



EUROPEAN BIODIVERSITY
OBSERVATION NETWORK

SELECTED
OUTCOMES



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No 308454.

Contents

3

The European Biodiversity Portal

An ideal partner for biodiversity science and policy

8

PlutoF

Data mobilisation and curation platform

11

GBIF - IPT

Supporting the publication of sampling-event data

16

EU BON Taxonomic Backbone

Access to taxonomic information systems

19

ARPHA-BioDiv

A Toolbox for scholarly publishing and dissemination of biodiversity data

24

AquaMaps

Possible climate change impact on bony fish diversity in Large Marine Ecosystems

28

GeoCAT

Geospatial Conservation Assessment Tool

33

TreatmentBank

Access to taxonomic treatments and observation records through text and data mining of taxonomic publications

37

DINA-Web in EU BON



Introduction

This booklet summarises the major outcomes of the EU-funded project EU BON “Building the European Biodiversity Observation Network” which aims to advance biodiversity knowledge by building a European gateway for biodiversity information and by integrating and harmonising a wide range of biodiversity data. Other goals of the project are the establishment and adoption of new data standards, the development of tools, the integration of advanced techniques for data analysis and the development of new approaches and strategies for future biodiversity monitoring and assessment. The project represents a joint effort of 31 partners from 15 European countries, Israel, the Philippines, Brazil and more than 30 associated partners.

In this booklet, you will find concise descriptions of products and services developed by project members during the period 2012-2017. It includes a range of tools for data analysis, such as GeoCAT - a tool that performs rapid geospatial analysis to ease the process of Red Listing taxa and AquaMaps – a toolkit that models the distribution and makes predictions of where aquatic species occur naturally. Another group of tools – the GBIF Integrated Toolkit and ARPHA-BioDiv facilitates the process of data sharing, integration and publishing. Furthermore, in this booklet one can also find information about novel services. Worthy of special mention is the EU BON Unified Taxonomic Information Service (UTIS) that allows the running of a federated search on multiple European taxonomic checklists by scientific name or vernacular name strings. These include the Pan-European Species directories Infrastructure (EU-Nomen), the European Nature Information System (EUNIS), the Catalogue of Life, the World Register of Marine Species (WoRMS), the GBIF Checklist Bank and the Plazi Treatment Bank.

All these web applications are bound together under the umbrella of the EU BON European Biodiversity Portal (<http://biodiversity.eubon.eu>), with highly relevant contributions to the aims of the Group on Earth Observation’s Biodiversity Observation Network (GEO BON). Besides applied tools and software, the portal also serves as an online library for manuals, guidelines, factsheets, case studies, policy recommendations and other documents.

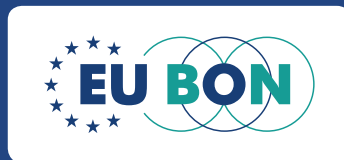
This booklet is aimed at scientists, policy and decision-makers, citizen scientists, members of governmental and non-governmental organisations, data managers, administrators, environmental managers, students and other expert groups.

We sincerely hope that you will find this booklet interesting to read and the EU BON products and services useful for your work.

Dr. Christoph Häuser
Project Coordinator, on behalf of the EU BON consortium

THE EUROPEAN BIODIVERSITY PORTAL

An ideal partner for
biodiversity science and policy



Hannu Saarenmaa

✉ hannu.saarenmaa@uef.fi

Antonio García Camacho

✉ antonio.garcia.camacho@csic.es

Overview

The European Biodiversity Portal (<http://biodiversity.eubon.eu>) is an online platform, which facilitates the delivery of relevant biodiversity information and analysis to a range of end users including researchers, policy-level users and other biodiversity stakeholders, thus supporting the biodiversity science and policy interface. The Portal provides easy access to insights in biodiversity trends and modelling, through visualisation modules and infographics (Fig. 1).

The Portal is an access point to evidence-based solutions and decision-support tools, enabling next-generation data management and the use of biodiversity information at national and regional levels. For example, the Portal can be used to identify data sets relevant to Essential Biodiversity Variables (EBV). Forecasting and the development of early warning systems will be facilitated by EBVs, aiding policy makers to develop evidence based conservation strategies. Visualisation of the status and trends of EBV candidates such as “Species distribution” and “Species abundance” are facilitated using the Spatial Dataset Browser (Fig. 2) and the Species Population Trend Browser (Fig. 3), respectively.

