each chapter also includes a comprehensive list of references (online, literature) and additional resources for detailed explanations and further information.

TOPIC 1

BIODIVERSITY INFORMATION SYSTEM COMPONENTS

Biodiversity information system concept and definitions

A **Biodiversity Information System** (BIS) consists of various thematic databases, applications, processes, protocols and web services intended for storage, maintenance and exchange of data related to biodiversity and nature protection. It is important not to perceive a BIS as "one big database" but more as an integrated system of multiple different and interconnected **thematic databases** (flora, fauna, habitats, protected areas etc.) which do not have to be hosted in the same institution and managed by the same authority. By displaying data and metadata in accordance with the relevant technical standards, each of these databases can function on its own and also act like a part of a complex system.



▲ Figure 1: *Biodiversity Information System - general concept scheme based on the example of the Croatian Nature Protection Information System (NPIS)*

The figure above represents a general concept scheme of a BIS, based on the example of Croatia and their Nature Protection Information System (NPIS). The information system might also include various other more-or-less specific databases, such as databases for marine data, speleological objects, fungi and lichens, and other specific databases and corresponding services.

Biodiversity data is stored across various thematic databases and is maintained via separate applications/ interfaces. Data is displayed via web services in order to facilitate data exchange between different BIS components and with various external information systems.

Thematic databases that are most relevant to BIS include species (floristic and faunistic) and habitat databases, nationally designated protected areas, Natura 2000 ecological network, and speleological objects such as pits and caves.

Species database (flora/fauna) – consists of at least two components:

A. **Flora/fauna checklist** - relevant, recent and continuously updated and taxonomically validated by multiple experts for different taxonomic groups. Apart from species systematics and taxonomy, this component also includes species description data, including species distribution areas, species conservation status, threats, endemic status and similar data.

B. **Species occurrence records** which include field inventory, museum collections and literature data covering species occurrences and any other data that includes evidence of a named organism found in nature.

Habitat database – is a thematic database intended for storage, maintenance and exchange of data related to habitat types. This database contains a catalogue of habitat types relevant for the national habitat classification, which includes information such as unique habitat code and name, description, endangerment, other classification codes, typical species, threats, etc.

Optionally, and depending on the use of other classification systems, the database needs to incorporate a coding scheme and a method for matching/converting habitat types to/from other classification systems.

A habitats database also includes spatial data, both raster and vector, representing distribution areas of habitat types as well as various other raw spatial data related to field observations.

Protected areas database – includes borders of nationally designated protected areas with corresponding description data from the registry of protected areas. This data includes information about the protection category, designation date, area border description, international protection status, etc.

Natura 2000 database – consists of spatial data of SCI and SPA areas, as well as description data in accordance with the Standard Data Form (SDF). This data covers lists of target species and habitat types for each Natura 2000 site as well as other related data.

Cadastre of speleological objects – includes a repository of speleological objects with corresponding description data, covering topics such as geomorphology, hydrology, archaeology, palaeontology etc. It also includes data from field research. Ideally the cadastre should also include cave plans and drawings which are necessary for further exploration and planning.

Common databases/catalogues – consist of various smaller-scale databases which provide information that accompany biodiversity species and habitats data, such as literature references, multimedia files, researchers/ experts and project catalogues, lists of georeferenced localities, and similar. This data represents common data required by different BIS components. To ensure efficient management, data is maintained and updated in one place (within the scope of one database) and is

	Module for flora	Module for fauna	Module for habitats	Natura 2000	Protected areas	Module for speleological objects
Natura 2000	■	■	■	■	■	■
CBD	■	■	■	■	■	■
CITES	■	■				
CDDA					■	
Eurobats		■		■		■
IPA	■		■		■	
Ramsar	■	■	■		■	
IBA[1]		■			■	

▲ **Table 1:** *BIS components and data required for various reporting purposes*

shared between various BIS components via web services. It is important and mandatory to plan these as an integral and important part of a BIS.

Each of the BIS thematic components includes information necessary for various reporting purposes. The above table indicates BIS components that include data required for some of the most common reporting purposes.

What to consider when planning to develop a BIS?

Development and planning of biodiversity information systems is a complex and long-lasting process that requires an interdisciplinary approach and inputs from IT experts as well as from biology and nature conservation experts. The following section provides an overview of several aspects to consider when planning, implementing and managing a BIS.

Planning phase

Ensure adequate number of information technology experts as well as biodiversity experts to support the information system in terms of planning its content and functionality, system design and architecture, and other technical, financial and operative aspects

Building a BIS is a process that requires a multidisciplinary approach. From the earliest stages of BIS planning it is necessary to ensure involvement and commitment of an adequate number of IT and nature conservation and biology experts. While IT experts are essential to the process in terms of providing support to all technical aspects of BIS development, from system architecture and administration to planning and management of technical resources, biology experts are necessary in the process of planning BIS content and defining the functionalities and services that the BIS is expected to provide. A BIS is developed to, among other things, support everyday work of employees whose

1 Natura 2000 - http://ec.europa.eu/environment/nature/natura2000/index_en.htm
CBD - https://www.cbd.int
CITES - https://www.cites.org
CDDA - https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-12
Eurobats - http://www.eurobats.org
IPA - http://www.plantlifeipa.org/reports.asp
Ramsar - http://www.ramsar.org
IBA - http://www.birdlife.org/worldwide/programmes/sites-habitats-ibas-and-kbas

work involves nature conservation, so communicating and clearly understanding their business processes (which the BIS has to support) is essential for good system design.

Ensure good cooperation between IT and biodiversity experts

In order to build a BIS that adequately supports business processes, it is necessary to ensure good cooperation between IT and nature conservation and biology experts.

With support from nature conservation and biology experts, IT experts have to make an effort to comprehend the specific requirements and business processes that BIS has to support. This is essential for prevention of oversights and mistakes that can result in an information system that is not able to fulfil its purpose, despite invested efforts and funds.

Ensure good cooperation with the related stakeholders - faculties, museums, institutions, NGOs, etc.

Good cooperation must be established with related stakeholders, such as faculties, museums, ministries, institutes, NGOs and similar, in order to understand their requirements and to ensure trust and mutual support in provision and exchange of data.

Consider external systems' requirements and needs

The purpose of a BIS is not only to support internal processes of institutions that are responsible for its establishment; their purpose is also to provide content and services to various external information systems. For example, firms/institutions that deal with environment and nature impact assessment studies require open access to georeferenced, precise data on species and habitats and other biodiversity data, to include in their studies; Geoportals for physical planning, forestry or agriculture might need access to Natura 2000 borders and nationally designated protected areas to show along with other available data and layers; Rescue service authorities might need precise data on speleological objects, etc. External systems' requirements have to be considered when planning and designing a BIS, so special emphasis must be placed on ensuring easy access to data and services.

Ensure design of common components (species catalogue, multimedia, references ...)

Some content of a BIS is common to many BIS components. This includes a database of references (literature), catalogue of species or habitat types, and similar data. For example, both the habitats database and the Natura 2000 database might need the species catalogue and there is no need to create and maintain a separate species catalogue in both databases. The goal is to ensure that such data is maintained in only one place/database and that it can be easily shared and used by different BIS components.

Implementation phase

Consider implementing open-source technologies

When planning the software platform for the BIS it is important to consider available open-source technologies and solutions, since many such solutions often provide equal possibilities, functionalities and, most importantly, technical support as corresponding commercial software solutions. Switching to open source technologies often requires time and funds for education and training of IT experts responsible for maintenance of the BIS, for them to learn new technologies and software solutions. However, this is a good investment that ensures a substantial reduction in the cost of IS maintenance and licensing. In addition to the initial investment in development of the BIS, IS maintenance requires long-term sustainable financing to cover software, hardware and service costs, so it is necessary to consider any means that will ensure more cost-effective and efficient management of the IS in the long run.

Consider implementing georeferencing rules for data entry

A lot of biodiversity data, such as species or habitats occurrences, is available only in literature and museum collections and is not even available in digital format. Most data available in digital format is not georeferenced and cannot be used to overlay with other related biodiversity spatial data and for other analyses. Furthermore, much of the historic literature data contains information about the locality where some species or habitat type occurred but this data rarely contains precise information about the location, such as its GPS coordinates. Localities are rather just descriptive, broad localities such as areas, towns, toponyms, lakes, rivers or similar. Even though this information is often too general, it is nevertheless very valuable to use in analysis. To transform this valuable data into georeferenced data, georeferencing rules must be applied to generate the coordinates for a certain locality together with the uncertainty/precision of these coordinates expressed in meters.

Ensure good cooperation with the State Geodetic Administration

A lot of biodiversity and nature conservation data requires data and inputs related to cadastral parcels, administrative units, addresses, maps, digital elevation model and the coordinate reference system. For example, since some protected areas or Natura 2000 areas often share a part of their border with, for example, county or state, to be able to delineate such areas there is a need for information about administrative units (e.g. state border, municipality or county borders). Furthermore, there is also a need for topographic maps, digital orthophoto maps and other cartography. All such data is under the responsibility of the national geodetic authority that is responsible for official cartography and cadastre, creation of topographic, cartographic and cadastral databases, and for coordination of the National Spatial Data Infrastructure (NSDI). It is important to ensure good cooperation with the national geodetic authority in order to facilitate access to the necessary geodetic and cartography data.

Management phase

Establish clear protocols for data maintenance

Data is the most valuable component of any information system, so it is important to conduct regular data updates in order to ensure good quality data that can be used for decision-making purposes. This requires clear protocols and instructions covering who is responsible for data maintenance, how often is data maintained, how is data quality ensured, and how the changes are documented (i.e. metadata).

Ensure content administration by biology experts

Data maintained within the scope of a BIS is very versatile and requires maintenance by various biology experts. For example, the part of the BIS that covers species systematics and taxonomy requires content administration by experts for many taxonomic groups. In that regard, species taxonomy and systematics data is particularly challenging to maintain as it requires engagement of experts for flora as well as experts for various fauna taxonomic groups (birds, mammals, amphibians, etc.). If the possibilities for regular content administration by an adequate number of biology experts are limited or absent, institutions responsible for BIS maintenance should consider taking over and connecting to already established international species catalogues, such as the Catalogue of Life and similar, which are regularly maintained by numerous taxonomy experts worldwide.

Ensure an adequate number of IT personnel to technically support IS administration

Apart from content administration, it is mandatory to ensure an adequate number of IT experts to cover various technical aspects of BIS maintenance. This includes

IT personnel that can provide timely technical support as well as IT experts to deal with strategic planning for the BIS, secured long-term financing, software and hardware procurement and maintenance, licensing costs, etc.

It is always important to consider engaging at least two experts who could cover any given field, in order to avoid situations where only one employee knows how to perform some key element of system maintenance and is, as such, irreplaceable. Such situations expose a significant system weakness since there is only one person capable of performing specific tasks. If this person leaves the organization, there is no one else who can take over their tasks and the entire system is jeopardized. In addition to this, and in order to mitigate this risk, it is also important to document business processes to ensure that the knowledge and know-how for specific tasks is available to many employees.

Ensuring continuous long-term financing

Maintenance of any IS requires significant funds in regard to many aspects - from hardware and software purchase, licensing costs, to technical support and expert administration. Information systems are always evolving through development of new components and improvements to existing ones, and all of this requires continuous availability of funds to cover the related costs.

It is often the case that funds for developing BIS components are ensured sporadically and periodically when institutions responsible for BIS establishment receive significant funds through various projects financed by EU pre-accession funds, different national funds, or similar. While such funding is of course very valuable, it is not sustainable in the long run. It is often the case that some IS components are built and established and the necessary IT equipment is purchased, but when time comes to finance mandatory administration and regular maintenance of these components there are no funds available and such situations seriously jeopardize the IS. It must always be considered that an information system is always evolving and as such requires continuous funding. So, while ensuring funds from various sporadic projects is important and valuable, it is far more important to ensure funding from a stable and continuous source such as the state budget or similar.

Further information

Example of different Biodiversity Information Systems

GBIF
http://www.gbif.org

BISE
http://biodiversity.europa.eu

OBIS
http://www.iobis.org

TOPIC 2

TAXONOMIC DATABASES AND SPECIES CATALOGUES

One of the principal aims of current efforts in biodiversity informatics is to foster sharing of electronically available information about organisms from a wide variety of sources. This information was produced at different times and places and with differing aims and is normally earmarked by means of the organism's scientific name. Correct (accepted) names are formed according to the rules of taxonomic nomenclature, without regard to the concept or circumscription of the taxon itself. Potentially, correct names stand for differing concepts (potential taxa), but the names used in scientific communication frequently do not provide a reliable index for biodiversity information. For efficient and reliable use of data in biodiversity databases, an authoritative index (taxonomic list) for information access is essential.

National checklists (national-level species catalogues or inventories) are invaluable resources which serve to co-ordinate, consolidate and disseminate basic taxonomic information for a particular economy and are usually needed by a range of users (not only those from the academic community) for research and biodiversity-related activities. Ideally, national checklists should be integrated, coordinated and disseminated from a single platform compiled by expert taxonomists. Since this is not always possible, non-experts with basic knowledge of biology can also adequately carry out the best part of this task. Expert taxonomists can then perform only a final review of the compiled list.

What is a taxonomic database and/or species catalogue?

A **taxonomic database** is a database created to hold information related to biological taxa - for example groups of organisms organized by species name or other taxonomic identifier, for efficient data management and retrieval of occurrence information as required. Taxonomic databases are routinely used for automated construction of biological checklists, e.g. for flora and fauna, but there are other applications that require such a database, the most important being databases of specimens observed in nature or stored in biological collections. The main goal of a taxonomic database is to accurately model the characteristics of interest for organisms that lie within the scope of intended coverage and usage of the BIS.

Taxonomic databases need to encode conventions from the International Code of Botanical/Animal Nomenclature (botanical code for fungi, algae, bryophytes and higher plants and zoological code for animals and most protists), to model the relevant taxonomic hierarchy for any taxon and to fit this model into the relational model employed in the database systems. In addition to encoding organism identifiers (most frequently a combination of scientific name, author, and - for zoological taxa - year of original publication), a taxonomic database may frequently incorporate additional taxonomic information such as synonyms and other taxonomic opinions, literature sources or citations, plus a range of desired biological attributes for each taxon, such as geographic distribution, ecology, descriptive information, threatened or vulnerable status, etc.

The first successful international initiative designed to provide consistent and reliable information on the taxonomy of biological species started in mid ‘90s as an American partnership of federal agencies together with Canadian and Mexican government agencies. This is the Integrated Taxonomic Information System (ITIS) with a database drawn from a large community of taxonomic experts. A number of other taxonomic databases specializing in particular groups of organisms appeared in the 1970s and continued to the present. They jointly contribute to the Species 2000 project, which is partnering with ITIS since 2001 to produce a combined product, the **Catalogue of Life** (**CoL**). The Global Biodiversity Information Facility (GBIF) constructs a freely-available ‘backbone taxonomy’ that integrates global checklists such as CoL and ITIS with additional sources including new taxonomic treatments, with the aim of organizing species data as efficiently and comprehensively as possible, incorporating synonyms and up to date nomenclature (GBIF Secretariat (2017). GBIF Backbone Taxonomy. Checklist Dataset **https://doi.org/10.15468/39omei accessed via GBIF.org** on 2017-10-16).

While the CoL concentrates on assembling basic name information as a global species checklist, numerous other taxonomic database projects such as **Fauna Europaea**, **Euro+Med** and **Index fungorum** are focused on a particular group of organisms or a particular area. Many taxonomic database projects on the national level were released recently and currently serve national biodiversity infrastructures in several countries. All these databases are based on some of the global/regional databases and fine-tuned for the specificities of a particular economy in terms of endemic species and local interpretation of their taxonomy.

One interesting initiative at the European level for building national biodiversity infrastructures in SEE economies is the **Pan-European Species Directories Infrastructure** (**PESI**). PESI is a European species catalogue (also required by **INSPIRE** directive’s theme - Species distribution) and represents the next step in integrating and securing taxonomically authoritative species name registers that underpin the management of biodiversity in Europe. PESI integrates the three main all-taxon registers in Europe, namely the European Register of Marine Species, Fauna Europaea, and Euro+Med PlantBase, in coordination with EU based nomenclature and the network of EU based Global Species Databases. It is a standards-based, quality-controlled, expert-validated, open-access infrastructure for research, education, and data and resource management.

Why are species catalogues important?

Scientific names are labels for taxa that are governed by formalized rules of nomenclature. These rules were introduced to establish clarity, stability, efficiency and uniqueness to the fragmented landscape of pre-Linnaean nomenclature. Scientific names also serve to label biodiversity information related to species or habitats i.e. biological observations. Biodiversity data about organisms comes from a wide variety of sources and, for preserving consistency in the biodiversity information system, it is essential to overcome all doubts when referring to a particular taxon. Biodiversity information systems are frequently developed at the national level, so it is also very important to embed all specificities of the local floras and faunas into a national biodiversity infrastructure. It sounds simple, but this is a complex task that implies national consensus about a species list living in a particular area and requires support and coordination by national authorities responsible for biodiversity, e.g. the relevant ministry, academy of science and/or other academic institutions.

The technical and organisational aspect of creating, using and maintaining a taxonomic database is a crucial aspect that has to be clarified at the initial phases of planning biodiversity information infrastructure. Establishment and maintenance of the catalogue of species is a significant challenge in both technical and expert/administrative sense. For such systems it is of crucial importance to ensure regular database maintenance and administration by taxonomic experts for many groups of species. Overall, this is a challenging task, both operationally and financially, and there are several options for implementation:

1. Connect (via an open Application System Interface - API) to existing systems such as the CoL, the GBIF Backbone Taxonomy or PESI (which are regularly maintained and administered by many taxonomic experts worldwide);
2. Develop and maintain a custom/own catalogue of species.

Depending on the taxonomic group, one or the other option can be used. Development of a custom catalogue is recommended for groups of organisms that show a high level of local taxonomic diversity, like plants or beetles, while the first solution is better (and more economical) for globally well-known taxonomic groups, such as most vertebrata or butterflies and dragonflies. In case of a custom catalogue, it is necessary to form expert groups for particular taxa composed of representatives from the academic community (national or international institutions) who will work under a clear mandate obtained from a responsible decision-maker at the national level. It is important to reach national (official) consensus on the list of species common at the state level, through involvement of key stakeholders. The agreed list(s) must be made publicly available for browsing and downloading to enhance biodiversity data sharing between data providers on the national and international level.

Using external catalogues such as CoL or PESI ensures compatibility with many other systems, such as GBIF, Encyclopedia of Life (EOL), IUCN Red List of Threatened Species portal and many other systems that are also based on the CoL. If developing own catalogue, it is necessary to provide full compatibility with international catalogues and find a solution for embedding data on local flora and fauna into internationally accepted catalogues. Local, national or thematic species checklists may also be published openly through the GBIF and thus enrich the globally-available species information including vernacular names, geographic distributions and characteristics such as endemism or invasiveness.

Who uses taxonomic databases and species catalogues?

Taxonomic databases (global, such as the CoL, as well as regional/national) are widely used by organisations and individuals worldwide. As a backbone of any biodiversity data infrastructure, these databases are integrated into BISs at all levels by technical staff of a particular IS. At user level, individuals like research scientists, policy and decision makers, citizen scientists and participants of any biodiversity programmes can use this data on daily basis for:

- **Checking** the scientifically accepted name, spelling, alternative names and distribution of a species;
- **Finding** the place of an organism in a consistent and integrated taxonomic hierarchy;
- **Compiling** checklists of species in a particular area or taxonomic group via downloads;
- **Downloading** an electronic list for use in own systems and portals;

- **Providing** an electronic taxonomic backbone for indexing and compiling other information;
- **Carrying** out biodiversity analyses.

A good example of enabling species data to be efficiently managed and made available electronically is the National Biodiversity Network (NBN) in the United Kingdom. The NBN is a collaborative partnership created to exchange biodiversity information between members that include many UK wildlife conservation organisations, the government, national agencies, environmental agencies, local environmental records centres and many voluntary groups. To its members NBN provides two controlled vocabularies (inventories) crucial for enabling species and habitats data to be managed in the shared environment: the UK Species Inventory and the NBN Habitats Dictionary.

The UK Species Inventory is managed by the Natural History Museum, although much of the actual content is created and provided by others, usually currently accepted experts in a taxonomic area. The Inventory also holds earlier checklists for taxonomic groups, derived from, for example, published taxonomic reviews. The Natural History Museum maintains a reporting arrangement on the current status of different taxonomic checklists, which can be referred to via the UK Species Inventory Project website.

The NBN Habitat Dictionary is accessible directly from the NBN website as a reference source and includes information on 16 classifications in use in the UK as a single publicly accessible information resource, and allows a user to compare them and select a classification suitable for his or her purpose. The NBN Habitat Dictionary is accessible directly from the NBN website as a reference source, and a reference should be made to the used classification and website and/or publication source of the classification itself for more detailed information.

How can we know that this data is reliable?

Scientific names of taxa are commonly used for searching, retrieving and integrating information about species, but different users often do not use names unambiguously to match a same taxon. Due to subjectivity, such names, and their underlying taxonomic definitions, are unstable, subject to change, and are frequently ambiguous, thereby creating difficulties in the integration of information from different sources and negatively impacting the utility of names as identifiers and effective indexing tools in biological informatics.

There are a number of reasons that interfere with unequivocal matching of scientific names and taxons in practice. The most common are:

- **Synonyms** - multiple names associated with a single taxon (usually as a result of a change in circumscription that occurs when two formerly distinct taxa are merged).
- **Homonyms** - two identically-spelled names that refer to two distinct taxa.
- **Polysemes** (literally “many meanings”) - a single name that refers to two or more taxa with different but related meanings. (The difference between homonyms and polysemes is subtle and a full explanation lies outside the scope of this guide).

These difficulties can be solved only by a person with experience and good knowledge of the particular taxon. The most frequent method to ensure the necessary degree of reliability and verifiability of names on the taxonomic list is to provide a name (and year) of the last revisor or reference(s) of paper(s) (or name of the database) where the nomenclatural status of every single scientific name was published. A good practice to eliminate ambiguities in scientific names in the taxonomic database is to use Life Science Identifiers (LSIDs).

Linking a particular name on the taxonomic list with its widely accepted interpretation of status and meaning, ensures unambiguous use of names and ensures the required reliability of a biodiversity database. Linking a taxonomic database in the local information system with some of the publicly available databases (like CoL or PESI) is also good practice for ensuring reliability of taxonomic vocabularies.