Other analytical interfaces include the Species Richness Tool and the Business Analytics Dashboard, which offer charts and aggregated data. Via the virtual laboratory, eLab, researchers can work on the same project, share documents and use the same web interface while accessing a set of tools offered as a Service in an elastic cloud environment, which also ensures the data privacy inside each eLab. GeoServer, RStudio and other analytical tools can be deployed as Docker containers, thanks to an OpenShift environment and an OpenStack Hybrid cloud.

Other functionalities accessible through the Portal include a Dataset Search Tool and the EU BON Product¹ and Document Repository, which is searchable by target audience and keywords. Additionally, specific child websites are available, such as the Citizen Science site, dedicated to provide information, tools and guidelines for citizen scientists; and also websites for the EU BON test sites, which include data, geospatial layers, results and specific use cases where EU BON tools were applied in a testing site.

Expected advantages

As a gateway to European biodiversity information, the Portal connects a vast number of biodiversity data repositories to capture an exceptional level of coverage in biodiversity data. Users can harvest and simultaneously access data from several directories, including Global Biodiversity Information Facility (GBIF), Long-Term Ecological Research Network (LTER), EU-wide monitoring methods and systems of surveillance for species and habitats of community interest (EuMon), the Pan European Species directories Infrastructure (PESI), which provides taxonomic information on species, and Global Earth Observation System

¹ A full list of EU BON products can be downloaded at: http://wcmc.io/EUBON_Products.



Figure 1. Screenshot of the European Biodiversity Portal.

of Systems (GEOSS) sources, making the fragmented data more useful and providing a more precise account of the current state of biodiversity status and trends. The analytical tools interface allows insights into the complexity of biodiversity information, whilst delivering synthesised, actionable facts and insights, illustrated by informative graphic content.

Applicability

The Aichi Biodiversity Targets of the United Nations Strategic Plan for Biodiversity (2011-2020) recognise the importance of improved access to knowledge, the science base and technologies relating to biodiversity. The European Biodiversity Portal is a contribution to Aichi Target 19, which specifies the need to increase the amount and quality of relevant biodiversity information and tools at the disposal of policy-makers and the general public.

Building on the Millennium Development Goals, the 17 Sustainable Development Goals (SDGs) were adopted by countries at the United Nations General Assembly in 2015 and officially came into force on January 1st 2016. Jointly, SDG 14 (“Life below water”) and SDG 15 (“Life on land”) aim to conserve and sustainably use ocean and land resources (such as fisheries and forests), combat desertification, halt and reverse land degradation,

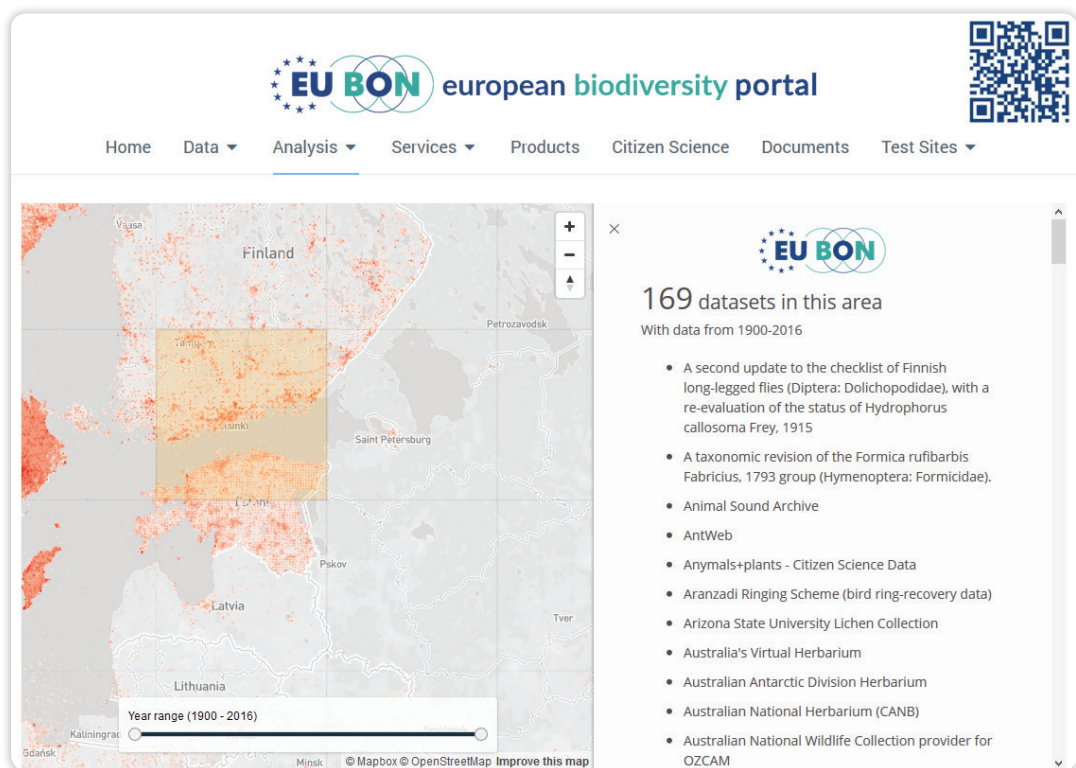


Figure 2. Spatial Dataset Browser helps to identify and access datasets of interest when generating Essential Biodiversity Variables (EBVs) for species distribution.

and halt biodiversity loss. The implementation of both SDGs depends on the availability and accessibility of comprehensive, sound and up-to-date data and information: this is where the European Biodiversity Portal aims to contribute.

Potential users

The European Biodiversity Portal was built by Information and communication technology (ICT) professionals for key stakeholders in the European biodiversity science and policy interface. It aims to facilitate easier access to multiple biodiversity (meta)data warehouses and encourages data mobilization using best-practice standards.

- The unique combination of user-friendliness, state-of-the-art research analytics and informative visualisation modules, make the European Biodiversity Portal an ideal partner for policy-makers, strengthening the political commitment to biodiversity monitoring and conservation in Europe.
- The various toolboxes assist scientists and data managers to upload and publish their data, helping to close knowledge gaps and increase the taxonomic, geographic and historic coverage and accuracy of trend forecasting and status evaluations for European biodiversity.

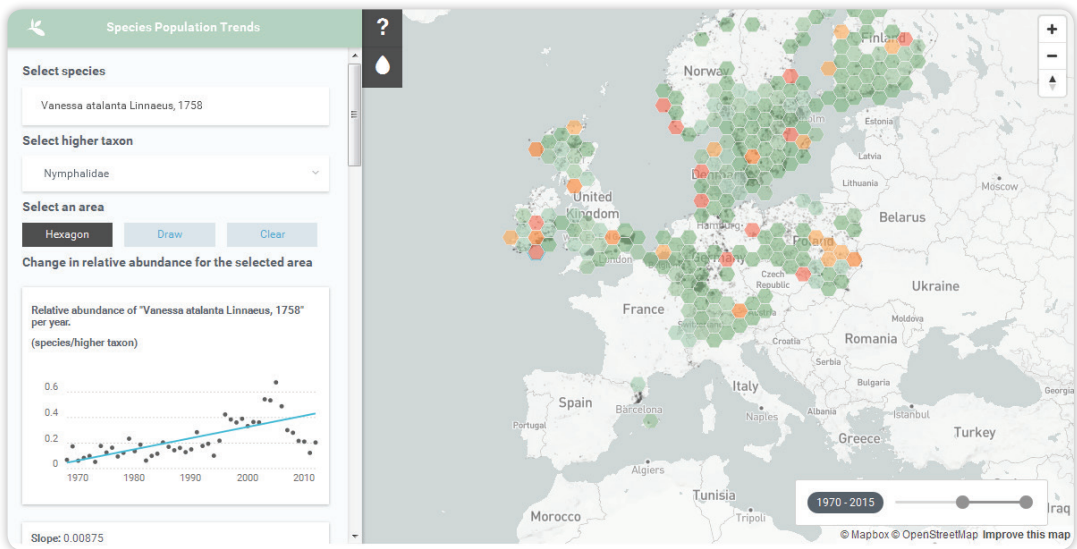


Figure 3. Species Population Trend Browser helps to analyse species trends in GBIF mediated data. Example: the butterfly species ‘Red Admiral’ *Vanessa atalanta*, green: increasing populations, red: declining.

PLUТОF

Data mobilisation and curation platform



Urmas Kõljalg

✉ urmas.koljalg@ut.ee

Overview

PlutoF platform (<https://plutof.ut.ee/>) provides online workbench and computing services for the biology and related disciplines (Fig. 1). The purpose of the platform is to provide synergy through common modules for the taxon occurrences, classifications, geography, projects, agents, analytical tools, etc.