Why are some of the taxons missing?

The CoL website[2] is a gateway to an online database of the world's ALL known species of animals, plants, fungi and microorganisms, and is freely available to all interested parties. The database is a result of a remarkable global partnership of over 200 expert taxonomic databases world-wide (Global Species Databases - GSD), involving over 3,000 taxonomic specialists that contribute to the content. GSDs are validated for inclusion by independent peer review, ensuring that the best available sources are identified. A few of the datasets contain data only for specific regions, where global coverage has not yet been achieved: these are clearly identified in the data. Completeness of data within individual databases is indicated within the dataset, based upon an assessment by the contributor. Expert teams peer-review the databases and integrate them into a single coherent catalogue, and have established a single hierarchical classification. The CoL is far from complete and covers only 84% of the world's known diversity. It is not complete because digital resources are not yet available for all taxa in the world. The global partnership in building a comprehensive list of all known species of the world is open, and a list of expert associates and databases involved is expanding daily.

2 PESI is a similar resource for European flora, fauna and fungi.

Compiling national checklists

National checklists (economy-level species catalogue or inventory) are invaluable resources which serve to coordinate, consolidate and disseminate basic taxonomic information commonly required by a range of users for research and biodiversity-related activities. The checklist should serve not only taxonomists, but also enable non-taxonomists to readily access (or create their own) information about species occurrences. National checklists should facilitate these activities, including which scientific name to use for any given taxon, without having to understand the scientific intricacies, nuances and debates surrounding the names and classifications.

Ideally, national checklists should be integrated, coordinated and disseminated from a single platform compiled by expert taxonomists, but this is not always possible. Non-experts with knowledge of nomenclature, taxonomy who have access to the relevant literature and online biodiversity data resources such as CoL or GBIF can also adequately carry out most of the activities. However, the final review of the compiled list by expert taxonomists is highly recommended.

National checklists can be published in hard copy or in electronic format on a website. National or economy-wide checklists are generally not published in scientific journals because they are usually very long, but components of national lists such as order or family level checklists may be accepted for publication. The benefit of publishing lists in journals is that they are peer-reviewed and therefore have more credibility, and they can still be incorporated into and distributed via a national checklist website. As a result of additions or changes to checklists, published versions quickly become outdated, but they still have some use as an assessment of the status of a taxon at a particular point in time. Because of the dynamic nature of checklists, it is recommended that a national checklist intended to identify the most up-to-date status of names be provided electronically.

Depending of the available human, organisational and material capacities there are several scenarios for compiling national checklists. Some important things must be kept in mind for each of these scenarios.

Scenario 1: Using taxon experts to compile the checklist

Ideally, a taxonomist with specialist knowledge of the relevant taxon should compile the checklist. Or, review a checklist prepared by non-experts before its dissemination, which is also an acceptable practice.

Scenario 2: Using global checklists to extract a taxon list for a particular area

The CoL may be a useful starting point for compiling a national checklist either in its entirety or for a particular taxon, but it must be recognized that the CoL list is far from complete.

Scenario 3: Extracting a checklist from a specimen database

Extracting a list of species from a global (e.g. GBIF) or a national specimen occurrence database is a good starting point for preparing a preliminary species list, but it should be taken into account that occurrence databases in electronic format for some regions are scarce and highly incomplete.

National checklists can be compiled without great expense. The main costs will be to provide incentive grants to specialist taxonomists, subscriptions to citation and abstract databases, access to publications through journals which might require purchasing PDFs and, possibly the main expense, salaries for checklist compilers where they are contracted specifically to compile the lists and/or for integrating smaller checklists into larger, more comprehensive ones and for formatting checklists submitted by different contributors.

More details about the steps in compiling national checklist can be found in a separate document (see References).

How to use data from online catalogues in an IS?

There is no single point of contact for all scientific names (plants and animals), therefore if looking for a species list, classification or basic species information in a particular group for which no expertise exists, it is recommended to consult several global or regional indexes of species. The most significant are the CoL (for the whole world) or PESI (for European species). Administrators of these indexes provide several options for access and/or use of their data:

1. Download a full (or filtered) list of species

To use species lists in the local environment, the actual list can be downloaded from the CoL. There are two distinct versions of the Catalogue: the Catalogue of Life (monthly edition) and the Annual Checklist. Each year there is a fixed annual edition which is additionally published on a DVD. All editions are available for free use by all individuals, and by arrangement for organisations: **http://www.catalogueoflife.org/content/annual-checklist-archive**.

2. Compare own checklist with a species list at CoL or PESI

If a user wants to check his/her own checklist and compare it with some internationally accepted lists, he/she can use some publicly available tools to do it. The most important are the CoL, PESI and GBIF matching services.

CoL provides a tool for comparing list of species with its dynamic checklist in order to retrieve data such as accepted taxon name and synonyms. The tool is called **List Matching Service**. To check names in the user's species list with the CoL database, it is possible to submit a file of species names in a comma delimited format. The service will compare the names with the CoL database and return a list of correct scientific names. List matching service for CoL is available at this web address: http://www.catalogueoflife.org/listmatching/.

PESI EU Nomen provides a similar tool, the Taxon Match Tool, that can be used for matching any list of species or taxon with PESI in order to get valid names, authorities, PESI classification, GUID, taxon status and other selected outputs. The PESI Taxon Match Tool interface is available at: http://www.eu-nomen.eu/portal/taxamatch.php. For performance reasons, the limit is set to 1,000 rows. For matching larger files, non-marine or multiple data sources, the user can use the Lifewatch Taxonomic Backbone service available at: http://www.lifewatch.be/data-services/.

GBIF also provides a species matching service for matching and comparing a species list with the GBIF backbone. This tool is available at: https://www.gbif.org/tools/species-lookup.

3. Direct linking to records from CoL or PESI databases

If you simply want to link species records from your database with a record for that species in the CoL or PESI database, just use the following link(s):

- http://www.catalogueoflife.org/col/details/species/id/[LSID]
- http://www.eu-nomen.eu/portal/taxon.php?GUID=[GUID]

and replace a particular [LSID] or [GUID] by a corresponding identifier of a particular taxon from the CoL or PESI database.

You can look up particular LSIDs (or GUIDs) via the search interface at CoL and PESI websites.

- http://www.catalogueoflife.org/col/search/scientific or
- http://www.eu-nomen.eu/portal/search.php

4. Embed data from international lists using web services

Catalogue of Life Web Service

CoL provides means to connect and retrieve data via web services. Data can be searched by name or ID and can be retrieved in several structured formats (json, xml, PHP array).

This link provides a full web service specification and capabilities:

- http://webservice.catalogueoflife.org/col/webservice

Here is an example how to search data by name and retrieve all available data using an example of *Vipera ammodytes*:

- http://webservice.catalogueoflife.org/col/webservice? name=Vipera+Ammodytes

This example call returns 15 results of all taxonomic ranks (species and infraspecies, in this example) that contain *Vipera Ammodytes* in their name.

Each result can be an accepted (infra)species name, an (infra)species synonym, a common name for an (infra) species, or a higher taxon.

PESI EU Nomen Web Service

PESI EU Nomen also provides web service to retrieve PESI taxonomy data. Web service can be used to get the GUID for a taxon, check the spelling of a taxa, find the authority for a taxa, resolve an unaccepted name to an accepted one, and much more.

Available methods of accessing the service include REST and SOAP services.

This link provides the full web service specification and capabilities: **http://www.eu-nomen.eu/portal/webservices.php**.

This is an example how to retrieve data by specifying the GUID:

i.e. the GUID for *Vipera ammodytes* is urn:lsid:faunaeur.org:taxname:214762: **http://www.eu-nomen.eu/portal/taxon.php?GUID=urn:lsid:faunaeur.org:taxname:214762**.

GUID can be looked up via the search interface **http://www.eu-nomen.eu/portal/search.php**.

GBIF Species API

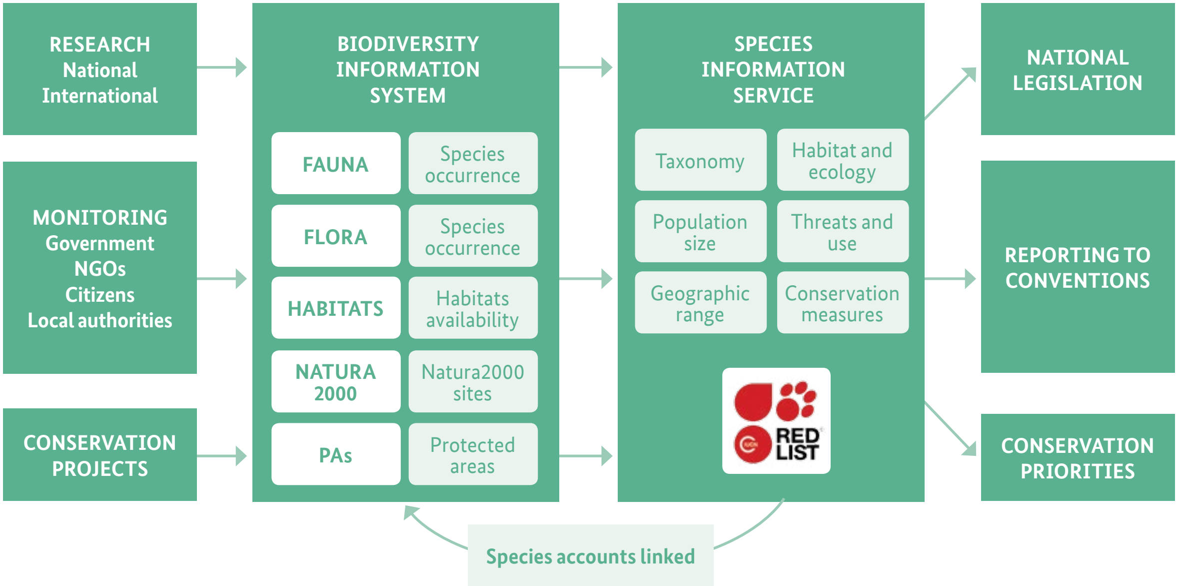
There are other additional species taxonomy web services available, such as the GBIF Species API (**https://www.gbif.org/developer/species**) which works against data kept in the GBIF Checklist Bank that taxonomically indexes all registered checklist datasets in the GBIF network. Full checklist can be found here **https://www.gbif.org/dataset/search?type=CHECKLIST**.

Red Lists in relation to the BIS

Lists of protected species are a core tool in species conservation legislation across all geographic scales. Protected species are species that have been prioritized for conservation, based on the threats they face. Red Lists provide such information and are recognized as an important tool to help prioritize species for conservation. Note that a Red List by itself is not a priority list for conservation action. The information in the Red List constitutes only one part of broader information necessary for prioritization. Other factors include endemism, phylogenetic rarity, practical concerns about the size of the effect that conservation measures may have, overall costs involved in conserving a particular species, etc. In many economies, national legislation either requires the establishment of a national Red List (e.g. in Macedonia), and/or lists of protected species that partially rely on a national Red List. Knowledge about the conservation status of species inside the economy, and conservation actions undertaken, is also necessary for reporting under international conventions (e.g. Convention on Biological Diversity, Bern Convention) and EU directives (e.g. Birds and Habitats Directives). At the national and global level, the aim is to provide scientific information, to convey the urgency of conservation issues to the public and policy makers, and to guide national and global priorities for conservation action.

Information stored in a national BIS can form a starting point for Red List assessments, either for the creation of a national Red List, or as a contribution to a region-wide (e.g. SEE) or even the global IUCN Red List. In particular, one of the first steps in determining the conservation status of a species is to gather historic and current information about its occurrence and population size in the economy. Occurrence localities and associated information stored inside the BIS can easily be compiled for the species of interest over the relevant area and timeframe in order to get this core data together for a Red List assessment. Further steps include gathering information about the species' taxonomy, habitat and ecology, threats and current conservation measures. Again, information about taxonomy, protected areas and international conventions stored in the BIS can be used to inform conservation assessments (see Figure 2). The compiled and summarized information is stored in a database (such as the IUCN Species Information Service) and serves to document the Red List assessments. Ideally, this database is publicly available and accessible online, providing easy access to any particular species' assessment and the rationale behind its conservation status.

▼ **Figure 2:** *Potential data flow between a Biodiversity Information System and a Species Information System containing Red List assessment information*



But information exchange can also happen in the other direction. When data (e.g. occurrences) for a particular species is entered in the BIS, the BIS could provide a link from the BIS species' account to an available Red List assessment for the species, such that users who are entering data have immediate access to its conservation status and data on habitat, ecology, threats and conservation actions. For example, links to the global IUCN Red List are provided for all species in the GBIF database (see Figure 3). An analogous set-up could be used on a national level, provided that the Red List assessments are stored in a publicly accessible database.

▲ **Figure 3:** *Highlighting the link between a species account in GBIF and in the IUCN Red List of Threatened Species (red outline)*

Economies interested in compiling a national Red List should consult the IUCN Red List of Threatened Species website (**www.iucnredlist.org**) for general information about the Red List Categories and Criteria, as well as the **National Red List** website (**www.nationalredlist.org**) for more specific information on the national red-listing process.

Further information

IUCN ECARO
https://www.iucn.org/regions/eastern-europe-and-central-asia

National Red List
www.nationalredlist.org

CBD
https://www.cbd.int/convention/

Bern Convention
https://www.coe.int/en/web/bern-convention

EU Birds Directive
http://ec.europa.eu/environment/nature/legislation/birdsdirective/index_en.htm

EU Habitats Directive
http://ec.europa.eu/environment/nature/legislation/habitatsdirective/index_en.htm

Fauna Europaea - All European Animal Species on-line
https://fauna-eu.org/

Index Fungorum
http://www.indexfungorum.org/

Euro+Med PlantBase
http://www.emplantbase.org/home.html

AlgaeBase
http://www.algaebase.org/

Pan-European Species Directories Infrastructure
http://www.eu-nomen.eu/

Species 2000
http://www.sp2000.org/

Catalogue of Life
http://www.catalogueoflife.org/content/tools

GBIF Global Biodiversity Information Facility
https://www.gbif.org/

National Biodiversity Website
https://nbn.org.uk/

UK Species Inventory
http://www.nhm.ac.uk/our-science/data/uk-species.html

TOPIC 3

BIODIVERSITY DATA STANDARDS

Data about living beings on Earth can be collected and stored in various forms and types. Data about taxonomy, species-by-occurrence (or sample-based data), geotagged images, or specific biological data like phylogenies, gene sequence, protein sequence, genomics data, etc. are generally called "*biodiversity data*", but the methods used for collection, structuring, storage in digital form and usage of such data are very different. Biodiversity data, information and knowledge are held by many different kinds of people (academics, managers, amateurs), in different capacities and for different purposes. There are many existing initiatives concerned with removal of access barriers to data, information and knowledge, and it is important that any new initiatives for organised biodiversity data management ensure coordination and collaboration with existing activities, supporting and augmenting what is already being done.

Data fundamental to biodiversity research, conservation, reporting, natural resource management and conservation policy development is the data on species (and habitats) distribution. Standards for international collaboration between biological database projects oriented towards **exchange** of biological / biodiversity data (such as Darwin Core and Access to Biological Collections Data - ABCD schema), although continuously improving, are already mature and widely used.

The Global Biodiversity Information Facility (GBIF) was established in 2001 with the explicit objective of enabling free and open online access to occurrence data, and is operating very actively in this area ever since. All relevant standards for biodiversity data management are implemented in their system, which is why it is considered the best model and the starting point in planning biodiversity information systems.

Types of biodiversity data

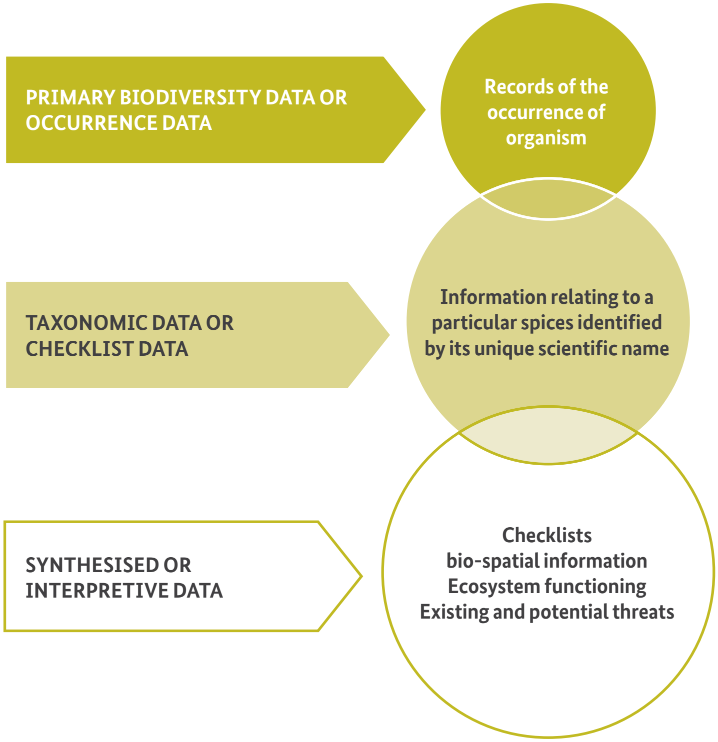
There are several categories of biodiversity data, i.e. levels at which data can be gathered and used, and it is important to distinguish between these levels. Biodiversity data can relate to a particular area or group of living organisms; it may store specimen-level information, species-level information, ecosystem-level information, information on nomenclature, or any combination thereof. The main distinction to be drawn is between primary biodiversity data (species occurrence data), taxonomic data (checklists and information about the identity of organisms), and synthesised or interpretive (secondary) data (indicators) (see Figure 4).

Primary (raw) biodiversity data (or occurrences) is the data from observation (in the field or of a vouchered/labelled specimen in a collection) of a taxon (or ecosystem/biological community) at a particular place on a specified date (possibly enriched with other attributes of the collecting/sampling event, such as collector name, number of specimens, etc).

Checklists are lists of scientific names of organisms grouped into taxonomic hierarchies that are common in a particular area.

Synthesized or interpretative data (or indicators) are statistical descriptions of biodiversity which help

▶ **Figure 4:** *Core categories of biodiversity data*
Source: GBIF, 2011. Best Practice Guide for publishing primary biodiversity data



scientists, managers and politicians understand the state of biodiversity and the factors that affect it. Indicators are usually a result of some form of processing, like grouping, categorising, pulling or mathematical transformations of primary biodiversity data.

Other types of biodiversity data important for reaching better conclusions and making the right decisions include metadata and environmental data that can influence the living world. *Metadata* are structured descriptions of other datasets i.e. the descriptive information that accompanies a dataset – they are the data about data. *Supporting biodiversity data* is data that is not directly related to biodiversity but is useful for better understanding of biological patterns, it explains biological processes and is a valuable resource for monitoring the state of nature. This data is usually collected and maintained by institutions entrusted with management of natural resources and is available in the form of orthophoto or satellite images, land use maps, geological maps, meteorological data, etc.

What are biodiversity data standards (TDWG)?

Biodiversity Information Standards (TDWG), also known as the **Taxonomic Databases Working Group**, is a non-profit scientific and educational association affiliated with the International Union of Biological Sciences. TDWG was formed to establish international collaboration between biological database projects by promoting wider and more effective dissemination of information about the world's heritage of biological organisms. TDWG now focuses on the development of standards for **exchange** of biological/biodiversity data. TDWG works in task groups focusing on various aspects of structuring and maintenance of different types of biodiversity data, such as: biological descriptions, collections of similar data, collections and/or field observations data, geospatial data, globally unique identifiers, multimedia illustrations of organisms, phylogenetics, providing biodiversity data to the world, and the system for labelling organisms. In the past decades, TDWG task groups developed a vast body of standards that facilitate structuring and sharing of biodiversity data for individuals, organisations, governmental and international bodies. The most important are:

- **Darwin Core** (DwC) is body of standards that includes a glossary of terms used in attributing the *occurrence of taxa in nature*, intended to facilitate sharing of information about biological diversity.
- **Access to Biological Collections Data** (ABCD) schema is an evolving comprehensive standard for access to and exchange of *data about specimens and observations*. ABCD Schema strives to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated. Versions 1.2 and 2.06 are currently in use with the GBIF and BioCASe networks. A separate standard for storing DNA sequence data (ABCDDNA) was developed as an extension to ABCD.
- **Natural-language descriptions of a taxon** (or of an individual specimen, occasionally) are semi-structured, semi-formalised texts that may be simple, short and written in plain language (if used for a popular field guide), or long, highly formal and using specialised terminology when used in a taxonomic monograph or other treatment. The goal of the standard is to allow capture, transport, caching and archiving of *descriptive data commonly used in scientific communication*, using a platform- and application-independent international standard.
- The **Audubon Core** (Biological Multimedia Metadata Standard) is a set of vocabularies designed to represent metadata for biodiversity *multimedia resources and collections*. These vocabularies aim to represent information that will help to determine whether a particular resource or collection will be fit for some particular biodiversity science application before acquiring the media.
- **TAPIR**, a computer protocol for discovery, search and retrieval of distributed data over the Internet, was designed to enhance and standardise technical aspects of *data flow between different biodiversity data providers*. TAPIR consists of a specification that determines how client applications seeking information should communicate with server applications hosting data, i.e. how distributed data providers with different database systems and different data structures but with the same type of content, could share biodiversity data[3].
- In addition to these standards, a large number of standards have been developed for specific purposes, such as the **Taxonomic Concept Transfer Schema**, **Natural Collections Descriptions** - NCD (data standard for exchanging data that describes natural history collections), as well as several standards developed especially for botanical data.

Technical specifications and updated documentation for all TDWG standards are available at: **http://www.tdwg.org/standards/**.

3 Consider a group of natural history museums or scientific institutions that want to make their biological specimen data available from a single web search page. Each organization has its own database in a different computer environment and stored in different formats. Their common requirement is for all searches via a single web page to access all databases and return a combined and integrated response to the searcher.

Collections and/or observations data

In the process of global integration of primary biodiversity data, it was necessary to standardize the terminology used to model information about biological collections and to provide a general format for exchange and retrieval of data on biological collections. The most used are **Darwin core** (DwC) and **Access to Biological Collections Data** (ABCD schema) - a comprehensive and commented schema for biological collection records.