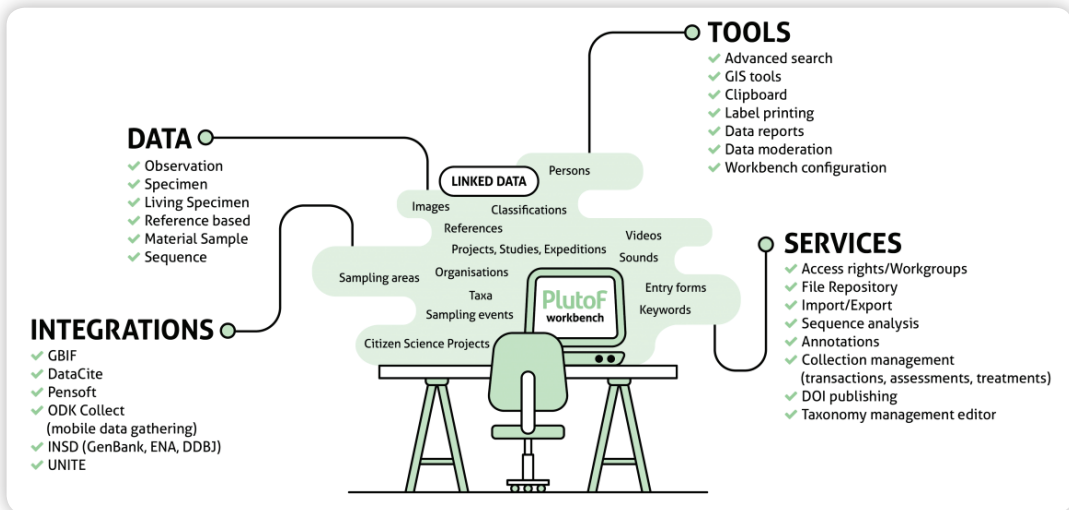


Figure 1. PlutoF provides tools for the full data lifecycle in one place.

Expected advantages

PlutoF secures that all your data are stored, managed and published in a non-fragmented way. The advantage of PlutoF is that the data in your monitoring, ecology, taxonomy, molecular biology, etc. projects are automatically linked and therefore can be searched, analysed and published as a unified dataset. All data from other public or shared projects are also available in the same way. Sharing data with selected users is just a few mouse clicks away. Datasets are easily published through GBIF or Pensoft journals, and Digital Object Identifiers (DOIs) may be obtained through the system. Users can also create a variety of reports and checklists.

Applicability

Useful platform for biological research, nature conservation, data collection management, and monitoring biodiversity. Easy to engage citizen scientists, teachers and children in your biodiversity projects.

Potential users

Research, private, or governmental institutions, workgroups, researchers, citizen scientists, and public servants who need to manage, share and publish their biological/bio-diversity data.

Examples of tools

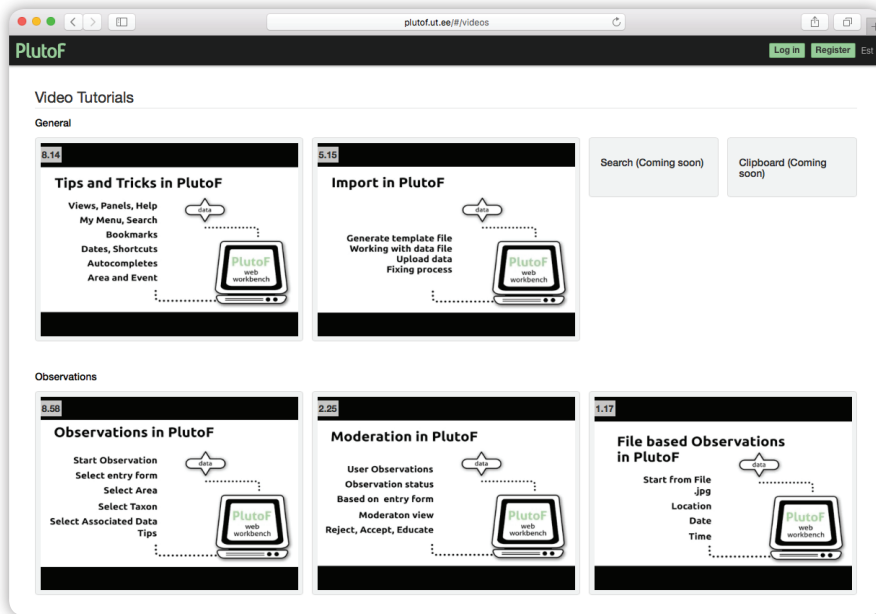


Figure 2. Screenshot with selection of video tutorials accessible at <https://plutof.ut.ee/#/videos>.