Darwin core

Darwin Core is the most widely used standard that facilitates exchange of information about the geographic location of organisms and associated collection specimens. It includes a glossary of terms intended to facilitate sharing of information about biological diversity by providing standard reference terms that include definitions, examples, and commentaries. In other contexts these terms might be called properties, elements, fields, columns, attributes, or concepts that describe taxa and their occurrence in nature, as documented by observations, specimens, samples, and related information.

A ratified metadata standard was officially released on 9 October 2009 and is maintained by a TDWG task group. At this moment it represents a glossary of almost 600(!) terms that describe different aspects that can be used to clarify field occurrences of individuals of a particular taxon in nature.

Basic occurrence data

- Basic collecting/observation event information describes **what, who**, **where**, **when** and **how** data about the particular taxa was recorded in the field

Additional occurrence data

- Details about specimen identification and taxonomic classification
- Locality details
- Biological data about the collected or observed specimen(s)
- Record-level information
  - Data cataloguing
  - Data on processing (e.g. preparation) and storage of specimen(s)
  - Authorship and data usage
- References such as images and or citations
- Verbatim original data (data published in scientific articles and/or catalogues).

The *Simple Darwin Core* is simple in that it assumes (and allows) no structure beyond the concept of rows and columns, which might be thought of as attributes and their values, or fields and records. The words *field* and *record* are used to refer to the two dimensions of the *Simple Darwin Core* structure. Think of term names as field (Excel column) names. In other words, a *Simple Darwin Core* record could be captured in a spreadsheet or in a single database table.

Darwin Core Archive

Darwin Core Archive (DwC-A) is a biodiversity informatics data standard that makes use of Darwin Core terms to produce a single, self-contained dataset for species occurrence or checklist data. Essentially it is a set of text (CSV) files with a simple descriptor (meta.xml) to inform others how your files are organized. The central idea of an archive is that its data files are logically arranged in a star-like manner, with one core data file surrounded by any number of 'extension' data files. Core and extension files contain data records, one per line. Each extension record (or 'extension file row') points to a record in the core file; in this way, many extension records can exist for each single core record.

Enhancements of the DwC exchange standard

Darwin Core was developed for exchange of information about occurrence of organisms in nature and only recently the standard was broadened by new terms for detailed reporting on other aspects of inventories such as biomass and the number of individuals as proxies of abundance. To enable data holders to share structured survey data, such as population time series or presence–absence data capture of sample effort, it now includes reporting and identifiers relating the parent–child relationships between events they represent. However, standard details about the sampling method and effort are not implemented in these Darwin 'Event' Core fields and therefore other unique standards that meet this requirement were developed. The most important are the **Humboldt Core** (a community-developed standard for representing critical information about scope, method and completeness of biological inventories) and **Ecological Metadata Language** (EML) - a set of controlled vocabularies (i.e. predefined, authorized terms) that specifically target long-term observation data to describe the dataset's spatial, temporal and taxonomic coverage.

Identify biodiversity data globally

Our world is numbered. Books have ISBNs and products have barcodes. Cars have VINs, even people have social security numbers. Numbers help us reference items unambiguously. The concept of **Universal Unique Identifier** (UUID) or **General Unique Identifier** (GUID) is designed with the intention to uniquely identify some object or entity on the Internet by assigning a specific number to it. UUID is a large, enormous number (128-bit) generated according to certain rules that are almost guaranteed to be unique. In a similar manner, **Life Science Identifiers** (LSID) are a way to name and locate pieces of biologically significant resources including species names, concepts, occurrences, genes or proteins, or data objects that encode information about them. Essentially, a LSID is a unique, persistent and location-independent identifier for some types of data, and the LSID protocol specifies a standard way to locate the data (as well as a standard way of describing that data). They are like **Digital object identifiers** (DOI) used by many publishers. To put it simply, LSIDs are a way to identify and locate pieces of biological information on the web. The importance of unique identification of biodiversity data accessible on the web is obvious.

Collecting, maintaining and storing biodiversity data cannot and does not need to be centralized. Governmental institutions, scientific institutes, museums or NGOs (and even individual scientists) organise collection, structuring, use and access to their data in accordance with their business policy and intentions, but, for clear regional or global insight into biodiversity processes and patterns, processing of all relevant data is very important. Integration or pooled processing of data from different sources is possible only if the meaning of base vocabularies (like referenced species and/or localities) is identical and unambiguous. The concept of the LSID was developed to facilitate this requirement.

Biodiversity data protocols

Huge amounts of new primary biodiversity data are being collected, and the digitisation of existing data from field surveys and museum collections is advancing in many countries. However, data resources remain fragmented, isolated by the software used to store it and by the ability of providers to make their information available online in a suitable format. Widely distributed data can only be integrated and delivered to a larger audience through a system of integrated data networks. To build such networks, that enable search and retrieval of data across multiple providers, some common protocols for data flow were developed. The most used are:

- Distributed Generic Information Retrieval (DiGIR) – an open source communication protocol for accessing distributed biodiversity databases via the Internet. Used primarily in North America, it works with the DwC exchange format.
  http://digir.sourceforge.net/

- Biological Collection Access Service (BioCASE) Protocol – an open source communication protocol for accessing distributed collection and observational databases via the Internet. Primarily used in Europe and works with ABCD.
  http://www.biocase.org/index.shtml
- TDWG Access Protocol for Information Retrieval (TAPIR) – an open source communication protocol for distributed queries of heterogeneous biodiversity databases. It was created as an integration of the DiGIR and BioCASE protocols, serving as an international standard.
  http://wiki.tdwg.org/TAPIR/

Why is it important to use TDWG biodiversity standards?

Data about the world's biodiversity is complex. For centuries, scientists and researchers studying the natural world have recorded a wealth of information about the organisms they observe or collect. Public and private institutions around the world manage this information and, profiting from the same technological advances that enable scientists and amateurs alike to contribute to our body of scientific evidence, it's possible to establish stronger connections between all this data. It is important to encourage data holders to publish the richest data possible to ensure its use across a wider range of research approaches and questions. Not every dataset includes information at the same level of detail, but sharing what is available is valuable, because even partial information can answer some important questions. Standardizing all their sources and formats is an overwhelming task.

Digital primary biodiversity data could be stored in different formats but for any kind of processing it must be structured. Spreadsheets (or tables) are the easiest and most frequently used form of structuring biodiversity data, where a single row represents a particular occurrence and the columns represent attributes of that occurrence, such as the taxon, locality, observation date or name of the observer. Spreadsheets are a good solution for structuring biodiversity data and the tools for manipulation of data organised in tables (like Excel) are very sophisticated nowadays. However, when faced with larger amounts of data or attributes these programs face significant degradation of performance and difficulties in data filtering. In such cases the data has to be transferred to a database management system. Modern database management systems have a well-developed system which avoids data redundancy by storing unique data (e.g. taxon names) in separate tables and referencing this data in other tables via numeric identification keys. The system of interconnected tables is known as a relational database and it significantly increases the amount of data that can be manipulated without reducing performance. Modern database management systems are powerful tools for storing and using large amounts of data, on local computers as well as online on a computer network or even the Internet. Most of them can be downloaded free of charge from the Internet.

Who uses TDWG standards?

Biodiversity data structured in a standard form has a diverse community of users comprising organisations and individuals around the world. The most important are:

- **Environmental authorities and agencies** in charge of reporting and monitoring;
- **Research scientists** in academia, institutes, industry and government;
- **Policy- and decision-makers** in governments and international organisations;
- **Citizen scientists** who are exploring biodiversity and are active in the education community.

They are usually engaged or affiliated with organisations or initiatives dealing with nature conservation:

- Biodiversity data holders (e.g. academic institutions, museums, libraries, herbaria, botanical gardens, government agencies, researchers, students, industry);
- International or national initiatives for archiving/compiling/managing biodiversity data;
- Biodiversity network developers (e.g. BioCASE, GBIF, IABIN);
- Consumers of biodiversity data;
- Developers of Collections Management Systems (e.g. Specify, Emu, Symbiota);
- GBIF Nodes;
- TDWG sub-groups and members.

Anyone planning to work or already working with biodiversity data who wants to be able to integrate their data with other providers' data for regional or global analyses, must match their data to the existing standards.

Practical aspects of DwC usage

Digitising biodiversity data

An important step to standardise large data sets and to make them available for analyses is to store them in a proper format that enables management and processing. Using a convenient desktop or web application, data can be entered by filing appropriate forms and the application will store the data for later use. This method is suitable for small amounts of data. To enter a larger amount of data it is better to fill tables (where the columns are standard attributes of an occurrence) and import data automatically from a table into a database. Digitisation of new biodiversity data is also possible directly in the field by recording data on a mobile device (by means of a data entry application) and storing it in a standard format. By exporting the recorded data from the mobile application and importing it into a desktop (or web) application, the digitization process can be significantly accelerated.

No matter which method of data entry is used, it is essential that the entered data is stored in a single (or interconnected) table(s) structured according to some accepted standard format (DwC or ABCD). When designing an information system, the book of standards must be defined, in the form of a list of attributes with mapping to the existing standards, at the very beginning. Based on this book of standards, it will be easy to define templates for data entry (i.e. Excel tables) and avoid mistakes and misleading interpretations of data which originates from different sources.

For efficient data filtering, where possible, it is always better to use sets of predefined values for particular attributes. With this strategy, the user that enters data can pick an appropriate value from a drop-down list. This will harmonise data from different sources, reduce the possibility of error and facilitate searching and locating data in a database. Not all attributes are suitable for configuration as a drop-down list. As a rule, only attributes with a limited number of values should be defined as pick lists. Some of DwC terms (attributes of an occurrence) that are highly recommended to be formatted as a pick list in entry forms are: *accessRights, basisOfRecord, taxonRank, identificationVerificationStatus, georeferenceProtocol, georeferenceVerificationStatus etc...*[4] However, attributes with a limited number of possibilities (e.g. *lifeStage, sex, preparation, identificationQualifier or even identifiedBy or rightsHolder*) are better formatted as a pick list to limit the user's possibility of submitting free-form data. For obvious reasons, use of pick lists is mandatory for identification (DwC term scientificName).

4 For explanation of DwC terms/attributes see Appendix 1 at the end of this document.

Biodiversity data publication

'Publishing' biodiversity data is defined as making biodiversity datasets publicly accessible in a standardized format via an online access point (typically a web address). Efficient access to biodiversity data is essential for a responsible approach to environmental management and planning (through regulation of land-use, planning, decision-making, and supporting implementation of global, national and sub-national biodiversity-related policies and strategies). Providers of biodiversity assets need educated personnel and proper infrastructure to organise harvesting, storage and access to their data in line with the needs of potential users. This infrastructure does not apply only to computers and other hardware but also to specialised software solutions for efficient management of biodiversity data. Development of a software for biodiversity data management can be an extremely complex and costly task. Fortunately, there are also free solutions in the form of fully functional platforms with excellent performance and genuine implementations of accepted biodiversity standards and protocols. The most powerful are **Specify**, **Symbiota**, **BEXIS2** software, and the **Integrated Publishing Toolkit** from GBIF.

Specify software

The **Specify software** is a database platform that manages species and specimen data for biological research collections. It records collection/observation attributes, tracks specimen transactions, links images to specimen records and publishes data to the Internet. It consists of several components: Specify 6 - a robust yet highly customisable software for Windows, MacOS, and Linux computers written in Java, the Specify 7 web platform - almost identical implementation of Specify 6 that operates within a web browser and the Specify Web Portal – which provides collection data, maps and images to public web users.

An interesting project for development and launching of a national BIS is the *DIgital Information system for NAtural history data* - DINA project. The DINA project is a joint initiative of the DINA consortium (founded in 2014 by six natural history collection institutions in Europe and North America, open to additional members) for developing an open-source web-based information management system for natural history data. At the core of the system is support for assembling, managing and sharing data associated with natural history collections and their curation. It is largely based on Specify 6/7 but also includes additional components developed within the DINA consortium.

Symbiota Software Project

The **Symbiota Software Project** is a constantly evolving library of web tools that aids biologists in establishing specimen-based virtual floras and faunas. This is a full-feature platform for creating voucher-based biodiversity information portals and communities. Originally conceived to promote small- to medium-sized, regionally and/or taxonomically themed collaborations of natural history collections, over time these taxonomically diverse portals have grown into an important resource for mobilization, integration and use of specimen- and observation-based occurrence records in North America and beyond. Symbiota is exclusively web-based and emulates functionality of a modern **Content Management System** (**CMS**), providing highly sophisticated yet intuitive user interfaces for data entry, batch processing, editing and visualisation of occurrence-based biodiversity data.

BEXIS2 software

BEXIS 2 is a free and open-source software supporting researchers in managing their data throughout the entire data life cycle, from data collection, documentation, processing, analysis, to sharing and publishing of research data. This is a modular scalable platform suitable for working groups and collaborative project consortia with up to several hundred researchers. It was designed to meet the requirements of researchers in the field of biodiversity but is generic enough to serve other communities as well.

Integrated Publishing Toolkit

GBIF is a global research infrastructure for biodiversity data. It integrates data by focusing on specific elements that tie all this varied and variable information together: evidence that a verified source found a specific organism at a specific time and place. By using GBIF tools and web services, it enables people to publish, discover and retrieve thousands of datasets containing hundreds of millions of species occurrences.

GBIF focuses on biodiversity data commonly used for biodiversity research, natural resource management and conservation policy and is therefore focused on species checklists, species occurrence data and metadata on biodiversity datasets. The term *dataset* in GBIF is used as a digital collection of logically connected facts (observations, descriptions, or measurements), typically structured in tabular form as a set of records, with each record comprising a set of fields, and recorded in one or more computer data files that together comprise a data package. In the domain of biodiversity, a dataset can be any discrete collection of data underlying a paper – e.g. a list of all species occurrences published in the paper, data tables from which a graph or map is produced, digital images or videos that are the basis for conclusions, an appendix with morphological measurements, or ecological observations. DwC was implemented for storing attributes of an occurrence and the preferred format for publishing data to the GBIF network, as a single, self-contained dataset for species occurrence or checklist data, is the **Darwin Core Archive - DwC-A**.

To facilitate efficient publishing of biodiversity data on the Internet, the GBIF Secretariat prepared an out-of-the-box set of tools as well as detailed instructions. The **Integrated Publishing Toolkit** (IPT) is a free open-source software tool written in Java, used for publishing and sharing biodiversity datasets through the GBIF network. Designed for interoperability, it enables publishing of content in databases, Microsoft Excel spreadsheets and text files using open standards. You can also use a 'one-click' service to convert your metadata into a draft data paper manuscript for submission to a peer-reviewed journal[5]. Core development of IPT takes place at the GBIF Secretariat, but the coding, documentation, and internationalization are a community effort and everyone is welcome to join in.

Development of this tool was guided by the need for a simple, general publishing solution that was platform independent, could be easily managed by institutions, and leveraged existing metadata tools and standards. The solution is a simple web-based publishing toolkit deployed as a Java application that can be downloaded from the GBIF site. IPT supports simple publisher workflows and includes the following features and steps that publishers need to complete:

1. Support for multiple users with distinct permissions for software administration and management of the resources it hosts.
2. Upload source data as a delimited text file or connect to a database.
3. Map the terms (*e.g.*, fields or headers in a database or spreadsheet) from the source dataset to the terms in the DwC.
4. Enter the metadata for the dataset, specifying the scope, methodology, ownership, rights, etc.
5. Produce a DwC-A and a publicly accessible web page that shows the metadata and links to the archive and other documents that were created.

5 For details about publishing metadata about primary biodiversity dataset in the form of a peer-reviewed article in a scholarly journal, see: http://www.gbif.org/publishing-data/data-papers.

6. Register datasets with the GBIF registry so they are discoverable and can be harvested for indexing by GBIF and others.

Other initiatives

PlutoF - a platform for storing and managing biodiversity data over the web by providing database and computing services for the taxonomical, ecological, phylogenetical, etc. research.

iDigBio (Integrated Digitized Biocollections) - a program of the National Resource for Advancing Digitization of Biodiversity Collections (**ADBC**) for digitization of biodiversity specimens by converting physical objects to high quality digital images, associated descriptive text to electronic records, analogue sound and motion recordings to digital representations.

EU BON (European Biodiversity Observation Network) - a project to build a substantial part of the Group on Earth Observation's Biodiversity Observation Network (**GEO BON**) by (inter alia) improving the range and quality of the methods and tools for assessment, analysis, and visualization of biodiversity and ecosystem information. In the framework of this project they launched the EU BON Portal, an online platform that facilitates delivery of relevant biodiversity information and integrates biodiversity observation systems under one search facility.

Publishing biodiversity data through the Global Biodiversity Information Facility (GBIF)

The Global Biodiversity Information Facility (**https://www.gbif.org**) is an open-data research infrastructure, funded by governments, aimed at providing anyone anywhere with access to data about all types of life on Earth. Coordinated by a Secretariat in Copenhagen, Denmark, GBIF enables data-holding institutions around the world to share information about where and when species have been recorded. This knowledge derives from many sources, including museum specimens dating back decades or centuries, current research data and monitoring programmes, as well as volunteer recording networks and citizen science initiatives.

By encouraging use of common data standards and open-source publishing tools, GBIF enables data from thousands of different collections and projects to be integrated, discovered and used to support research and policy. Data published through GBIF can be freely accessed at the global level via GBIF.org and associated web services, as well as through national and thematic portals making use of the shared infrastructure.

Through its network of national, regional and thematic nodes (see: **https://www.gbif.org/the-gbif-network**), GBIF also acts as a collaborative community of practice, sharing skills and best practices to encourage the widest possible participation.

Benefits of publishing data through GBIF

By sharing data using GBIF-compatible tools, researchers and institutions in the SEE can:

- Add value to the data by enabling its re-use across a wide range of research fields;
- Fill geographic, taxonomic and temporal data gaps, thus advancing biodiversity knowledge both within the region and beyond;
- Provide visibility for natural history collections and research projects, including individuals involved at all levels, e.g. field collection, identification, curation and data management;
- Track the uses and applications of data through citation in research and metrics on data downloads;
- Meet obligations concerning data management and access increasingly required by funding agencies and public authorities.

At the national level, participation in the network helps countries to meet obligations under the CBD, as mobilization of species occurrence records through GBIF is a recognized indicator of progress towards Aichi Biodiversity Target 19 on sharing of knowledge (see **https://www.bipindicators.net/indicators/growth-in-species-occurrence-records-accessible-through-gbif**).

Types of data shared through GBIF

GBIF supports four classes of datasets, listed here in progressive order of richness:

- **Resource metadata**, providing structured information about data included in a collection or project, even if the data records themselves are not yet digitized or organized in standard formats. This can alert users to existence of the data and provide links or contact information to enable access.
- **Checklist data**, in the form of a catalogue or list of named organisms, or taxa. These may include additional details such as local species names or specimen citations, and categorize information along taxonomic, geographic and thematic lines. Examples could include 'Endemic vertebrates of South-East Europe' or 'Invasive alien species in Macedonia'.
- **Occurrence-only data**, providing evidence of the location of individual organisms in time and space. This may derive from georeferenced specimens, time-stamped images, literature references, field observations or other sources. Data standardized in this format is invaluable in building models of species distributions and ecological niches to help assess impacts of climate change, risks of invasion and conservation priorities, among many other uses.
- **Sampling-event data**, providing additional detail to the classes above, by linking occurrence of species to particular monitoring or sampling activities. This will include information such as the methodologies or protocols used to collect or observe species, and the relative abundance of different species in the same sample. Such datasets enable direct comparison between samples using the same protocol at different times or locations, thus inferring trends over time or impacts of human activities.

Details of each of these dataset types, including templates for organizing them in appropriate formats, are provided at: **https://www.gbif.org/dataset-classes** .

Standards and tools

The use of the Darwin Core (DwC) and Darwin Core Archives (DwC-A) are highly recommended in publishing data to GBIF. The GBIF Integrated Publishing Toolkit (IPT) is the preferred method of storing and registering datasets. Details of each of these are provided elsewhere in these guidelines.

Becoming a GBIF data publisher

Before sharing datasets through GBIF, institutions must register as a GBIF Data Publisher. This ensures that all data is correctly attributed to the source institution, and provides users with additional information on the provenance of data to help establish its fitness for use. Registration is through a standard form available at **https://www.gbif.org/become-a-publisher**. This will include endorsement of the publisher by an existing GBIF node. In cases where an institution's economy is not yet participating in GBIF, the Secretariat will provide alternative endorsement options through the wider community of nodes.

Citation and visibility

Datasets published through GBIF are assigned a dedicated web page including all the information given in the metadata associated with each dataset. This comprises, for example, the full set of individuals involved in creating and maintaining the dataset, a map view of the georeferenced data, institutional logo and contact details, information on the taxonomic, geographical and temporal coverage of the data, as well as methodologies used in data collection. For an example of a dataset page, see **https://www.gbif.org/dataset/cb6e66f1-3056-404d-a341-bb856762c57c** .

Published datasets are automatically assigned a Digital Object Identifier (DOI) for use in the standard citation provided to users of GBIF.org. In addition, all data downloads delivered to users are also assigned a unique DOI for use in citation. This enables all datasets contributing to each download to be individually credited, and GBIF dataset pages include a 'citation' button that enables data publishers to view examples of research to which their data has contributed.

Further information

TDWG Standards
http://www.tdwg.org/standards/

Humboldt Core
https://mapoflife.github.io/humboldtcore/

Specify software
http://specifyx.specifysoftware.org/

DINA project
https://www.dina-project.net/wiki/Welcome_to_the_DINA_project!