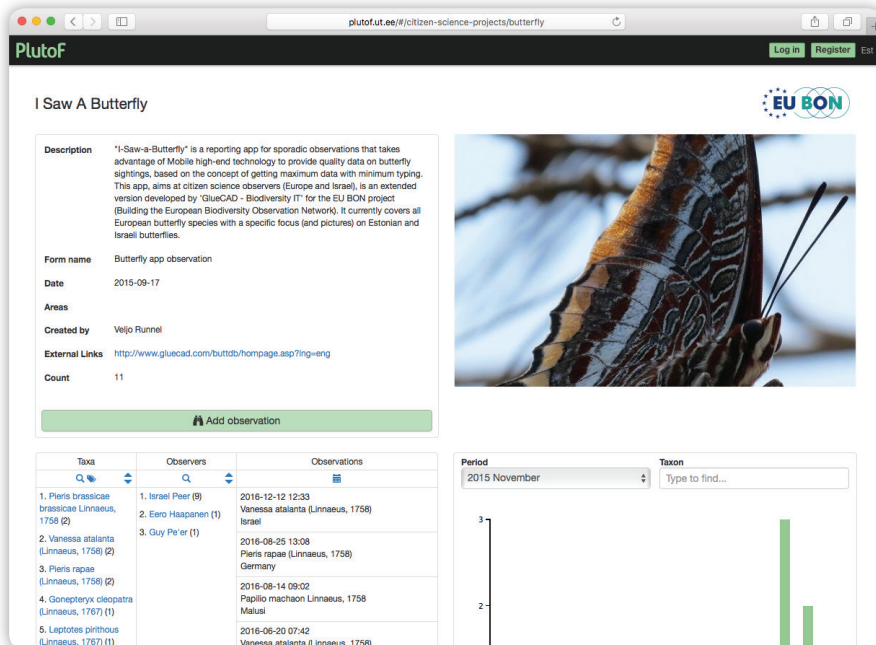


Figure 3. Screenshot from citizen science project, "I Saw A Butterfly" (<https://plutof.ut.ee/#/citizen-science-projects/butterfly>).

GBIF - IPT

Supporting the publication of
sampling-event data



Kyle Braak

✉ kbraak@gbif.org

Tim Robertson

✉ troberson@gbif.org

Overview

GBIF's **IPT, or Integrated Publishing Toolkit**, is free open source software used to share biodiversity datasets through the GBIF network. Designed for interoperability, the IPT enables the publishing of content from databases, Microsoft Excel spreadsheets, or simple text files using the open **Darwin Core (DwC)** and **Ecological Metadata Language (EML)** standards.

EU BON has supported improvements to DwC Standard (Table 1) that accommodate **sampling-event data** (Fig. 1). By addressing the major shortcomings of DwC occurrence data, the sampling-event core provides a standardized structure for communicating, accessing and sharing data from monitoring surveys and inventories - as well as initiatives like IPBES and GEO BON that rely on such data - through the addition of:

- **Data on abundance**, whether individual counts, percent biomass, dry mass, coverage etc.
- Information that improves the ability to infer the **significance of absence** (Fig. 2).

Many datasets already published as occurrence datasets do contain richer information—particularly those captured as part of carefully planned, structured surveying programmes that:

- **Revisit and resample** locations over time.
- Are comparable to **similar surveys** at other sites.
- Collect **additional measures** like abundance, level of effort, standard methodologies and protocols.

Structuring such information and making it openly available through DwC sampling-event datasets enables data users, tool builders, policy makers and others to:

- Discern patterns more quickly.
- Support new discovery and comparison services.
- Reveal new opportunities for survey and monitoring projects to collaborate.

Table 1. New fields added to Darwin Core standard to support sampling-event data.