Symbiota
http://symbiota.org/docs/

BEXIS2 software
http://bexis2.uni-jena.de/

GBIF Integrated Publishing Toolkit
https://www.gbif.org/ipt

Annual Science Review summarising research enabled by data accessed through GBIF
https://www.gbif.org/science-review

Quick guide to publishing data through GBIF
https://www.gbif.org/publishing-data

How to join GBIF as a national or organizational participant
https://www.gbif.org/become-member

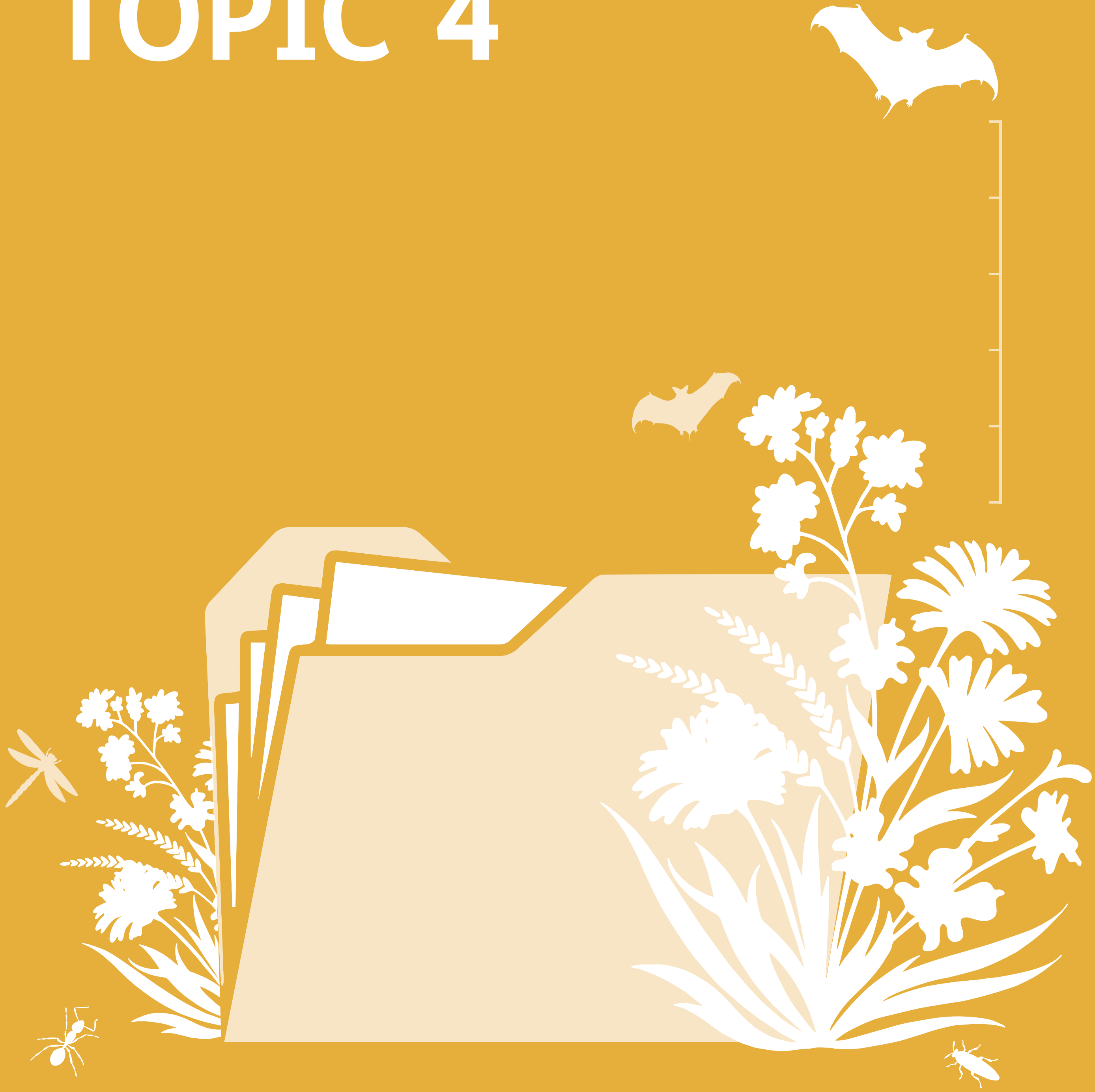
EU BON
http://eubon.eu/

EU BON Portal
http://biodiversity.eubon.eu/

PlutoF
https://plutof.ut.ee

iDigBio
https://www.idigbio.org/

TOPIC 4



STRUCTURED AND STANDARDISED FORMS FOR BIODIVERSITY DATA COLLECTION

Raw biodiversity data (recorded on the field or extracted for scientific publications) is the starting point for all future processing and reporting and it is essential to find a proper solution for data harvesting and storage that enables easy and intuitive procedures for querying and filtering a database. Using (raw) biodiversity data collection forms for recording biodiversity data, structured according to the accepted standards, ensures pooling of data from different sources and its processing for various reporting purposes, such as Natura 2000 Standard Data Forms (SDF), national reports on the state of biodiversity, preparation of scientific articles, etc. The popularity of smartphones and their ease of use provides an opportunity to include general public into scientific inventory and monitoring projects by using simple applications for recording biodiversity data. A number of different initiatives are active throughout the world in this moment, under the common name - **citizen science**.

Importance of collecting data in standardised forms

Biodiversity data based on field observations is essential for a correct overview of biodiversity structures and patterns, which are indispensable for efficient planning and reporting. The more (quality) data, the better. Programmes and projects for biodiversity assessment are, as a rule, long-term and complex operations involving large numbers of individuals with different professional and academic backgrounds. Also, results of previous surveys published in scientific papers and monographs or recorded in researchers' workbooks make a significant contribution to better assessments. It is extremely important to organise field data collection, digitisation of published data and listing of material available in collections in a way that will enable pooling of data from different sources without uncertainties and ambiguities.

Ensuring compatibility of biodiversity data from different sources is a twofold process. The first implies harmonization of the vocabulary of species occurrence attributes, while the second includes using digital or paper forms for recording data (on the field or in the office) structured in a harmonized list of attributes. Use of a common vocabulary of occurrence attributes (e.g. Darwin Core terms) reduces/eliminates uncertainties and ambiguities after data pooling, while use of standardized forms facilitates the integration of digitized data from a variety of sources.

Importance of collecting raw biological data

A widely accepted classification of different types of biodiversity data is commented in detail in Chapter 3 of this document (Importance and Application of Biodiversity Data Standards), It is important to understand essential differences between raw biodiversity data and data that underwent some form of processing (usually raw data recorded in the field or extracted from scientific papers and/or biological collections). Collecting and storing raw biodiversity data in a functional information system ensures its processing for various reporting purposes. Also, it enables a continuity of addition of new data for further fine-tuning and/or creation of updated reports.

For instance, in Chapter 3 of the SDF (ECOLOGICAL INFORMATION) the data for 'Population on the site' and 'Site assessment criteria' for a given species referred to in Annexes to the Birds and Habitats Directives (Type, Size, Population, etc.) are not collected directly in the field. These are estimates based on expert judgments. Reliability of these judgements is highly dependent on the number of single records of a particular species during field visits to the site and, in general, more records lead to higher reliability of judgments. Harmonised and standardised data collected by different individuals and stored in raw form in database(s), enables better overview and more reliable estimation of Natura 2000 site values for particular species.

Using mobile data collection solutions

A new generation of smartphones enables casual and professional nature watchers to contribute to building of the national information resource on environment. Designing an application that runs on mobile phones allows the recording of a photo, location and details of a species in the field. A number of applications were developed for collecting and digitising field biodiversity data in standardised form on the global, regional, even national level. Most of them provide the opportunity to download collected data in structured form for later upload to some data management or mapping system. There are a number of active applications in use by a number of users. **iNaturalist**, **Observado**, **Map of Life**, **eBird**, **SISS-Geo**, to mention a few, can be freely downloaded from the internet, customised for the intended purpose and then used for recording observations in the field. Although most of these applications struggle with a number of difficulties (e.g. maintenance of species vocabularies) and controversies about the quality of collected data, these applications are a useful tool for increasing the quantity and quality of biodiversity data for inventories and species monitoring.

Correct identification of species in an occurrence is the most important factor, of critical importance for reliability of biodiversity data. Other very useful applications for mobile devices that can significantly increase the quality of recorded data in the field are digital keys for species identification. Capabilities of mobile devices are raising the use of conventional field guides to a brand new level. Applications like **NatureGate** or **PlantNet** facilitate the process of identification of the species of a particular observed specimen in a simple and effective way, directly in the field.

Observado - **https://observation.org/index.php**

iNaturalist - **https://www.inaturalist.org/**

eBird - **http://ebird.org/content/ebird/**

Map of Life - **https://mol.org/**

Pl@ntNet - **http://identify.plantnet-project.org/**

NatureGate - **http://www.luontoportti.com/suomi/en/**

Practical examples of data collection forms

As emphasized, it is important to collect and store raw biodiversity data in a standardized structure and using unified forms. Usually this can be achieved by preparing templates in the form of an (Excel) table with column headings that are harmonised with terms from the appropriate TDWG.

At this time, DwC represents a glossary of almost 600(!) terms that describe different aspects that clarify field occurrence of individuals of a particular taxon in nature, which can be used for different purposes. The text below will only refer to those terms that describe data usually collected by practitioners and are necessary for assessing natural values in a particular area.

Required DwC terms

Basic collection/observation event information describes **what, who**, **where**, **when** and **how** data about the particular taxa recorded in the field. These attributes enable a wide range of processing and interpreting operations on the collected data. The essential DwC terms[6] that are necessary for describing a collecting/ observation event are: ***scientificName***, ***identifiedBy***, ***recordedBy***, ***locality***, ***eventDate*** and ***basisOfRecord***. Without these data a particular collection/observation cannot be meaningfully interpreted and processing of incomplete data cannot provide usable results.

Recording and storing a number of attributes for each field collection/observation highly increases the usability of recorded data. The DwC terms that explain more closely an identification of the observed taxon are: *scientificNameID*, *dateIdentified* and *identificationRemarks*. Terms that clarify the basic taxonomy of the identified taxon are: *kingdom* and *taxonRank*. An important term related to the collection/observation event are: *occurrenceStatus* and *occurrenceRemarks*. It is useful to provide more details about the locality, such as: *locationRemarks*, *locationID*, *decimalLatitude*, *decimalLongitude* and *geodeticDatum*. An important term related to biological data about the collected or observed specimen(s) are: *individualCount*, *organismQuantity* and *organismQuantityType*.

6 Explanations and examples of usage of frequently used DwC terms are provided in the Annex.

Recommended DwC terms

DwC vocabulary is extremely rich, with a number of terms defined so as to capture specific aspects of a field observation. Some terms are extremely specific while others are common and logical. Recording and storing as many attributes of the observation as possible is highly recommended because it will increase the quality and interpretability of data and enable the use of this data for reporting in different contexts. Some frequently used biodiversity data attributes of a field observation, in addition to required DwC terms, are as follows:

1. Identification and taxon details: *scientificName Authorship*, *identificationReferences*, *identification Qualifier;*
2. Locality: *coordinateUncertaintyInMeters*, *minimumElevationInMeters*, *maximumElevationInMeters*, *minimumDepthInMeters*, *maximumDepthInMeters;*
3. Biological data: *sex*, *lifeStage*, *behavior* and *associatedTaxa;*
4. Record level information:
    a. Authorship and data usage: *license, accessRights* and *rightsHolder;*
    b. Data cataloguing: *occurrenceID, catalogNumber;*
    c. Data on processing and storage of specimens: *institutionCode, collectionCode, modified, preparation;*
5. References (such as images or citation): *associatedMedia*, *associatedReferences* and *bibliographicCitation;*
6. Verbatim (original) data: *verbatimEventDate*, *verbatimLocality*, *verbatimCoordinates*, *verbatimLatitude*, *verbatimLongitude*, *verbatimCoordinateSystem*, *verbatimElevation* and *verbatimDepth*.

Data on identification and taxon details, locality and biological data are usually collected in the field by experts and practitioners and the volume of collected occurrence attributes depends of the survey design and/or requirements of the funders. In the BIS it is also necessary to keep further data that are important for evaluation of the value and reliability of every single occurrence. Darwin Core enables recording of a wealth of data about occurrences that are stored in some database or published in electronic and paper articles. **Record level information** is information about authorship and mode of use for every single piece of data in the BIS as well as information about denotation of occurrences, if the data is stored in another database. For museum specimens there is a possibility to record data about methods of their processing and storing in a collection. Details about published occurrence data can be recorded also - not only the data that defines authorship and source of the data (**references**) but also attributes of the occurrences recorded in the form in which they were published - **verbatim data**.

EXAMPLE

Implementation of DwC in the Centre for Biodiversity Informatics CBI, Faculty of Biology, Belgrade

Within the framework of projects “Establishment of an ecological network in the Republic of Serbia” and “Development of the Red Book of Plants, Animals and Fungi in the Republic of Serbia”, supported by the former Ministry of Agriculture and Environment, CBI was engaged to prepare a database of biodiversity data on occurrences of species (and habitats) of conservation concern in Serbia and to provide a database in standardised form (simple DwC) to the Institute for Nature Conservation in Belgrade. The project started in 2016 and was planned to last two years: the first year for integrating published data and the second for collecting field data for targeted species in territories recognised in the first phase as important for improvement of the Serbian ecological network and documentation of pSCIs for Natura 2000. Due to administrative difficulties, the project was suspended during 2017 and its continuation is uncertain. However, experiences from project implementation in 2016 (related to organisation and coordination of biodiversity data collection from various sources, including implementation of internationally accepted standards on data storage) can be a useful guidance in the planning of improvements to biodiversity data management and reporting in regional economies.

The start of this project coincided with the CBI’s plan to configure and launch its BIS, so the project encouraged definition of some functional aspects of the CBI - which is an organisational unit of the Faculty of Biology in Belgrade. Two important documents were discussed and accepted: a) **Protocol for Collection, Processing, Organization and Management of Biodiversity Data in CBI** and b) **CBI Standards: Classification, Names, Definitions, Formats and Examples of Biodiversity Data**. The Protocol defines the main aspects of CBI’s operation and possible forms of cooperation, such as: a) data acquisition and ownership rights, b) access to data, c) data categories and expected formats for provision of data to the CBI, d) methods of gathering and digitizing the data, and e) financial aspects of cooperation. The Book of Standards contain Serbian names and detailed explanations of 215 attributes organised in 13 categories: Taxon, Identification, Locality, Georeferencing, Collecting event, Population, Habitat, Storage, Reference, Verbatim data, Proprietorship, Data entry and Data quality. Attributes were mapped to DwC (and ABCD) terms.

The first phase of the project (digitisation of already published biodiversity data) was successfully implemented in 2016. Over 500,000 records were digitised at the primary level, of which nearly 120,000 records were georeferenced. More than 80,000 records were submitted to the database of the Institute for Nature Conservation in Belgrade. The data was entered in Excel tables by more than one hundred collaborators and verified by experts for different groups, that were engaged on the project. Several Excel template tables for data entry were prepared for the project. The most important were:

1. Template for the taxonomic list;
2. Templates for occurrence data;
3. List of attributes with predefined values.

▼ **Table 2:** *Fragment of the CBI template structure for details about the taxon (species and/or subspecies)*

CBI Name [7]	DwC term	Example	Comment
TaksonID	Taxon	Equisetum arvense L. f. decumbens G. F. W. Mey	
TaksonFilum	phylum	Tracheophyta	
TaksonRed	order	Equisetales	
TaksonFamilija	family	Equisetaceae	
TaksonRod	genus	Equisetum	
TaksonRodAutor*	-	L.	Name of person(s) who described the genus
TaksonPodrod	subgenus	Hippochaete	
TaksonPodrodAutor*	-		Name of person(s) who described the genus
TaksonEpitetVrsta	specificEpithet	arvense	
TaksonVrstaAutor	scientificNameAuthorship	L.	
TaksonEpitetInfraspecijski	infraspecificEpithet	decumbens	
TaksonInfraspecijskiAutor*	-		Name of person(s) who described the infraspecific taxon
TaksonNarodniNazivVrsta	vernacularName		
TaksonReferenca*	-	Christenhusz, M. & Raab-Straube, E. von (2013): Polypodiopsida. – In: Euro+Med Plantbase - the information resource for Euro-Mediterranean plant diversity.	Reference or list of references on the basis of which taxonomic and nomenclatural status is defined
TaksonReferencaKlasifikacija*	-	The Angiosperm Phylogeny Group (1998). "An ordinal classification for the families of flowering plants". Annals of the Missouri Botanical Garden 85 (4): 531–553. doi:10.2307/2992015	Reference or list of references in which a classification system was published, applied to taxonomic categories above the genus level

7 Rows in the following tables correspond to columns in Excel tables used for data entry.

CBI Name [7]	DwC term	Example	Comment
TaksonNapomena	taxonRemarks		
ImePravilnikITD*	-		Verbatim taxon name from the Rulebook on Proclamation and Protection of Strictly Protected and Protected Wild Species of Plants, Animals and Fungi in Serbia
ZaštitaSR*	-		National conservation status
IUCN_Status*	-		IUCN status
Bern*	-		Bern Convention status
Natura2000*	-		Natura 2000 Annexes
Endemizam*	-		Level of endemism

** Additional fields not defined by DwC*

▼ **Table 2A:** *Fragment of the CBI template structure for details of occurrences - attributes of the collection event*

CBI name	DwC term	Comment
SakupljanjeSakupio	recordedBy	
SakupljanjeDatum*	-	Date of the collection event i.e. the date when the sample was collected or registered
NalazMetodaSakupljanja	samplingProtocol	
NalazBrojPrimeraka	individualCount	
NalazPol	sex	
NalazOntogenetskaFaza	lifeStage	
NalazStatusNalaza	occurrenceStatus	
NalazDigitalniZapisID*	-	URI of the digital record (photo, video, audio) that confirms the collected/observed specimen(s)
NalazDigitalniZapisAutor*	-	Author of the digital record

** Additional fields not defined by DwC*

▼ **Table 2B:** *Fragment of the CBI template structure for details of occurrences - attributes of the taxon identification*

CBI name	DwC term	Comment
IdentifikacijaPrva*	-	Taxon name of the first identification
IdentifikacijaPrvaDatum*	-	Date of the first identification
IdentifikacijaReferenca	identificationReferences	List of references used for identification of a taxon in the last accepted identification
IdentifikacijaPoslednja*	identifiedBy	Taxon name in the last accepted identification
IdentifikacijaPoslednjaDatum	dateIdentified	Date of the last accepted identification
IdentifikacijaPoslednjaIdentifikator*	-	Name of the person who provided the last accepted identification
IdentifikacijaOstale*	-	List of all identifications except the first and the last
IdentifikacijaVerifikacija	identificationVerificationStatus	
IdentifikacijaPouzdanost	identificationQualifier	

** Additional fields not defined by DwC*

▼ **Table 2C:** *Fragment of the CBI template structure for details of occurrences - attributes of the locality*

CBI name	DwC term	Comment
LokalitetGrafičkiPrikazID	locationID	Unique identifier - code of the locality (e.g. WP-039)
LokalitetID*	-	Alphanumeric, generated by concatenating the "Locality-Locality1", "LocalityLocality3", "LokalitetGrafičkiPrikazID" fields separated with underscore "_". (e.g. Kopaonik_Đorov most_WP-039)
LokalitetLokalitet1	locality	Wider area where the species was registered (e.g. Kopaonik)
LokalitetLokalitet2	locality	Closer location where the species was registered (e.g. Samokovska klisura)
LokalitetLokalitet3	locality	Narrowest location where the species was registered (e.g. Đorov most)
LokalitetVodenoTelo	waterBody	
LokalitetNadmorskaVisina(AltMin)	minimumElevationInMeters	
LokalitetNadmorskaVisina(AltMax)	maximumElevationInMeters	
LokalitetDubinaMin	minimumDepthInMeters	
LokalitetDubinaMax	maximumDepthInMeters	
StaništeOpis	habitat	

** Additional fields not defined by DwC*

▼ **Table 2D:** *Fragment of the CBI template structure for details of occurrences - locality georeferencing attributes*

CBI name	DwC term	Comment
GeoreferenciranjeGeoreferncirao	georeferencedBy	
GeoreferenciranjeMetoda	georeferenceProtocol	
LokalitetGeodetskiDatum	geodeticDatum	
LokalitetGeoreferenciranjePreciznost	coordinateUncertaintyInMeters	
GeoreferenciranjeVerifikacija	georeferenceVerificationStatus	

** Additional fields not defined by DwC*

▼ **Table 2E:** *Fragment of the CBI template structure for details of occurrences - collected specimen(s) storage attributes*

CBI name	DwC term	Comment
NalazInstitucijaKod	institutionCode	
NalazZbirkaKod	collectionCode	
NalazInventarskiBroj	catalogNumber	
NalazKolekcija*	-	Specially organized part of the collection that relates to a particular group of organisms, individual and/or group of collectors, or special area where collected
NalazKolektorskiBroj	recordNumber	
NalazIndividualniBroj	individualID	
NalazNomenklaturniTip	TypeStatus	
NalazMetodaPrepariranja	preparations	
NalazPreparator*	-	Name of the person who prepared the specimen(s)
NalazDatumPrepariranja*	-	Date of preparation

** Additional fields not defined by DwC*

▼ **Table 2F:** *Fragment of the CBI template structure for details of occurrences - attributes of the reference where the occurrence were published*

CBI name	DwC term	Comment
ReferencaAutorReferenca*	-	List of authors of the reference
ReferencaAutorCitat*	-	Short list of authors of the reference
ReferencaGodina*	-	Year of publication
ReferencaNaslov*	-	Title
ReferencaIzdanje*	-	Journal name or publisher of the publication
ReferencaTipPublikacije*	-	Publication type

** Additional fields not defined by DwC*

▼ **Table 2G:** *Fragment of the CBI template structure for details of occurrences - attributes of the citation and verbatim data, as published in the reference where the occurrence was published*

CBI name	DwC term	Comment
OriginalniDatum*	-	Verbatim date on which the sample/specimen was entered into the collection, published in literature, or recorded as a field observation
OriginalnaIdentifikacija	originalNameUsage	
OriginalniLokalitet	verbatimLocality	
OriginalniGeodetskiDatum	verbatimSRS	
OriginalniKoordinatniSistem	verbatimCoordinateSystem	
OriginalnaLatituda(N_lat)	verbatimLatitude	
OriginalnaLongituda(E_long)	verbatimLongitude	
OriginalnaNadmorskaVisina(Alt)	verbatimElevation	
OriginalnaDubina	verbatimDepth	
ReferencaCitatPun	associatedReferences	
ReferencaStrana*	-	Page or table in which the particular occurrence was mentioned

** Additional fields not defined by DwC*

▼ **Table 3:** *Examples of attributes with predefined values in the CBI database*

CBI Name	DwC term	Example	Comment
VlasništvoPravaPristupa	accessRights	1: Restricted Access 2: Semi-Restricted Access 3: Free Access	
KvalitetTipPodatka	type	1: Physical object 2: Picture 3: Video clip 4: Sound recording 5: Record from transmitter, echo-detector, etc. 6: Personal field observation 7: Text	
KvalitetPrirodaPodatka	basisOfRecord	1: Live specimen 2: Dead specimen 3: Product/Effect/Trace 4: Fossil remains	
TaksonRang	taxonRank	sp subsp var subvar f subf cv ×	

CBI Name	DwC term	Example	Comment
LokalitetGeoreferenciranjePre-ciznost	coordinateUncertaintyInMeters	1: not georeferenced 2: 5 m 3: 10 m 4: 100 m 5: 500 m (MGRS 1x1 km) 6: 1000 m 7: 5000 m (MGRS 10x10 km) 8:> 5000 m 9: Unknown	
GeoreferenciranjeMetoda	georeferenceProtocol	1: GPS device 2: GoogleEarth/Google Map 3: Digital gazetteer 4: Digital map/Ozi explorer 5: Topographic maps 6: Other	
GeoreferenciranjeVerifika-cija	georeferenceVerificationStatus	1: Verification necessary 2: Georeferenced by the collector 3: Georeferenced by the provider 4: Georeferenced by the verifier	
PopulacijaBrojnostOpseg	-	1: 1-5 2: 6-10 3: 11-50 4: 51-100 5: 101-250 6: 251-500 7: 501-1000 8: 1001-10000	Population abundance
IdentifikacijaVerifikacija	identificationVerificationStatus	1: Direct Personal Identification of the Provider 2: Accepted identification by indisputable authority 3: Direct reinterpretation of identification 4: Indirect reinterpretation of identification 5: Accepted identification without analysis	
IdentifikacijaPouzdanost	identificationQualifier	1: High reliability (> 70%) 2: Moderate reliability (30-70%) 3: Low reliability (<30%) 4: Reliability uncertain 5: Temporary or approximate identification 6: Identification impossible	

CBI Name	DwC term	Example	Comment
Endemizam*	-	1. Serbian steno endem 2. Balkan endem 3. Balkan subendem 4. Pannonian endem 4. Relict	Level of endemism

** Additional fields not defined by DwC*

Further information

DwC terms: Quick Reference Guide
http://rs.tdwg.org/dwc/terms/

TOPIC 5

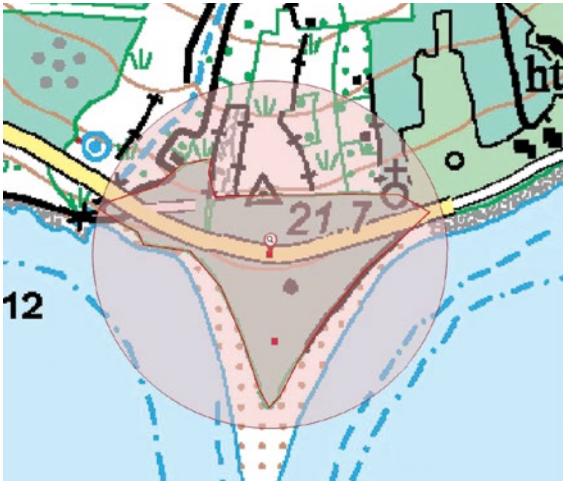
GEOREFERENCING

This topic is meant to assist BIMR users to better understand the importance and underlying principles of the georeferencing process and what it means to georeference localities from literature data. Today a lot of valuable biodiversity data is available only in paper form and is as such unavailable for analysis and processing. Finding a way to quickly and unmistakably georeference locations mentioned in literature is very important. So, in a nutshell, georeferencing is a process of assigning geographic coordinates and maximum error distances for those coordinates to locality descriptions found in literature.

What is georeferencing?

In the context of a BIS, georeferencing is the process of assigning real-world coordinates to biodiversity data with an accompanying estimation of precision. Recent biodiversity data is mostly georeferenced by default because it is collected either with GPS-enabled devices, mobile devices, or in the form of geotagged photos so the latitude and longitude coordinates are collected automatically and are immediately known. (For data to be fully georeferenced we also need to know how imprecise the data is; for GPS this is given by default but this is not the case with geotagged data). Data that needs to be georeferenced is data that is found in literature (reports, books and articles) or biological collections (museums and herbariums). At the time of collection, collectors did not have reliable means to assign coordinates to the locality where the species was collected so they collected descriptive information about the locality, and this information may have been mapped into maps but with little precision.

Although such data has questionable analytical value, it often makes for most of the data we possess and is the only proof of presence of certain taxa in some area and a valuable indicator of changes in biodiversity. Without means of assigning “exact” coordinates to the biodiversity “object”, biodiversity data from literature,



Georefild	6922
Name	Borova šuma, Zlatni rat
Canton	Split - Dalmatia Canton
Municipality	Bol
x	510,884.10
y	4,790,955.15
Uncertainty of locality	206.84 m
Tool	Uncertainty of polygon locality

▲ **Figure 5:** *Example of geotagged data Source: Croatian Agency for Environment and Nature (2017) & State Geodetic Administration (2017)*

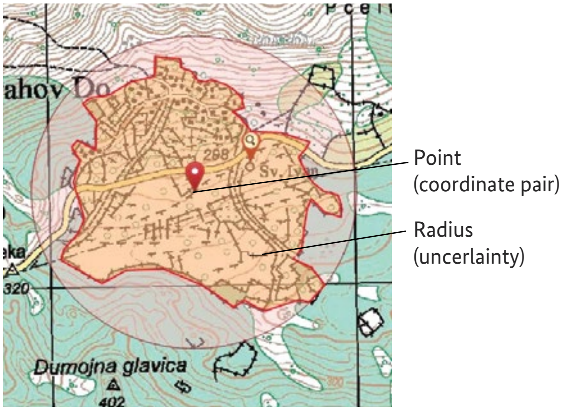
What we have?	What we want?				
	Id	Species	Latitude	Longitude	Uncertainty
Aquila chrysaetos, Bobotov kuk peak, Durmitor	1	Aquila chrysaetos	43.127781	19.034436	750m
Pelican, 1.5 km W of Remas, Karavasta lagoon	2	Pelicanus crispus	40.888248	19.491907	1250m
Dianthus scardicus, Around Great Rudoka peak, Šar Mountain	3	Dianthus scardicus	41.949910	20.798970	2300m
Acer heldreichii, 200 m SE from the crossroads of road 197 and road to Vučak, Golija	4	Acer heldreichii	43.356269	20.257470	125m
...	...	...	...	...	...

▲ **Table 4:** *Presence of certain taxa in the area with coordinates and uncertainty*

herbariums and collections cannot be used in further analyses, research and assessments, so it is important to first georeference such data.

Point-radius method

Although there are several methods to assign correct coordinates to a given textual locality, the point-radius method is widely used and accepted as the "de-facto" standard. This method tries to describe the species locality with two measurements, one is the **x, y (lat/lon) coordinate pair** which defines the centre point (a circle, that is) or centroid of the shape of the locality described in literature, and the other is the **distance** from that point to the farthest point of that shape which represents the possible uncertainty of these coordinates. In this manner we assume that the point and radius which define the circle will contain the actual collection locality. The key advantage of this method is that the uncertainties can be readily combined into one attribute, whereas the bounding box method requires uncertainty to be represented independently in each of the two dimensions (Wieczorek, 2004.)



▲ **Figure 6:** *Example of locality with point-radius method Source: Croatian Agency for Environment and Nature (2017) & State Geodetic Administration (2017)*

This method is not perfect but it allows the georeferencer to assign similar and consistent measurements to each locality found in literature, regardless of whether locality description was given as a named place or maybe uncertainty of direction and distance, and even for different maps and their datums.

How data is georeferenced

The **first step** in georeferencing a described locality is to determine the possible sources of uncertainty in locality descriptions and define the extent of the feature. Although descriptions found in literature and natural history collections can vary immensely in every imaginable way (language, details in describing a feature, different units of measurement, etc.), we are essentially dealing with only nine possible categories of descriptions, given in the table below (Wieczorek, 2004.).

This first step is very important because at the very start of the georeferencing process we should be aware and be able to immediately recognize localities that cannot be georeferenced. In some cases (like in the first three cases in the table below) it is possible that the locality cannot be identified, as per the rules in the below table, or can match more than one category. In that case locality should not be georeferenced and this information should be noted for other users to help them in future.

Second step of the georeferencing process is to determine coordinate points (coordinate pair) of the locality description. This could be obtained from various geographic maps, gazetteers, geographic name databases ("toponymic databases"), GPS devices (recent collections and their locality descriptions often include locality coordinates together with descriptions).

Table 5: *Description of nine possible categories of locality*

Type	Descriptions	Examples
Dubious	The locality explicitly states that the information contained therein is in question.	"Šar Planina?", "presumably central Bosnia"
Cannot be located	Either the locality data is missing, or contains information other than locality information, or the locality cannot be distinguished between multiple possible candidates, or the locality cannot be found with available references.	"locality not recorded", "Visoko brdo", "peak"
Demonstrably inaccurate	The locality contains irreconcilable inconsistencies.	"South of Fruška gora toward Danube"
Coordinates	The locality consists of a point defined by coordinates.	"42.988718, 19.136842", "6667134.94, 4616661.05"
Named place	The locality consists of a reference to a geographic feature (e.g., town, cave, spring, island, reef, etc.) having a spatial extent.	"Skadarsko Lake", "Peshna cave", "Lapušnik"
Offset	The locality consists of an offset (usually a distance) from a named place.	"3 km outside Tuzla"
Offset along a path	The locality describes a route from a named place.	"1 km NE of Orahovo via Road 105", "200 m NE from the crossroads of road 197 and road to Vučak toward Vučak"
Offsets in orthogonal directions	The locality consists of a linear distance in each of two orthogonal directions from a named place.	"1 km N and 2 km W of Tetovo"
Offset at a heading	The locality contains a distance in a given direction.	"1.5 km W of Remas, Karavasta lagoon"

Third step of the georeferencing process is to accurately calculate uncertainty, i.e. radius of the circle that encompasses the locality, to account for all its possible and associated uncertainties. All uncertainties related to georeferencing a locality come from six sources, which are:

- the extent of the locality,
- unknown datum,
- imprecision in distance measurements,
- imprecision in directional measurements,
- imprecision in coordinate measurements, and
- map scale.

When georeferencing, all the combined information on exact coordinates and calculated uncertainties are needed to accurately determine fitness of the final data and thus its quality. Detailed description of each of the sources of uncertainty described above and how to exactly and accurately calculate them are presented in several georeferencing articles, the most popular of which are given below. Also, together with written guidelines, different versions of georeferencing calculators were developed to assist with georeferencing of descriptive localities and especially with calculating uncertainties without the need for thorough understanding of the algorithmic complexity underlying the georeferencing process.

Example of georeferencing a locality description

For the purposes of these guidelines, 2 simple examples of georeferencing locality descriptions are presented: locality description as a named place, and locality description defined between two features (or named places). For all other rules (categories of different locality descriptions), more detailed instructions can be found in Wieczorek, 2004 or by using available georeferencing services.

Example 1: Feature (named place)

This is the simplest example and often the first one in georeferencing a locality. It usually consists only of a named place, which is often a feature listed in a standard gazetteer. This feature can probably be located on a map of the appropriate scale. Examples of such features includes: town, populated place, island, bay, airport, dock, cave, peninsula, hill, mountain, park, junction etc. The important thing to note is that regardless of how the features are presented on a map (or gazetteer), features are not points and they always have a spatial extent. If the spatial extent is obvious, the georeferencing procedure involves identification of the geographic centre (*e*) for the coordinates and determination of the distance from these coordinates to the farthest point within the named place as the extent (*e'*). If the spatial extent is not so obvious, e.g. mountains, road junctions etc., we need to define the extent manually using best judgment and also document it for later usage. If the locality is "exact" (such as a GPS reading), the accuracy of the GPS is used as the extent.

Locality: "Kumanovo" - Suppose the coordinates for Kumanovo came from Google Maps/Earth and the distance from the centre of Kumanovo to the outermost city limit is 3 km.

Coordinate System: decimal degrees

Latitude: 42.1323° N

Longitude: 21.7257° E

Datum: WGS84

Coordinate Precision: 0.0001 degrees

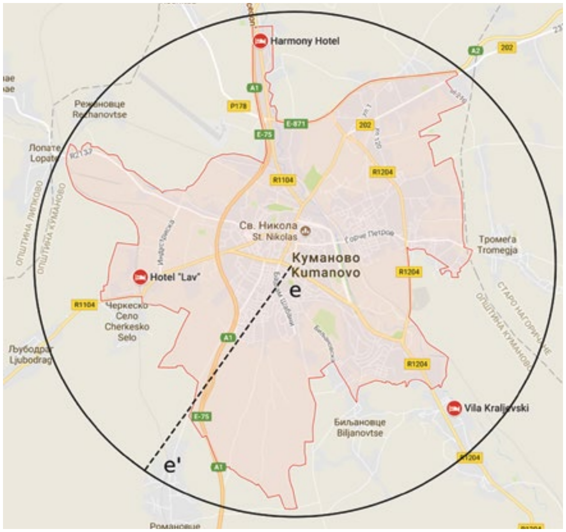
Extent of Named Place: 3 km

Decimal Latitude: 42.1323

Decimal Longitude: 21.7257

Maximum Uncertainty Distance: 3.014 km

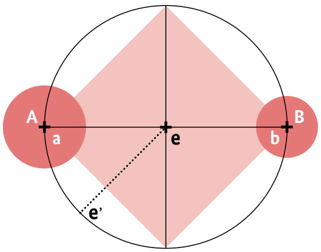
▼ Figure 7: *Locality „Kumanovo“*
Source: Google Maps/Earth



Example 2: Between two features (or named places)

This is example of a locality found in a literature description and defined as located between two features, e.g. “between Kumanovo and Vojnik”, and should be georeferenced in two steps. To georeference this type of locality description, the first step would be to find the coordinates (*a*, *b*) of both named places (*A*, *B*). The second step of the georeferencing procedure includes finding the midpoint between those coordinates (centres of named places), which will be the final coordinates (*e*) for the locality description. Extent is defined as one-half the distance between the centres of named places (*e’*). After that the uncertainty is calculated, same as for the feature (named place) in example 1 (see figure below).

▶ Figure 8:
Example of locality found in a literature and defined as located between two features, e.g. “between Kumanovo and Vojnik”



Resources needed for successful georeferencing

The resources needed for successful georeferencing depend on the scope and extent of the project. Also, depending on the size and type of the institutions involved, different resources will be needed but the basics include:

- A database and/or database software that will serve as a repository for storing georeferenced localities;
- Different topographic maps (electronic or paper) of various scales, years and types, any maps with various toponyms, terms, themes, etc. are welcome (topographic, military, vegetation, history, etc.);
- Access to a good gazetteer - many are available for free on the Internet, either as downloads or searchable online (usually available through the state-administered geoportal or similar service);
- Dedicated and trained personnel;
- Defined methodology for georeferencing (e.g. point-radius method described in…);
- Defined data cleaning and validation procedures;
- Defined documentation procedures for all of the above, which result in a document that covers the key aspects of the georeferencing process.

Important terms related to georeferencing (Chapman, 2006):

Gazetteer - a geographic dictionary or index of *locality names,* usually also including an indication of position on the Earth’s surface using one of several *geographic coordinate systems;*

GPS - Global Positioning System;

Lat (Latitude) - describes the angular distance that a location is north or south of the equator, measured along a line of longitude;

Lon (Longitude) - describes the angular distance east or west of a prime meridian (q.v.) on the earth's surface along a line of latitude;

Offset - a displacement from a reference point, named place, or other feature. Used here as the distance from a named place using the location of the named place as the starting point;

Uncertainty - measure for the maximum distance from the x,y coordinates of a described locality to the farthest part of the locality area (which is often described as a circle) in which the whole of the described locality must reside.

How to incorporate georeferencing-related fields into the database structure?

When planning a georeferencing project, it is usually not recommended to store the locality and other georeferenced data in the spreadsheet except for temporary purposes (or a single locality). This holds especially true for complex projects expected to generate large numbers of localities and volumes of georeferenced data. In such projects it is better to structure the data in advance and organise the georeferencing process to store the data in databases.

The table below shows some of the terms directly related to georeferencing. As already mentioned in topics 3 and 4, these terms refer to previously explained DwC terms in the "Locality" class. These terms are not mandatory but it is recommended to consider them in the database design phase for inclusion as a basic set of attributes related to locality georeferencing.

▼ **Table 6:** *Terms directly related to georeferencing*

Term	Definition	Description	Data type
decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.	Example: "-41.0983423"	number
decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive.	Example: "-121.1761111"	number
footprintSpatialFit	The ratio of the area of the footprint (footprintWKT) to the area of the true (original, or most specific) spatial representation of the Location. Legal values are 0, greater than or equal to 1, or undefined. A value of 1 is an exact match or 100% overlap. A value of 0 should be used if the given footprint does not completely contain the original representation. The footprintSpatialFit is undefined (and should be left blank) if the original representation is a point and the given georeference is not that same point. If both the original and the given georeference are the same point, the footprintSpatialFit is 1.	Detailed explanations with graphical examples can be found in Chapman and Wieczorek, eds. 2006.	string

Term	Definition	Description	Data type
footprintSRS	A representation of the Spatial Reference System (SRS) for the footprint of the Location. Do not use this term to describe the SRS of the decimalLatitude and decimalLongitude, even if it is the same as for the footprint - use the geodeticDatum instead.	Example: The WKT for the standard WGS84 SRS (EPSG:4326) is "GEOGCS["GCS_WGS_1984",DATUM["D_WGS_1984",SPHEROID["WGS_1984",6378137,298.257223563]],PRIMEM["Greenwich",0],UNIT["Degree",0.0174532925199433]]" without the enclosing quotes.	string
footprint	A representation of the shape (footprint, geometry) that defines the Location. A Location may have both a point-radius representation (see decimalLatitude) and a footprint representation, and they may differ from each other.	Example: the one-degree bounding box with opposite corners at (longitude=10, latitude=20) and (longitude=11, latitude=21) would be expressed in well-known text as POLYGON ((10 20, 11 20, 11 21, 10 21, 10 20))	
geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum or code of the ellipsoid, if known. If none of these is known, use the value "unknown".	Examples: "EPSG:4326", "WGS84", "NAD27", "Campo Inchauspe", "European 1950", "Clarke 1866"	string
georeferencedBy	A list (concatenated and separated) of names of people, groups, or organizations who determined the georeference (spatial representation) for the Location.	Examples: "Kristina Yamamoto (MVZ); Janet Fang (MVZ)", "Brad Millen (ROM)"	string
georeferencedDate	The date on which the Location was georeferenced. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E).	Examples: "1963-03-08T14:07-0600" is 8 Mar 1963 2:07pm in the time zone six hours earlier than UTC, "2009-02-20T08:40Z" is 20 Feb 2009 8:40am UTC, "1809-02-12" is 12 Feb 1809, "1906-06" is Jun 1906, "1971" is just that year, "2007-03-01T13:00:00Z/2008-05-11T15:30:00Z" is the interval between 1 Mar 2007 1pm UTC and 11 May 2008 3:30pm UTC, "2007-11-13/15" is the interval between 13 Nov 2007 and 15 Nov 2007.	timestamp

Term	Definition	Description	Data type
georeferenceProtocol	A description or reference to the methods used to determine the spatial footprint, coordinates, and uncertainties.	Examples: "Guide to Best Practices for Georeferencing" (Chapman and Wieczorek, eds. 2006), Global Biodiversity Information Facility.", "MaNIS/HerpNet/ORNIS Georeferencing Guidelines", "Georeferencing Quick Reference Guide"	string
georeferenceRemarks	Notes or comments about the spatial description determination, explaining assumptions made in addition or opposition to the those formalized in the method referred to in georeferenceProtocol.	Example: "assumed distance by road (Hwy. 101)"	string
georeferenceSources	A list (concatenated and separated) of maps, gazetteers, or other resources used to georeference the Location, described specifically enough to allow anyone in the future to use the same resources.	Examples: "USGS 1:24000 Florence Montana Quad; Terrametrics 2008 on Google Earth"	string
georeferenceVerificationStatus	A categorical description of the extent to which the georeference has been verified to represent the best possible spatial description. Recommended best practice is to use a controlled vocabulary.	Examples: "requires verification", "verified by collector", "verified by curator".	string
locality	The specific description of the place. Less specific geographic information can be provided in other geographic terms (higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description.	Example: "Bariloche, 25 km NNE via Ruta Nacional 40 (=Ruta 237)"	string
locationAccordingTo	Information about the source of this Location information. Could be a publication (gazetteer), institution, or team of individuals.	Examples: "Getty Thesaurus of Geographic Names", "GADM"	string
locationID	An identifier for the set of location information (data associated with dcterms:Location). May be a global unique identifier or an identifier specific to the data set.		URI, URL, DOI
pointRadiusSpatialFit	The ratio of the area of the point-radius (decimalLatitude, decimalLongitude, coordinateUncertaintyInMeters) to the area of the true (original, or most specific) spatial representation of the Location. Legal values are 0, greater than or equal to 1, or undefined. A value of 1 is an exact match or 100% overlap. A value of 0 should be used if the given point-radius does not completely contain the original representation. The pointRadiusSpatialFit is undefined (and should be left blank) if the original representation is a point without uncertainty and the given georeference is not that same point (without uncertainty). If both the original and the given georeference are the same point, the pointRadiusSpatialFit is 1.	Detailed explanations with graphical examples can be found in the Chapman and Wieczorek, eds. 2006	string
verbatimCoordinates	The verbatim original spatial coordinates of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem.	Examples: "41 05 54S 121 05 34W", "17T 630000 4833400"	string

Term	Definition	Description	Data type
verbatimLat-itude	The verbatim original latitude of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem.	Example: “41 05 54.03S”	string
verbatimLo-cality	The original textual description of the place.	Example: “25 km NNE Bariloche por R. Nac. 237”	string
verbatim-Longitude	The verbatim original longitude of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem.	Example: “121d 10’ 34” W”	string
verbatimSRS	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which coordinates given in verbatimLatitude and verbatimLongitude, or verbatimCoordinates are based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum or code of the ellipsoid, if known. If none of these is known, use the value “unknown”.	Examples: “EPSG:4326”, “WGS84”, “NAD27”, “Campo Inchauspe”, “European 1950”, “Clarke 1866”	string

Further information

GBIF
http://www.gbif.org/resource/80536

GEOLocate
http://www.museum.tulane.edu/geolocate/default.html

iDigBio
https://www.idigbio.org/

TOPIC 6

NATURA 2000 PREPARATION - BIMR REQUIREMENTS

Over 18% of the European Union land area and almost 6% of its marine territory is covered by the Natura 2000 network, the largest coordinated network of protected areas in the world. Natura 2000 network is designated under the Birds and Habitats Directives, with the main goal to ensure long term sustainability of over 230 habitats and 1,500 species of animals and plants and all bird species naturally occurring in the EU.