Relevant fields	Location in DwC-A
eml:studyExtent	Metadata
eml:samplingDescription	Metadata
eml:qualityControl	Metadata
eml:temporalCoverage	Metadata
eml:taxonomicCoverage	Metadata
eml:creators	Metadata
dwc:samplingProtocol	Event
dwc:habitat	Event
dwc:locationID / dwc:siteID	Event
dwc:siteTreatment	Event
dwc:sampleSizeValue	Event
dwc:samplingTaxaRange	Event
dwc:samplingEffort	Event
dwc:scientificName	Occurrences
dwc:individualCount / dwc:organismQuality	Occurrences
dwc:individualID	Occurrences
dwc:recordedBy	Occurrences
dwc:basisOfRecord	Occurrences

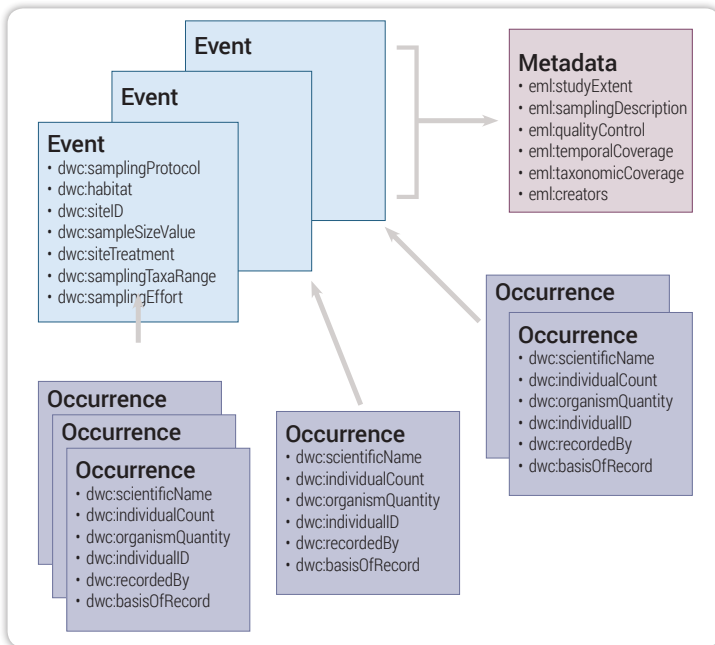


Figure 1. Diagram showing the structure of a Darwin Core sampling-event dataset.

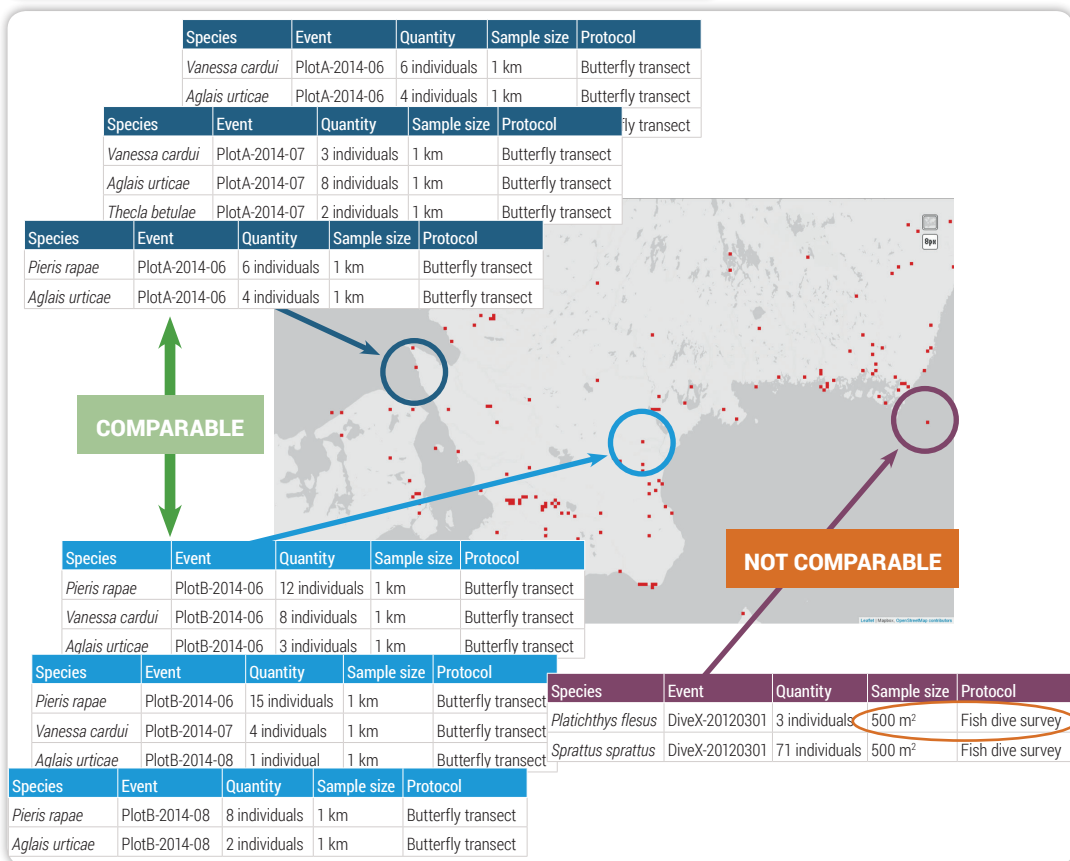


Figure 2. By including consistent information about the sampling protocols used in different survey and monitoring activities, standardized Darwin Core sampling-event datasets enable quicker comparison and more robust inference and analysis of ecological data.