IT solutions for management and reporting of Natura 2000 data

The Natura 2000 network designation process requires collection, processing and management of vast amounts of biodiversity data and, as such, a data-driven process, it requires an appropriate IT infrastructure/platform to efficiently handle and maintain this data.

For each Natura 2000 site, national authorities have to submit a Standard Data Form (SDF) that contains an extensive description of each site and its ecology. In order to be able to report Natura 2000 in the SDF format, national authorities responsible for reporting Natura 2000 have to have an appropriate database and IT solution for storage and maintenance of data for each Natura 2000 site.

Natura 2000 database model is already defined, it is in line with the Natura 2000 Standard Data Form (SDF - **http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32011D0484**) and is available for download on the Reference portal for Natura 2000 (**https://bd.eionet.europa.eu/activities/Natura_2000/reference_portal**), together with other supporting data such as codelists and various other documents.

Furthermore, DG Environment, European Environment Agency (EEA) has developed the Natura 2000 software, which is free and available for download at **https://bd.eionet.europa.eu/activities/Natura_2000/N2000_software**. The Natura 2000 software package provides a ready-to-use solution for preparing all information about Natura 2000 sites necessary for the reporting to the EC. This is important to consider, because it might not be necessary to invest funds in development of in-house solutions for reporting Natura 2000 data.

Descriptive data about each Natura 2000 site must be accompanied by corresponding spatial data representing the geographic borders of each site. It is therefore also necessary to have an adequate spatial database for storage of GIS data, as well as an appropriate GIS tool for data entry and editing.

The national delivery of Natura 2000 consists of three parts:

a. A **descriptive database** comprising the information from the SDF in the electronic format;

b. A **spatial dataset** comprising the electronic boundaries of the sites;

c. An **explanatory note** explaining the changes in the database concerning site boundary changes, as well as any additions/deletions of sites, species and/or habitat types, as compared to the previous database delivery.

The format of the descriptive database should be identical to the standard SDF template. The structure of the XML document should follow ('validate against') the XML schema (see Natura 2000 reference portal).

Spatial dataset:

- Electronic boundaries should be submitted in ESRI Shapefile format.
- Electronic boundaries must include a field called SITECODE, containing the Natura 2000 code of the site.
- Electronic boundaries should be dissolved according to SITECODE. Un-dissolved Shapefiles require additional processing steps.
- Shapefiles must contain a projection file (.prj) containing the necessary projection system to view the files.
- GIS files may consist of multiple files to accommodate different projections used by Member States.
- In order to minimise the risk of potential coordinate transformation errors in further data processing, economies are encouraged to deliver their GIS data in an ETRS89 (or WGS84) based coordinate reference system, where feasible.
- The consistency of data from different regions (e.g. no overlap), particularly in the case of Member States with deferral structure, must be ensured. [8]

It is important to understand that information that is reported for each Natura 2000 site in SDF format is processed and aggregated ecological information about each Natura 2000 site. Prior to the ecological assessment of each site, vast amount of raw species and habitat types occurrence data needs to be collected, processed and analysed by experts. Occurrence data is collected through organized field inventories conducted by many experts and it is therefore necessary to consider tools and IT solutions that will enable collection and storage of structured data that can be easily included in the analysis. This means that apart from Natura 2000 database that is in line with SDF structure, authorities that are responsible for Natura 2000 preparation have to provide databases and applications for collection and storage of raw field data. This is important not only to ensure the necessary quality of data important for the designation phase, but also to ensure adequate data for later reporting.

Data collected through the process of field inventory activities implemented in the scope of Natura 2000 projects is a very valuable and important input for many nature protection activities in general. It is therefore important to consider data usage rights when engaging and contracting various biology experts to perform the necessary field inventories in order to ensure that this data can be later used not only for Natura 2000 but also for other nature conservation activities.

Data resources for Natura 2000

When preparing the Natura 2000 proposal, candidate economies need to be aware that successful preparation of Natura 2000 requires more than just the recent species occurrence data. For better assessment of the conservation status of each species and habitat type, other data has to be available as well, such as occurrences from literature, species distribution data, red list data, etc. The important condition in the evaluation of proposed Natura 2000 sites is that this data must be available for verification, which will be done by experts from the European Topic Centre on Biological Diversity (ETC/BD) and interested public from the academic and NGO community.

8 Natura 2000 Reporting guidelines https://bd.eionet.europa.eu/activities/Natura_2000/Folder_Reference_Portal/Reporting%20guidelines%20update%201.3-March%202012.pdf

Additional datasets and data that should be consulted includes, but is not limited to, the following:

Species data

- Field research data - species occurrences (inventory data)
- Literature data on species
- Specific targeted field research data
- Species distribution areas
- Red list data
- Endemic status of species

Habitats

- Field research data
- Literature data on habitats
- Habitat map
- Floristic data

Protected areas - nationally designated protected areas.

Other - speleological data, marine data

When planning and creating the Natura 2000 boundaries for each potential site, another set of spatial datasets is important for effective planning, implementation and management of Natura 2000 sites. This data includes accurate and recent digital orthophotos, accurate and recent administrative borders, various topographic maps and satellite imagery, and especially cadastre data to accurately mark-off each site.

Basemaps and base layers include, but are not limited to, the following:

- Biogeographical regions
- NUTS regions
- National administrative borders
- Water bodies
- Digital orthophoto
- Satellite imagery (for example European Space Agency (ESA) imagery - public institutions are eligible for free access to this data)
- Cadastre - it is important to note that close cooperation is needed with official cadastre in each economy
- Topographic maps
- Digital elevation model
- National Grids (1km, 5km, 10km, 50km)
- Corine Land Cover
- Traffic data (roads, railways etc.)

Further information

Reference portal for Natura 2000
https://bd.eionet.europa.eu/activities/Natura_2000/reference_portal

Standard Data Form
http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32011D0484

NUTS regions overview
http://ec.europa.eu/eurostat/web/nuts/overview

NUTS regions GIS data (Shapefile/Personal GDB) download
http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administative-units-statistical-units/nuts

Biogeographical regions overview
https://biodiversity.eionet.europa.eu/activities/Natura_2000/chapter1

Biogeographical regions GIS data
https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3

Natura 2000 Software Package
https://bd.eionet.europa.eu/activities/Natura_2000/N2000_software

Natura 2000 Database and GIS
http://ec.europa.eu/environment/nature/natura2000/db_gis/index_en.htm
http://ec.europa.eu/environment/nature/natura2000/data/index_en.htm

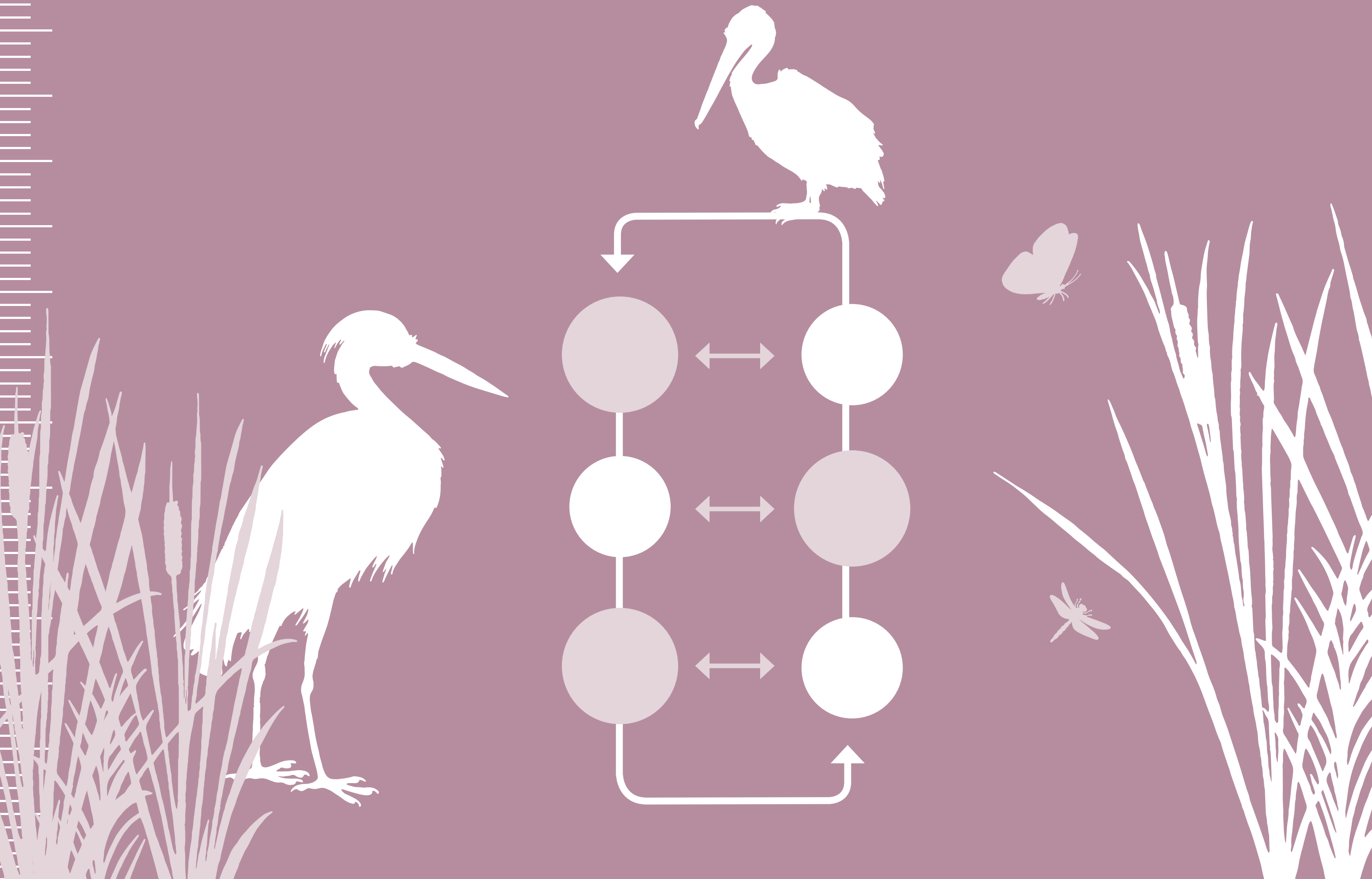
Access to Natura 2000 Data
http://ec.europa.eu/environment/nature/natura2000/access_data/index_en.htm
https://www.eea.europa.eu/data-and-maps/data/natura-8

EIONET GIS - Geospatial data - Maps
http://www.eionet.europa.eu/gis/

European Topic Centre on Biological Diversity
https://bd.eionet.europa.eu/

Natura 2000 species code list + EUNIS
https://www.eea.europa.eu/data-and-maps/daviz/sds/list-of-species-with-natura/@@view

TOPIC 7



COMPLIANCE WITH THE EU INSPIRE DIRECTIVE

Spatial Data Infrastructure (SDI) is a framework of specific policies, technologies, data, institutional arrangements and people that support sharing and effective usage of spatial (geographic) data. It is achieved by applying standardized formats and protocols for data exchange and interoperability.

The goals of SDI are to: 1) reduce duplication of efforts among governments, 2) lower the costs related to geographic information while making geographic data more accessible, 3) increase the benefits of using available spatial data, and 4) establish key partnerships between states, counties, cities, academia, and the private sector.[9]

What is INSPIRE Directive?

Infrastructure for Spatial Information in the European Community (INSPIRE) is a European Union (EU) directive which aims to create EU spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment.

European Spatial Data Infrastructure enables sharing of environment-related spatial data between public sector organisations, facilitates public access to spatial information across Europe, and assists in policy-making across boundaries.

The INSPIRE Directive binds EU members to establish a spatial data infrastructure via the Internet, to facilitate sharing of geographic information in a standardized way.

Making this information widely accessible enables many industries and public agencies to add value and reduce costs. A national SDI (NSDI) is a foundation for producing, sharing, and consuming geospatial information, thus improving decision making and service delivery across many sectors.

The Directive came into force on 15 May 2007 and will be implemented in stages, with full implementation required by 2020.

For the INSPIRE Roadmap please follow the link **http://inspire.ec.europa.eu/inspire-roadmap/61.**

Main principles and components of the INSPIRE Directive

Main principles of the INSPIRE directive:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe

9 “Tonchovska, Rumyana; Stanley, Victoria; De Martino, Samantha. 2012. Spatial Data Infrastructure and INSPIRE. Europe and Central Asia knowledge brief;issue no. 55. World Bank, Washington, DC. © World Bank. **https://openknowledge.worldbank.org/handle/10986/17082** License: CC BY 3.0 IGO.”

and share it with many users and applications.

- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

Main components of the INSPIRE directive:

- metadata,
- interoperability of spatial data and services,
- services (discovery, viewing, downloading, transformation and invoke),
- joint use of spatial data and services,
- coordination and supervision and reporting measures.

INSPIRE Directive Themes

The Directive addresses 34 spatial data themes needed for environmental applications and successful construction of environmental information systems.

INSPIRE spatial data themes are arranged into three annexes (Annex I, II and III) (see page 79)[10].

INSPIRE Directive Themes related to biodiversity

Themes related to biodiversity and nature conservation include the following:

10 http://inspire.ec.europa.eu/webarchive/index.cfm/pageid/2/list/7.html

ANNEX I

Protected sites. This theme includes Natura 2000 Ecological Network as well as nationally designated protected areas. INSPIRE directive defines protected sites as areas designated or managed within a framework of international, Community and Member States' legislation to achieve specific conservation objectives.
https://themes.jrc.ec.europa.eu/groups/profile/1737/protected-sites

ANNEX III

Habitats and Biotopes. Habitats and Biotopes is a biodiversity theme that deals with habitats and biotopes as areas and their distinct boundaries. INSPIRE defines habitats and biotopes as Geographical areas characterised by specific ecological conditions, processes, structure, and (life support) functions that physically support the organisms that live there. Includes terrestrial and aquatic areas distinguished by geographical, abiotic and biotic features, whether entirely natural or semi-natural.
https://themes.jrc.ec.europa.eu/groups/profile/1757/habitats-and-biotopes

Species Distribution. Species Distribution is a biodiversity theme focused on geographical distribution of occurrence of biological organisms aggregated by grid, region, or any administrative or analytical unit. Distributions may be represented in a wide range of formats, such as points, grid cells at different scales or polygons of specifically defined areas. To achieve harmonization EU-Nomen is the preferred reference list for species (taxon) names to be used, the second choice is European Nature Information System and finally Natura2000.
https://themes.jrc.ec.europa.eu/groups/profile/2017/species-distribution

Bio-geographical regions. Bio-geographical Regions describe areas of relatively homogeneous ecological conditions with common characteristics. The INSPIRE

▼ Figure 9: *Annexes I, II and III with INSPIRE Directive themes*

ANNEX 1

- Addresses
- Administrative units
- Cadastral parcels
- Coordinate reference systems
- Geographical grid systems
- Geographical names
- Hydrography
- Protected sites
- Transport networks

ANNEX 2

- Elevation
- Geology
- Land cover
- Orthoimagery

ANNEX 3

- Agricultural and aquaculture facilities
- Area management / restriction / regulation zones & reporting units
- Atmospheric conditions
- Bio-geographical regions
- Buildings
- Energy Resources
- Environmental monitoring Facilities
- Habitats and biotopes
- Human health and safety
- Land use
- Meteorological geographical features
- Mineral Resources
- Natural risk zones
- Oceanographic geographical features
- Population distribution and demography
- Production and industrial facilities
- Sea regions
- Soil
- Species distribution
- Statistical units
- Utility and governmental services

theme Bio-geographical Regions has a strong linkage to other "biodiversity themes". The Habitats Directive (EEC/92/43) is the most important guiding document regarding to Bio-geographical Regions, which contains a list of 'bio-geographical regions' (Article 1.iii). These bio-geographical regions are the basis of a series of seminars evaluating the Natura 2000 network and for reporting on the conservation status of the habitats and species protected by the Directive.

https://themes.jrc.ec.europa.eu/groups/profile/1762/bio-geographical-regions

Who is obliged to comply with the INSPIRE Directive?

The INSPIRE Directive binds EU members to establish a national spatial data infrastructure (NSDI) that facilitates sharing of geographic information in a standardized way.

NSDI includes spatial data owned by government bodies or by other institutions for government bodies and spatial data used by government bodies for completion of their public duties.

NSDI subjects are public authorities whose competences, i.e. scope of work, include establishing or maintaining spatial data and which are obliged to participate in NSDI establishment, maintenance and development.

Public authorities, NSDI subjects are:

- state bodies,
- national, regional and local level bodies, and
- legal persons with public responsibilities.[11]

In the environment and nature conservation sector there will be various public institutions, governmental agencies or other academic institutions that will be identified as NSDI subjects and as such will be responsible for the provision of datasets covered by the INSPIRE Directive and national NSDI laws and regulations. For example, environmental and nature protection agencies are usually responsible for provision and maintenance of a dataset related to species distribution areas, habitat maps or Natura 2000 Ecological Network and nationally designated protected areas.

Compliance and data conformance with the INSPIRE Directive

To ensure that the spatial data infrastructures of the Member States are compatible, the Directive requires that common Implementing Rules (IR) are adopted in a number of specific areas (Metadata, Data Specifications, Network Services, Data and Service Sharing and Monitoring and Reporting).

Metadata

Metadata is data that provides detailed information about the content and quality of a data resource (datasets or data services). Metadata should include sufficient detail about a spatial information resource so as to allow a user to make an accurate judgment on its content, quality, currency and conditions of access and reuse of certain data source or data set.

INSPIRE metadata includes various information such as name of dataset, description of the content, data source, referenced coordinate system, data format, spatial extent of data, information about dynamics of data updates and maintenance, usage rights etc.

More information and resources: **http://inspire.ec.europa.eu/webarchive/index.cfm/pageid/101.html**

11 **http://www.nipp.hr/default.aspx?id=263**

Data specifications

Each of the INSPIRE data themes has its own Data Specification associated with it. The Data Specification documentation aims to harmonise datasets according to the INSPIRE spatial data theme models, thus ensuring their interoperability between public organizations and the wider European INSPIRE network.

Data specification for INSPIRE theme can be found here: **http://inspire.ec.europa.eu/webarchive/index.cfm/pageid/2.html**

Network services

In order for users to be able to find data according to specific search criteria, view the metadata and the spatial data themselves, and download them on their own computer for further use network services have to be published conforming to the INSPIRE network services implementing rules. The INSPIRE network services build further on existing international standards from W3C, ISO and OGC.

More information and resources: **http://inspire.ec.europa.eu/webarchive/index.cfm/pageid/5.html**

Data and service sharing

Public bodies holding INSPIRE data will have to ensure compliance with the requirements of the INSPIRE Data and Service Sharing regulation[12].

The main points of the Regulation are the following:

- Metadata must include the conditions applying to access and use for Community institutions and bodies; this will facilitate their evaluation of the available specific conditions already at the discovery stage.
- Member States are requested to provide access to spatial data sets and services without delay and at the latest within 20 days after receipt of a written request; mutual agreements may allow an extension of this standard deadline.
- If data or services can be accessed under payment, Community institutions and bodies have the possibility to request Member States to provide information on how charges have been calculated.
- While fully safeguarding the right of Member States to limit sharing when this would compromise the course of justice, public security, national defence or international relations, Member States are encouraged to find the means to still give access to sensitive data under restricted conditions, (e.g. providing generalized datasets). Upon request, Member States should give reasons for these limitations to sharing.

More information and resources: **http://inspire.ec.europa.eu/webarchive/index.cfm/pageid/62.html**

Monitoring and reporting

In order to have a solid basis for decision making related to the implementation of INSPIRE Directive and to the future evolution of INSPIRE, continuous monitoring of the implementation of the Directive and regular reporting are necessary.

Monitoring and reporting have to cover the 4 main fields of the INSPIRE Directive: metadata, spatial data sets and services, network services, data sharing.

More information and resources:
http://inspire.ec.europa.eu/webarchive/index.cfm/pageid/5022.html

SEE Regional SDI initiatives

IMPULS, a cross-border SDI project in SEE: the IMPULS project is a regional SDI project financed by Sida (Swedish International Development Cooperation Agency). Project beneficiaries include the cadastre

12 http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32010R0268

authorities of SEE region, and the project was approved for financing by the European Commission (EC) under the 2011 Instrument for Pre-Accession Assistance (EU IPA) Multi-Beneficiary Programme. The project is managed by Lantmäteriet, the Swedish Mapping and Land Registration Authority together with the Croatian counterpart State Geodetic Administration as junior partner. The project aims to increase institutional capacities for SDI development in the countries involved (in accordance with the INSPIRE guidelines and other EU legislation), specifically focusing on their capacity to collect, process, exchange, and create available spatial data, thus better preparing these countries for EU membership. [13]

Further information

INSPIRE Directive Portal
http://INSPIRE.jrc.ec.europa.eu/

INSPIRE Geoportal
http://www.INSPIRE-geoportal.eu/

13 "Tonchovska, Rumyana; Stanley, Victoria; De Martino, Samantha. 2012. Spatial Data Infrastructure and INSPIRE. Europe and Central Asia knowledge brief;issue no. 55. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/17082 License: CC BY 3.0 IGO."

TOPIC 8

DATA USAGE AND AUTHORSHIP RIGHTS, INCLUDING THE CROATIAN EXAMPLE

Proper management of biodiversity data can play an important role in environmental protection, policy development and proper management of natural resources, including also the economic benefits from all these fields. However, much biodiversity data and information often remains unused due to unsolved issues related to copyright and authorship rights.

Many public institutions finance various biodiversity inventory projects and research. One of the main outputs of such projects is both raw and processed data that is very valuable for nature conservation expert work. If usage rights are not clearly tackled or not covered at all in formal agreements/contracts, institutions are not allowed to share such data and the data remains unavailable for further use by various stakeholders (from public institutions to firms that prepare impact assessment studies).

The importance of properly addressing data usage rights and authorship was recently taken into consideration in the legislation of European countries, including the economies in SEE. The EU BON - Building the European Biodiversity Observation Network Project with 31 partners from 18 countries is dedicated, inter alia, to understanding the usage of biodiversity information and data in a wider context, from the social, economic, as well as scientific perspective. It aims to foster the development of a new open-access platform for sharing biodiversity data and tools, a system that facilitates open access to taxonomic data and will allow sustainable provision of high quality data to the public. The project aims to withdraw biodiversity data from the context of copyright property rights. In this regard it is considered that the "biodiversity data is not considered a protected immaterial good and is thus not subject to intellectual property legislation. Biodiversity data, if gained in a systematic way and in agreement with standards and conventions, is unlikely to have any individual imprint in itself and thus any copyright issue", Egloff et al., 2016. This is also in line with the EU Directive 2001/29/EC, which stresses the importance and right of general interest in using intellectual property works. According to this directive, the author has the right to decide who shall be allowed to reproduce his work or communicate it to the public but it removes from the author to a wide degree the intellectual property rights for the benefit of the general interest. "The EU Database protection is not part of copyright but is *a sui generis* (special case) right that applies whether copyright relating to the database exists or not. Database rights do not prevent the use of individual data elements or minor parts of the data collection. The EU Database protection also provides for exceptions and limitations in the general interest, for example in the interest of scientific efforts.", Egloff et al., 2016. In this context, database rights apply only to the use of the whole dataset and do not prevent the use of particular elements or smaller portions of the data collection. According to this understanding, this is not only valid for scientific data but also for a large set of biodiversity data presented as images because these are understood to represent facts according to standardized conventions.

The goals and principles of making biodiversity data openly and universally available have been defined in several web portals dealing with biodiversity and environmental data, such as GBIF (Global Biodiversity Information Facility) for example. GBIF has formulated a set of principles and rules for publishers who want to publish data by using GBIF facilities such as: "Biodiversity data accessible via the GBIF network is openly and universally available to all users within the framework of the Data User Agreement; The GBIF Secretariat does not assert any intellectual property rights in the Data and datasets that are made available through its network. The Data Publisher warrants that it has made the necessary agreements with the original owners of the Data and other material which may be subject to a third party's copyright or other similar rights, that it can make the Data available through the GBIF network; Responsibility regarding the restriction of access to sensitive Data resides with the Data Publisher". GBIF promotes free dissemination of biodiversity data and, in particular: (a) should not assert any proprietary rights to the data in databases that are developed by other organisations and that subsequently become affiliated to GBIF; (b) should seek, to the greatest extent possible, to make freely and openly available, with the least possible restrictions on reuse, any data commissioned, created or developed directly by GBIF; and © should respect conditions set by data publishers that affiliate their databases to GBIF. When establishing affiliations or linkages with other databases, GBIF should seek to ensure that the data so made available will not be subject to limitations on further non-commercial use and dissemination, apart from due attribution of its source.

Other institutions dealing with biodiversity data such as IUCN for example, also tend to, wherever possible, ensure open access to biodiversity data for all publicly funded initiatives and research, and acknowledge that open access to, and effective use of, conservation data, information and knowledge resources by all sectors of society is essential both to enable sound decision making and to empower those concerned with the conservation of biodiversity and the natural world. However, IUCN has produced the Policy for Commercial Use of IUCN Biodiversity Data which seeks to standardize protocols and procedures surrounding the use of data by commercial enterprises. "Under this Policy, commercial entities, and other entities using the data for financial gain or for the benefit of a commercial company, are required to enter into a negotiation with IUCN prior to receiving data. This is partly to ensure that the intended use of the data is in line with the conservation purpose outlined below, and partly to explore options for obtaining cost recovery for the generation, processing and management of the data from the commercial entity in question, or from any other entity using the data for financial gain or for the benefit of a related commercial company."[14]

Croatian example

In Croatian law, as well as in most economies in the SEE region, copyright work is defined as any intellectual creation that has an individual character (e.g. reports, individual portions of a report such as tables, graphs, photographs, drawings). In this context, and in relation to biodiversity and the environment, copyright work includes presentations of a scientific or technical nature such as drawings, plans, sketches, tables, etc., as well as databases arranged according to a certain system or method, the elements of which are individually accessible by electronic or other means.

In Croatia, use of data and authorship rights are regulated by specific contracts, terms of use and other documents formulated in accordance with the Croatian Copyright and Related Rights Act. In this regard, the Croatian Agency for Environment and Nature (CAEN)

14 Policy for Commercial Use of IUCN Biodiversity Data, Annex 15 to Decision C/78/24.

has regulated the use of data in information systems developed by CAEN through Terms of Use and authorship rights stipulated in four different documents: 1. Agreement between the Employer and the Executor/ Author(s), 2. Annex to the Agreement, 3. Author Statement and 4. Co-Author Statement.

Use of data is regulated through a consolidated set of rules and standards for biodiversity and environmental databases such as the BIOPORTAL - Nature Protection Information System geoportal (**http://www.bioportal.hr/gis/**), the Croatian Speleological Cadastre Portal, etc. BIOPORTAL is a web information system which is accessible to the public only in a restricted edition with less information and precision. However, it may be made available to any party or the general public through a Request for Data and Publications, which can be submitted online to CAEN. Multimedia content in this information system is available for viewing only, however the complete version may be downloaded subject to written agreement of the authors. BIOPORTAL data can be used, subject to above conditions, as .csv and .xlsx file types and through WFS services. Material available through Bioportal can be used and published with proper attribution, with the exception of use for any activities in violation of the Law on Nature Protection and other applicable Croatian legislation. Data obtained via public access or upon request cannot be used in scientific publications prior to expiry of two years from the time the observation was made, except with prior written permission of the authors. By accepting the Terms of Use, BIOPORTAL users agree not to change the content of the used information in any way, including but not restricted to observers, temporal and spatial data, assessment of species population, species observations, etc. BIOPORTAL Terms of Use indicate that data quality assurance is a continuous process and that CAEN cannot be held liable for any damages arising as a result of outdated data or inaccuracies in the data available on the portal.

On the other hand, the internal instance of the BIOPORTAL and the Cadastre of Speleological Objects can only be used by authorized representatives of parties that have a signed Cooperation Agreement with CAEN. In this case, access to information requires a registration process and users are prohibited from sharing their access information with third parties. Use of information at both portals is also subject to authorship rights, as mentioned above.

Authorship rights and transfer of rights are regulated by a contract between the author and the other party, as well as a statement of the author of the information. The Agreement between the Employer and the Executor/ Author(s) stipulates the conditions of use and transfer of rights to the outputs produced by the Executor/ Author(s) under the Agreement, which apply, without limitation, to reports, data contained in the reports, research and data collection results and analyses as well as their interpretation, photographs, recordings, video recordings, audio recordings, sketches, drawings, tables, etc. This Agreement establishes the Employer's right to use the Author's work, particularly the right to: multiplication (reproduction), publication and further distribution in any medium and in any format known today or in the future, alteration, development, translation, adaptation, processing or any modification of the copyright work in question, regardless of the purpose for which the work was altered, without further consent, permission and/or approval of the Author. However, by this Agreement "*the Employer agrees not to alter the original data (e.g. data about the observer, observations of species and habitats, estimates of population abundance and distribution of species and/or habitats, opinions and conclusions) in the Copyright Work, and that in case of use of the Copyright Work in its entirety, shall not modify the complete report*"[15]. Through this Agreement, the Author(s) gives the Employer the right

15 Internal document of the Croatian Agency for Environment and Nature, "Agreement on the Transfer and Regulation of Author(s)' Rights.

to grant further non-exclusive right to use the data to any third party, with the provision that "*the content and scope of rights may not be greater than the rights that were transferred to the Employer or were granted in favour of the Employer*". Under this Agreement, the Executor/ Author(s) must inform the Employer of the original author or co-authors of the work for the purposes of citing their name(s) in the course of further use of this information.

The second document, Annex to the Agreement is signed between two parties for the purposes of regulation of mutual rights and obligations covered by the Agreement. The Annex further puts the Agreement into the context of the Copyright and Related Rights Act (Official Gazette 167/03, 79/07, 80/11, 125/11, 141/13, 127/14).

The Author's Statement includes information about granting the non-exclusive right to use the Copyright Work to the other party (association / institution / company) without limitations related to the duration, space and content, as related to particular issues mentioned in the Agreement. This statement includes a mandatory declaration by the Author that the Copyright Work is not subject to exclusive third party copyright and that the Author will not grant exclusive right to use of the Copyright Work to a third party. This statement stipulates that the Author must not limit in any way the rights of the other party regarding the Copyright Work. Similarly, the Co-Author's Statement defines the other party's right to use the Copyright Work in the same way and under the same conditions as the Author's Statement.

Further information

Data Policy Recommendations for Biodiversity Data. EU BON Project Report

Policy for Commercial Use of IUCN Biodiversity Data, Annex 15 to Decision C/78/24

Terms of Use for BIOPORTAL

Terms of Use for the Cadastre of Speleological Objects

Terms of Use of the Internal Portal of Bioportal

The Contract Agreement of Transferring and Regulating Author(s) Rights

The Statement of the Author

The Statement of the Co-author

TOPIC 9

NATIONAL LEGISLATION RELATED TO BIMR

A Biodiversity Information System (BIS) is a powerful tool to disseminate biodiversity information and improve management efficiency and decision-making impacts. It is the basis for monitoring, analyses, planning and implementation of biodiversity conservation actions and sustainable use. More details about the BIS concept and definitions are given in Topic 1. In general, the aim of the BIS is to make information on wildlife, their habitats and important sites readily available to those who need it. It supports implementation of the EU strategy and Aichi targets in Europe (BIS for Europe) and also supports reporting towards different multilateral environmental agreements.

Well-developed legislation for biodiversity information system on a national level is a prerequisite for establishment of an effective NBIS.

The National BIMR Assessments developed for SEE countries indicated a legal obligation to establish and maintain a BIS in the national nature legislation, however this obligation is scarcely mentioned and clear and adequate legislative acts are lacking. Usually, environmental legislation provides much better information about the environmental information system, where BIS is often presented as a sub-module. For example, Macedonian environmental legislation defines the national environmental information system as a comprehensive database that includes information about different media (water, air and soil) and different areas of the environment (nature, waste, noise, vibration, ionizing and non-ionizing radiation, climate, and other elements).

In fact, national environmental legislation provides a good legal basis for the environmental information system and provides details of its content, database maintenance, data provision obligations for different institutions/organizations, and data exchange.

In order to establish a functional BIS, there are several topics that needs to be clearly defined in national legislation (laws or secondary legislation), including:

- What are the main thematic components of the system?
- Who is responsible for system maintenance?
- How is the system financed?
- Who is responsible for data collection?
- Who is obliged to provide data to the system?
- How is data managed?

As explained in Topic 1, BIS consists of various thematic databases, web services, protocols etc. and presents an integrated system of different interconnected thematic databases. For example, it can consist of different databases of flora, fungi, fauna, habitats, forests, protected areas (usually organized as a cadastre of nationally protected areas), speleological objects, Natura 2000 network and/or other ecological networks, internationally designated areas (Important Bird Areas, Important Plant Areas, etc.). Structure of the databases of internationally important areas are usually designed by international organizations or responsible agencies and unified for all participating countries, e.g. Natura 2000 Network (designed by EC), Emerald Network (designed by the Bern Convention, Council of Europe), CDDA-Common Database on Designated Areas (designed by European Environmental Agency), IPAs (designed by Plantlife Int., IBAs (developed by BirdLife Int.), network of Ramsar sites created by Ramsar convention, etc. This is supposed to ease the process of data collection and reporting, however sometimes many different reporting obligations can create additional burdens in terms of the economy's obligations.

In general, the body responsible for establishment and maintenance of the BIS database is a government institution (ministry of environment, environmental protection agency or other relevant body such as a state institute for nature protection). This is usually clearly defined in environmental or nature legislation. Due to complexity of the system, the responsible institution needs to establish a network of other institutions/organizations that should take part in the BIS or make arrangements for coordination with other databases, e.g. forest information system, pastures database, databases in scientific institutions, etc. This will ensure that decisions which may affect natural heritage are made on the basis of best available knowledge.

To establish an effective BIS, a monitoring network based on clear and well identified indicators needs to be in place and the capacities of the relevant stakeholders must be developed (training of personnel) to operate, update, and maintain the system. For example, the Macedonian Nature Law gives the body responsible for execution of expert tasks in the field of nature protection the option to delegate nature status monitoring to accredited legal entities that meet the conditions prescribed by the law/secondary legislation, in accordance with the monitoring methodology. This monitoring network can be easily reassigned to tasks required for establishment and maintenance of the NBIS.

Provision of adequate long-term financial resources for the NBIS is a prerequisite for its long-term functioning. Even though its establishment can be supported by different international donors, it is of great importance to secure an annual national budget for its maintenance. This is a demanding process, so it is extremely important to include clearly defined strategic objectives related to the establishment of the NBIS into the relevant strategic documents (e.g. NBSAP, National Environmental Action Plan, etc.) in order to facilitate the process of obtaining the necessary funding from different international sources (GEF, EU IPA funds, etc.). Project applications which demonstrate a clear connection to strategic documents are often requested by providers of financial support. It should be noted that a local contribution by the economy and demonstrable project sustainability is requested by the international donors for most projects.

Obligation to provide data to the system needs to be clearly defined in the legislation in order to ensure content and have a functional BIS. Environmental legislation is usually more precise on this issue and could be used as good example when developing nature legislation.

Example 1

Croatian environmental legislation could be used as a good example for development of national legislation for BIS. The Croatian Agency for Environment and Nature (CAEN) is the institution responsible for maintenance of the EIS, data analysis and reporting towards EU and international bodies.

The Law on Protection of the Environment prescribes the purpose of the EIS, details about the data and information it contains, mandatory data exchange between different institutions and organizations, as well as national reporting obligations towards the EU and other international organizations. It is important to know that data, information and reports provided to the Agency are subject to verification of quality in order to ensure credibility, completeness and reliability. Also, competent authorities must provide information and data to the Agency without compensation and, upon request from the Agency, must supplement these in the reporting and auditing phases carried out by the competent authorities of the European Union and intergovernmental bodies and organizations. More information can be found in Articles 148-152 of the Croatian Law on Protection of the Environment.

In fact, more details about the structure, content, form and method of operation, administration and maintenance of the EIS are usually prescribed in secondary legislation, in this case the **Regulation on the Environmental Information System** from 2008. According to this ordinance, the aim of the EIS is to connect all existing data and information streams by using modern tools such as the Internet and satellite technology and to ensure that reporting in paper form is replaced by a system in which data is accessible to users at source, in an open and transparent manner. It enables collection and provision of information and data which was processed and analysed in accordance with international and European methodologies and enables exchange of environmental data with similar existing systems. The EIS is organised in 4 basic groups: environmental components, environmental pressures, impact on human health and safety, and societal responses, organized into thematic areas and sub-areas.

Another important step necessary for operationalization of the EIS Regulation and its implementation (i.e. for effective functioning of the EIS) is to develop a programme for its administration. In fact, CAEN is obliged to prepare and adopt the **Environmental Information System Administration Programme** for the purposes of establishment, administration, development, coordination and maintenance of the single EIS. Its contents are: organisation, method of administration and maintenance of the IS; list of reporting entities and method of data submission according to thematic areas and sub-areas; method and deadlines for submission of environmental data and information; method of environmental data and information management; cost estimate; necessary measures and activities for thematic areas and sub-areas.

Example 2

The Macedonian Law on Nature Protection (Official Gazette of the Republic of Macedonia No. 67/04, 14/06, 84/07, 35/10, 47/11, 148/11, 59/12, 13/13, 163/13, 41/14, 146/15, 39/16, 63/16) prescribes mandatory keeping of records related to the protection of nature in the economy. This comprises the **Cadastre of protected areas** (includes all categories of national protected areas as well as areas under temporary protection due to on-going procedure for their proclamation) and the **Registry of natural heritage** (includes strictly protected and protected wild species, speleological objects, minerals, fossils and natural rarities); both are part of the NBIS.

The content of both the cadastre and the registry, method of assignment of a unique registration number, maintenance instructions and data sharing are prescribed in the secondary legislation. More information can be found in the Rulebook on Keeping Records for Nature Protection (Official Gazette of the Republic of Macedonia No. 102/2012).

Development of the Red List is an obligation under the Law on Nature Protection. According to Article 35 of the same law, species can be declared as protected or strictly protected based on the Red List assessments, thus they are becoming a part of the **Registry of natural heritage**.

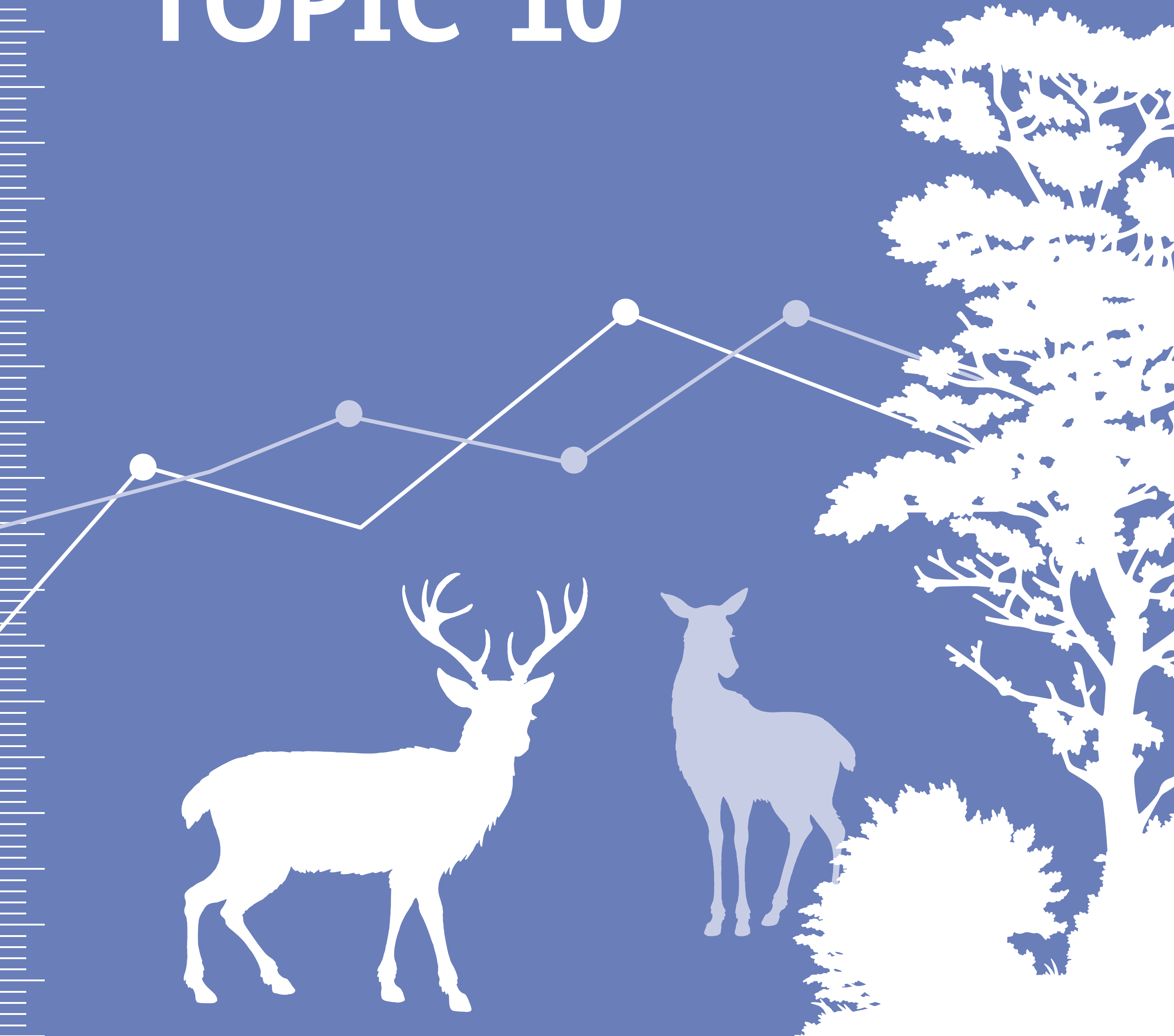
Further information

Croatian Agency for Environment and Nature (information systems)
http://www.haop.hr/hr/informacijski-sustavi

Ministry of Environment and Energy of the Republic of Croatia (regulations and international treaties in the field of nature protection)
http://www.mzoip.hr/hr/priroda/propisi-i-medunarodni-ugovori.html

Macedonian Law on Nature Protection, 2004
http://www.moepp.gov.mk/?page_id=901

TOPIC 10



German Development Cooperation and Gender Equality

The Ministry for Economic Cooperation and Development (BMZ) formulates its objectives with regard to gender equality in the development process in the inter-sectorial policy document "Gender Equality in German Development Policy/Politics" (*Gleichberechtigung der Geschlechter in der deutschen Entwicklungspolitik* – Übersektorals *Konzept*). This policy document was revised and the updated version was published in February 2014. The concept is binding for all implementing organizations, including GIZ.

The revised document reflects recent progress in development thought and practice on gender equality. It rests on three pillars: (i) gender mainstreaming through integrating a gender equality perspective into all development policy and action; (ii) empowerment through targeted elimination of gender-based discrimination and women-specific support measures in support of women' rights; and (iii) integration of women's rights and gender equality in bi- and multi-lateral political dialogue, sector dialogue, and policy advice. This third element is an addition and clear innovation compared to the concept's fore-runner and of particular relevance for ORF, since it explicitly refers to the strategic level of development cooperation. Sex-disaggregated data collection and the use of gender-sensitive indicators are supporting elements which enable monitoring of gender equality and re-adjustment when needed.