The changes implemented to the Darwin Core through EU BON support mobilization of data for GEO BON's Essential Biodiversity Variables (EBVs), particularly the 'species populations' class. EBVs generally function as intermediate layers between 'raw' data and the indicators established by governments under the CBD and may also support IPBES assessments.

What more is needed now?

- A consistent, referenced catalogue of sampling protocol definitions.
- New proposed DwC terms.
- More extensive and improved documentation for publishers.
- Proposed 'Humboldt Core' standard to catalogue sampling protocols.

Mobilizing data through the IPT

As of February 2017, ten European institutions are publishing nearly 70 sampling-event datasets, enriching the hundreds of millions of occurrence records already shared via IPT installations around the world (Fig. 3). The free open-source software's main features include:

- Ability to connect databases, upload spreadsheets or simple CSV style files for data.
- Use forms to document the dataset descriptive metadata.
- Support for checklists (e.g. taxonomies, species lists etc), occurrence datasets and sampling event datasets. See more on the Dataset classes supported by GBIF.
- Manage processes for registration, licensing, publication and updating datasets for search and download services through GBIF.org.
- Ensure datasets receive a Digital Object Identifier (DOI) to enable consistent citation and help track use. Publishers can configure the IPT to issue DOIs directly through a DataCite account or by using a GBIF-issued DOI.
- A 'one-click' service allows users to convert dataset metadata into a draft data paper manuscript for submission to a peer-reviewed journal.

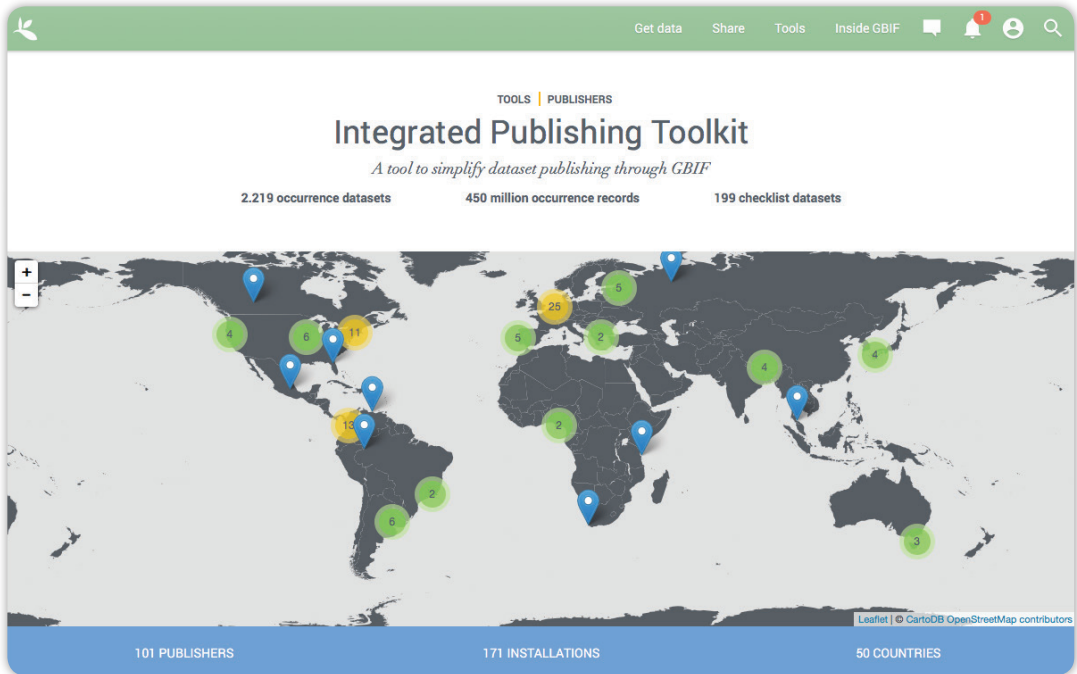


Figure 3. In addition to serving millions of occurrence records, GBIF’s free open-source IPT is the only tool available for publishing sampling-event data.