As an implementing organization of the BMZ, GIZ operationalizes the Ministry's binding concepts in its own strategies. At present, the relevant strategic document is the 3rd corporate "Gender Strategy 2010-2014" which highlights the need for promoting gender equality first and foremost within the organization, i.e. within GIZ. The strategy further elaborates gender mainstreaming for projects and programmes.[1] The major difference from its two predecessors is that the 3rd version *must* be translated into specific action. Furthermore, the "provisions of the Gender Strategy relating to gender equality within the company are mandatory for all GIZ staff members." The strategy's foreseen indirect results are noteworthy and substantive: (i) men and women benefit to an equal extent from the development inputs of technical cooperation; and (ii) men and women can play an equal and active part in shaping the development inputs of technical cooperation. The gender strategy is closely interlinked to the GIZ concept of sustainability which emphasizes the need for equal opportunities as a clear prerequisite.[2]

Although the topic of biodiversity conservation, sustainable use of natural resources and gender related aspects are facing a complex set of challenges at strategic, policy, legal and operational levels, ORF BD and GIZ are fully committed to gender equality and ensuring the mainstreaming of gender into the processes of overall ORF BD and ORF BD sub-projects planning and implementation. The project developed an ORF BD gender mainstreaming strategy and will proceed with its implementation in the course of the one year remaining until the end of the first phase. The strategy itself consists of three components addressing the perspectives of gender equity, gender equality and gender mainstreaming. The implementation of the gender equity component has already been initiated, hence enabling the project to already deliver results through the well-established mechanism for collection, recording and analysis of sex-disaggregated data.

[1] GTZ (2010). Gender Strategy 2010-2014. Eschborn
[2] GTZ (2008). Sustainability Concept. Eschborn

Introducing Gender

The term gender refers to the **socially-constructed expectations** about the characteristics, attitudes and behaviours associated with being a woman or a man. Gender defines what is feminine and masculine. Gender shapes the social roles that men and women play and the power relations between them, which can have a profound effect on the use and management of natural resources.

Gender is not based on sex or the biological differences between women and men; rather, gender is shaped by culture and social norms. Thus, depending on values, norms, customs and laws, women and men in different parts of the world have adopted different gender roles and relations. Within the same society, gender roles also differ by race/ethnicity, class/caste, religion, ethnicity, age and economic circumstances. Gender and gender roles then affect the economic, political, social, and ecological opportunities and constraints faced by both women and men.

Gender mainstreaming is a comprehensive strategy aimed at achieving greater gender equality. This is obtained by integrating a gender perspective into existing mainstream institutions and all programmatic areas or sectors, such as trade, health, education, environment, and transportation.

Linking Gender and Biodiversity

Considering gender issues in relation to biodiversity involves identifying the influence of gender roles and relations on the use, management and conservation of biodiversity. Gender roles of women and men include different labour responsibilities, priorities, decision-making power, and knowledge, which affect how women and men use and manage biodiversity resources. As a result, women and men develop different knowledge about different species, their uses as well as how to manage them.

The roles and responsibilities of men and women in the management of biodiversity, and the ability to participate in decision-making, vary between and within countries and cultures. However, in most circumstances there are gender-based differences and inequalities, which tend to favour males. Stark gender differences are evident in economic opportunities and access to and control over land, biodiversity resources and other productive assets, in decision-making power, as well as in vulnerability to biodiversity loss, climate change and natural disasters.

To inform efficient policies regarding biodiversity conservation, sustainable use and the sharing of its benefits, we need to understand and expose gender-differentiated biodiversity practices, gendered knowledge acquisition and usage, as well as gender inequalities in control over resources. We need to consider the influences of gender differences and inequalities on the conservation and sustainable use of biodiversity, and the ways in which these differences and inequalities influence how women and men are affected by biodiversity policies, planning and programming.

Exposing and understanding the gender-differentiated biodiversity practices and knowledge of women and men enhances biodiversity conservation. Many case studies from around the world have demonstrated that in empowering women and vulnerable groups to participate as equals in decision-making related to information sharing and generation, education and training, technology transfer, organizational development, financial assistance and policy development, biodiversity conservation efforts become more effective and efficient.

How do we mainstream gender into biodiversity projects?

Some of the key gender issues in biodiversity conservation and management are meaningful participation and voice, education and access to information, access to gender-disaggregated data. Therefore, it is necessary to document the differential knowledge of women and men about biodiversity resources; women and men have complementary knowledge about biodiversity resources which reflects their shared responsibilities. Gender-disaggregated data on the conservation, use, and management of biodiversity and women's and men's differential needs and control over resources need to be documented. Besides, biodiversity programs and projects must strive to increase and encourage women's participation in decision making related to biodiversity conservation. Women's capacity to participate in management of local, community based institutions implementing conservation initiatives should be increased through increased access to information and equitable participation in training and extension services.

Gender aspects to consider when planning, implementing and managing a BIS

Planning phase

- Ensure that gender aspects of the reporting requirements are identified.
- Ensure that relevant gender statistics are part of the data to be collected.
- Ensure gender balance in internal and external project personnel and associates (including IT and biology experts to be engaged on the BIS).

Throughout the project cycle, a range of stakeholders should participate. Stakeholder engagement must be gender-sensitive. This means that:

- Both women and men are represented.
- Women's voices are heard. Typically, women are less forthcoming in meeting settings and are less represented in organizations that are typically consulted in biodiversity projects, such as different associations and groups. Steps need to be taken to ensure that women are willing and able to participate.
- A wide range of women's perspectives are represented. There will not be one female perspective, and efforts should be made to integrate and triangulate a range of female points of view.
- The particular constraints that women may face are taken into account. For example, meetings held at a time when women can leave the home, and in locations that consider women's security concerns.
- In some places, women's organizations may not be adequately organized to ensure their coordinated participation. In these cases, building the capacity of women's associations should be considered, to ensure they are adequately represented and heard in decision-making processes.

- Civil society is not automatically more gender sensitive than other actors. Allocate adequate resources for strengthening the capacities of selected civil society organizations to address and substantiate the nexus between gender equality and biodiversity conservation and management. Pro-actively support the linking of such organizations to regional and international knowledge hubs and networks. Also, engage civil society organizations that represent women's interests.

Implementation phase

- Ensure equal access to training, capacity building and other resources.
- Develop **gender-sensitive and stereotype-free PR**, information dissemination, and awareness raising materials and activities. Alert partners, contracted **designers, DTP specialists, editors and journalists** to the requirement of using gender-sensitive language, avoiding gender stereotypes, and portraying women and men as equally relevant for biodiversity conservation and management. On a routine basis, **check PR materials for their gender content** prior to publication.
- Identify and seek **collaboration with Gender Focal Points** in partner organizations and institutions.

Management phase

- Collect gender-disaggregated data throughout all activities.
- Alert counterparts and partners to their gender equality obligations whenever non-compliance is identified/apparent.
- Ensure that systems, processes and budget commitments are in place to so that the collection of gender-disaggregated data continues once the project ends.
- Document good practice of engendering biodiversity. Assign clear responsibility for such documenting. On a routine basis, present the collected cases, initially internally (e.g. in planning and management meetings) and subsequently in regional events in order to enable learning from experience and increase exchange on approaches, challenges.

In conclusion

Understanding the linkages between gender relationships and the environment means achieving a better analysis of patterns of use, knowledge and skills regarding conservation and sustainable use of natural resources. Only with a gender perspective in place can a more complete picture of human relationships and ecosystems be seen. Gender equality is clearly a matter of fundamental human rights and social justice. Taking gender into account in relation to environmental management is a precondition for sustainable development.

Without the participation of women and the realization of their full creative and productive potential, it will not be possible to attain the Sustainable Development Goals (SDGs), including those related to environmental protection. Sustainable Development Goals emphasize clear linkages between gender equality, poverty alleviation, biodiversity conservation and sustainable development. Such insights should be included into our outlook and approach to reversing biodiversity loss, reducing poverty and improving human well-being.

REFERENCES

Anonymous (2012): A review of barriers to the sharing of biodiversity data and information, with recommendations for eliminating them. Prepared by the UNEP World Conservation Monitoring Centre in its capacity as Secretariat of the Friends of the Conservation Commons. Convention on Biological Diversity.

Chapman, A.D. and J. Wieczorek (eds). (2006): Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility.

Croatian Agency for Environment and Nature (2017): Nature Protection Information System – Application for Georeferencing, Available on: http://georef.iszp.hr/. Accessed: July 12, 2017. [in croatian]

Croatian Environment Law (2013).

Croatian Environment Information System Ordinance (2008).

Egloff W, Agosti D, Patterson D, Hoffmann A, Mietchen D, Kishor P, Penev L (2016): Data Policy Recommendations for Biodiversity Data. EU BON Project Report. Research Ideas and Outcomes 2: e8458.

Environmental Protection Agency (2014): Management Program for Environmental Information System of the Republic of Croatia for the period 2014-2017. [in croatian]

European Environment Agency (2012): Streamlining European biodiversity indicators 2020: Building a future on lessons learnt from the SEBI 2010 process. Technical report No. 11/2012, Copenhagen.

Hamer, M., Victor, J., Smith, G.F. (2012): Best Practice Guide for Compiling, Maintaining and Disseminating National Species Checklists, version 1.0, released in October 2012. Copenhagen: GBIF, 40 pp.

GBIF (2012): GBIF-ICLEI-CBD Best Practice Guide for Publishing Biodiversity Data by the Local Governments, (contributed by Cadman, M.J.; Chavan, V.; Patrickson, S.; Galt, R.; Mader, A.; Sood, R.; Hirsch, T.) Copenhagen: Global Biodiversity Information Facility, pp. 62.

GBIF (2011): Promoting biodiversity data inclusive EIA: Best Practice Guide for publishing primary biodiversity data, (contributed by Cadman, M., Chavan, V., King, N., Willoughby, S., Rajvanshi, A., Mathur, V.B., Roberts, R. and Hirsch, T.) Copenhagen: Global Biodiversity Information Facility, pp 50.

GBIF (2010): Best Practices in Publishing Species Checklists. Contributors: Remsen D, Döring M, Robertson T, Ko B. Copenhagen: Global Biodiversity Information Facility, 10 pp.

Kissling, W. D., J. A. Ahumada, A. Bowser, M. Fernandez, N. Fernandez, E. A. García, R. P. Guralnick, N. J. B. Isaac, S. Kelling, W. Los, L. McRae, J.-B. Mihoub, M. Obst, M. Santamaria, A. K. Skidmore, K. J. Williams, D. Agosti, D.

Amariles, C. Arvanitidis, L. Bastin, F. De Leo, W. Egloff, J. Elith, D. Hobern, D. Martin, H. M. Pereira, G. Pesole, J. Peterseil, H. Saarenmaa, D. Schigel, D. S. Schmeller, N. Segata, E. Turak, P. F. Uhlir, B. Wee and A. R. Hardisty (2017): Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biol. Rev. Camb. Philos. Soc., 2017 Aug 2.

Lakušić, D., A. Ćetković, G, Mesaroš (2016): Protocol for collecting, processing, organization and management of biodiversity data in the Centre for Biodiversity Informatics. Internal document. [in serbian]

Lakušić, D., A. Ćetković, G, Mesaroš (2016): Centre for Biodiversity Informatics Standards: Classification, names, definitions, formats and examples of biodiversity data. Internal document. [in serbian]

Remsen, D. (2016): The use and limits of scientific names in biological informatics. ZooKeys 550: 207–223.

Secretariat of the Convention on Biological Diversity (2014): Global Biodiversity Outlook 4, Montréal.

State Geodetic Administration (2017): Topographic Map TK25, Downloaded from: http://georef.iszp.hr/. Accessed: July 12, 2017. [in croatian]

Wetzel, F. T., H. Saarenmaa, E. Regan, C. S. Martin, P. Mergen, L. Smirnova, É. Ó Tuama, F. A. Garc Ía Camacho, A. Hoffmann, K. Vohland & C. L. Häuser (2015): The roles and contributions of Biodiversity Observation Networks (BONs) in better tracking progress to 2020 biodiversity targets: a European case study. Biodiversity, 16:2-3, 137-149.

Wieczorek & Braak (2015): The GBIF Integrated Publishing Toolkit User Manual, version 2.3. Copenhagen: Global Biodiversity Information Facility.

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. (2012): Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715.

Wieczorek, J., Q. Guo, and R. Hijmans (2004): The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*. 18: 745-767.

van der Sluis, T., Foppen, R., Gillings, S., Groen, T.A., Henkens, R.J.H.G., Hennekens, S.M., Huskens, K., Noble, D., Ottburg, F.G.W.A., Santini, L. and Sierdsema, H. (2016): How much Biodiversity is in Natura 2000? (No. 2738). Alterra Wageningen UR.

ANNEXES

ANNEX 1.
DEFINITIONS OF FREQUENTLY USE DARWIN CORE TERMS

Definitions of frequently use Darwin Core terms
(for full list of term definitions see **http://rs.tdwg.org/dwc/terms/**)

**** Required DwC terms***

Term	Definition
acccessRights	Information about who can access the resource or an indication of its security status. Access Rights may include information regarding access or restrictions based on privacy, security, or other policies (Example: “not-for-profit use only”)
associatedMedia	A list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of media associated with the Occurrence.
associatedReferences	A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the Occurrence.
associatedTaxa	A list (concatenated and separated) of identifiers or names of taxa and their associations with the Occurrence (Examples: “host: Quercus alba”, “parasitoid of: Cyclocephala signaticollis | predator of Apis mellifera”)
basisOfRecord*	The specific nature of the data record (recommended best practice is to use a controlled vocabulary such as the list of Darwin Core classes. Examples: “PreservedSpecimen”, “FossilSpecimen”, “LivingSpecimen”, “HumanObservation”, “MachineObservation”.)
behavior	A description of the behaviour shown by the subject at the time the Occurrence was recorded. Recommended best practice is to use a controlled vocabulary (Examples: “roosting”, “foraging”, “running”)
bibliographicCitation	A bibliographic reference for the resource as a statement indicating how this record should be cited (attributed) when used. Recommended practice is to include sufficient bibliographic detail to identify the resource as unambiguously as possible.
catalogNumber	An identifier (preferably unique) for the record within the data set or collection.
collectionCode	An identifier for the collection or dataset from which the record was derived.
coordinateUncertaintyInMeters	The horizontal distance (in meters) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term.
dateIdentified	The date on which the subject was identified as representing the Taxon.
decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.

Term	Definition
decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive.
eventDate*	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded
geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value "unknown". Examples: "30" (reasonable lower limit of a GPS reading under good conditions if the actual precision was not recorded at the time), "71" (uncertainty for a UTM coordinate having 100 meter precision and a known spatial reference system).
georeferencedBy	A list (concatenated and separated) of names of people, groups, or organizations who determined the georeference (spatial representation) for the Location.
geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value "unknown". Examples: "EPSG:4326", "WGS84", "NAD27"
georeferenceProtocol	A description or reference to the methods used to determine the spatial footprint, coordinates, and uncertainties. Examples: "Guide to Best Practices for Georeferencing. (Chapman and Wieczorek, eds. 2006). Global Biodiversity Information Facility.", "MaNIS/HerpNet/ORNIS Georeferencing Guidelines", "Georeferencing Quick Reference Guide".
georeferenceVerificationStatus	A categorical description of the extent to which the georeference has been verified to represent the best possible spatial description. Recommended best practice is to use a controlled vocabulary.
habitat	A category or description of the habitat in which the Event occurred. Examples: "grassland", "pannonian and pontic sandy steppe".
identificationQualifier	A brief phrase or a standard term ("cf.", "aff.") to express the determiner's doubts about the Identification (Examples: 1) For the determination "Quercus aff. agrifolia var. oxyadenia", identificationQualifier would be "aff. agrifolia var. oxyadenia")
identificationReferences	A list (concatenated and separated) of references (publication, global unique identifier, URI) used in the Identification.
identificationRemarks	Comments or notes about the Identification.
identificationVerificationStatus	A categorical indicator of the extent to which the taxonomic identification has been verified to be correct.
identifiedBy*	A list (concatenated and separated) of names of people, groups, or organizations who assigned the taxon name to the subject.
individualCount*	The number of individuals represented present at the time of the Occurrence.
infraspecificEpithet	The name of the lowest or terminal infraspecific epithet of the scientificName, excluding any rank designation. Examples: "concolor", "oxyadenia", "sayi"
institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.

Term	Definition
kingdom	The full scientific name of the kingdom in which the taxon is classified.
license	A legal document giving official permission to do something with the resource
lifeStage	The age class or life stage of the biological individual(s) at the time the Occurrence was recorded. Recommended best practice is to use a controlled vocabulary (Examples: “egg”, “eft”, “juvenile”, “adult”, “2 adults 4 juveniles”)
locality*	The specific description of the place. Less specific geographic information can be provided in other geographic DwC terms (higherGeography, continent, economy, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description. Example: “south of Martinovice village, 8 km E from Plav”.
locationRemarks	Comments or notes about the Location. Example: “under water since 2005”.
locationID	An identifier for the set of location information (data associated with DCterms: Location). May be a global unique identifier or an identifier specific to the data set.
minimumElevationInMeters	The lower limit of the range of elevation (altitude, usually above sea level), in meters.
maximumElevationInMeters	The upper limit of the range of elevation (altitude, usually above sea level), in meters.
minimumDepthInMeters	The lesser depth of a range of depth below the local surface, in meters.
maximumDepthInMeters	The greater depth of a range of depth below the local surface, in meters.
modified	The most recent date-time on which the resource was changed.
occurrenceID	An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique.
occurrenceRemarks	Comments or notes about the Occurrence (Example: “found dead on road”)
occurrenceStatus	A statement about the presence or absence of a Taxon at a Location. Recommended best practice is to use a controlled vocabulary (Examples: “present”, “absent”)
organismQuantity	A number or enumeration value for the quantity of organisms.
organismQuantityType	The type of quantification system used for the quantity of organisms (e.g. e.g., “27” for organismQuantity with “individuals” for organismQuantityType; “12.5” for organismQuantity with “%biomass” for organismQuantityType; “r” for organismQuantity with “BraunBlanquetScale” for organismQuantityType
preparation	A list (concatenated and separated) of preparations and preservation methods for a specimen.
recordedBy*	A list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence. The primary collector or observer, especially one who applies a personal identifier (recordNumber), should be listed first.
rightsHolder	A person or organization owning or managing rights over the resource
scientificName*	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined.
scientificNameAuthorship	The authorship information for the scientificName formatted according to the conventions of the applicable nomenclaturalCode. Example: “(Torr.) J.T. Howell”, “(Martinovský) Tzvelev”, “(Györfi, 1952)”

Term	Definition
scientificNameID	An identifier for the nomenclatural (not taxonomic) details of a scientific name (Example: "urn:lsid:ipni.org:names:37829-1:1.3")
sex	The sex of the biological individual(s) represented in the Occurrence (Examples: "female", "hermaphrodite", "8 males, 4 females".)
specificEpithet	The name of the first or species epithet of the scientificName. Examples: "concolor", "bosnicus"
taxonRank	The taxonomic rank of the most specific name in the scientificName. Recommended best practice is to use a controlled vocabulary (Examples: "subspecies", "varietas", "forma", "species", "genus".)
taxonRemarks	Comments or notes about the taxon or name. Example: "this name is a misspelling in common use".
TypeStatus	A list (concatenated and separated) of nomenclatural types (type status, typified scientific name, publication) applied to the subject. Examples: "holotype of Ctenomys sociabilis. Pearson O. P., and M. I. Christie. 1985. Historia Natural, 5(37):388", "paratype of Otiorhynchus subnivalis"
verbatimCoordinates	The verbatim original spatial coordinates of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem.
verbatimCoordinateSystem	The spatial coordinate system for the verbatimLatitude and verbatimLongitude or the verbatimCoordinates of the Location. Recommended best practice is to use a controlled vocabulary. Examples: "decimal degrees", "degrees decimal minutes", "degrees minutes seconds", "UTM".
verbatimDepth	The original description of the depth below the local surface (Example: "10-12 m".)
verbatimElevation	The original description of the elevation (altitude, usually above sea level) of the Location (Example: "1000-1500 m")
verbatimEventDate	The verbatim original representation of the date and time information for an Event ("spring 1910", "Marzo 2002", "1999-03-XX", "17IV1934")
verbatimLatitude	The verbatim original latitude of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem.
verbatimLocality	The original textual description of the place (Example: "Avala, Beograd")
verbatimLongitude	The verbatim original longitude of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem.
verbatimSRS	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which coordinates given in verbatimLatitude and verbatimLongitude, or verbatimCoordinates are based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value "unknown". Examples: "EPSG:4326", "WGS84", "NAD27"
vernacularName	A common or vernacular name. Examples: "Golden eagle", "Alpine salamander", "Serbian spruce", "Apollo butterfly".
waterBody	The name of the water body in which the Location occurs. Recommended best practice is to use a controlled vocabulary. Examples: "Adriatic sea", "Skadar lake", "Bosna river".

ANNEX 2.

MODEL TEMPLATE FOR SUBMISSION OF FIELD SURVEY DATA

Model template for submission of field survey data (minimum and recommended set of attributes). *The template can be adapted to the needs of the institution or individual who collects data in the field.*

scientific Name	**identified By**	organism Quantity	organism Quantity Type	**recorded By**	**event Date**	**locality**	decimal Latitude	decimal Longitude	geodetic Datum	**basis Of Record**	occurrence Remarks

Highly recommended additional attributes

for Taxon	kingdom
for Identification	dateIdentified, identificationQualifier, identificationRemarks
for Locality	coordinateUncertaintyInMeters, minimumElevationInMeters, maximumElevationInMeters, minimumDepthInMeters, maximumDepthInMeters, locationRemarks
for Observation	sex, lifeStage, behavior, associatedTaxa
for a specimen from a collection	catalogNumber, institutionCode, collectionCode, preparation
for occurrence ilustration	associatedMedia
for Database records	occurrenceID, accessRights, rightsHolder, modified

ANNEX 3.

MODEL TEMPLATE FOR SUBMISSION OF DATA FROM LITERATURE

Model template for submission of data from literature (minimum and recommended set of attributes). *The template can be adapted to the needs of the institution or individual who inputs the data from literature.*

scientific Name	**associated References**	**verbatim Event Date**	**ver-batim Locality**	verbatim Coordinates	verbatim Latitude	verbatim Longitude	verbatim Coordi-nate System	verbatim Elevation	verbatim Depth

NOTES

NOTES

NOTES

NOTES

german
cooperation
DEUTSCHE ZUSAMMENARBEIT

Published by:

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

GIZ Country Office Sarajevo
Deutche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH
Zmaja od Bosne 7-7a
Importanne centar 03/IV
71000 Sarajevo, Bosnia and Herzegovina
T +387 33 957 500
F +387 33 957 501
I www.giz.de/regional-funds-southeasteurope