Learn more about the IPT

- <http://gbif.org/ipt>
- http://www.gbif.org/sites/default/files/gbif_IPT-sample-data-primer_en.pdf
- <http://www.github.com/gbif/ipt/wiki/samplingEventData>

EU BON TAXONOMIC BACKBONE

Access to taxonomic information systems



Andreas Kohlbecker

✉ a.kohlbecker@bgbm.org

Overview

The EU BON taxonomic backbone is an instance of the Unified Taxonomic Information Service (UTIS) (Fig. 1) which is publicly available at <http://cybertaxonomy.eu/eu-bon/utis>. By building the EU BON taxonomic backbone on the foundations of the Data Refinement Workflow library, EU BON can benefit from optimisations and implementations made by BioVeL and vice versa. The service implements a unified protocol for simultaneous access to the following (INSPIRE-conform) taxonomic information systems:

- Pan-European Species directories Infrastructure (EU-Nomen, particularly Euro+Med Plantbase, Fauna Europaea, European Register of Marine Species (ERMS), Index Fungorum)
- EUNIS
- Natura2000

Furthermore it connects the Catalogue of Life (CoL), the World Register of Marine Species (WoRMS), the GBIF Checklist Bank and the Plazi TreatmentBank.

The EU BON taxonomic backbone (Fig. 2) has been developed as a federated search engine that connects an extensible set of taxonomic checklists. Requests sent to the taxonomic backbone are simultaneously distributed to the connected data providers. Result sets returned from the individual checklists are then mapped into a unified data model. Finally, the results are combined and presented to the client, who sent the request.

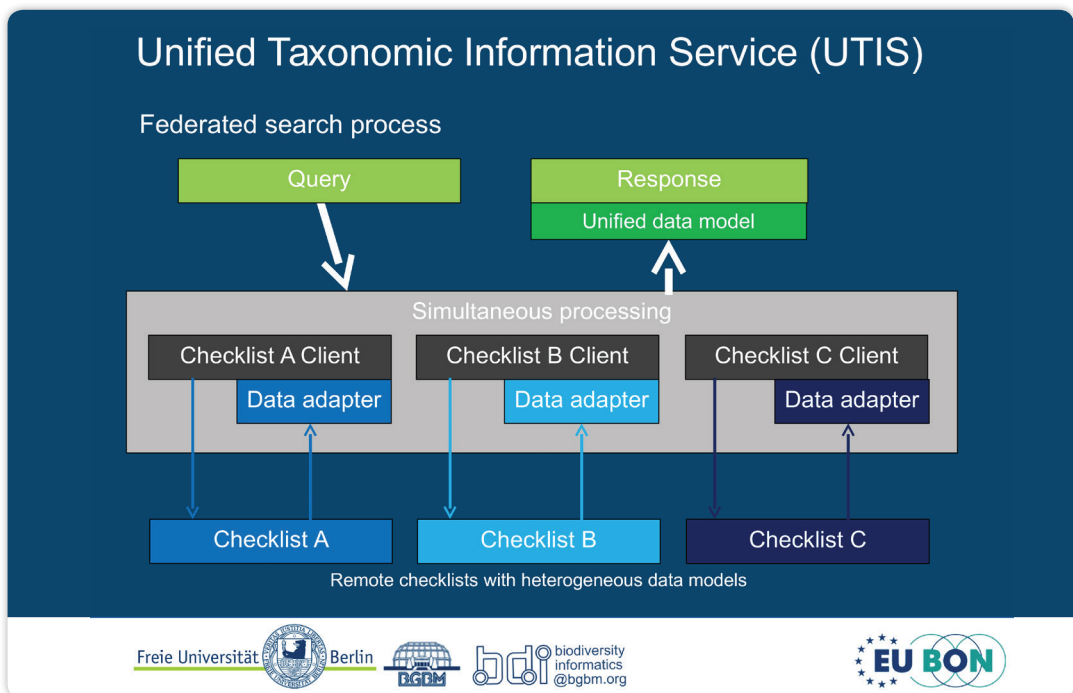


Figure 1. Basic UTIS information flow.

Expected advantages

The overall objective of the EU BON taxonomic backbone is to improve and harmonize European and national level taxonomic backbone data for European plant and animal species. The main sources for the European taxonomic backbone are the databases Euro+Med Plantbase, Fauna Europaea and the European Register of Marine Species (ERMS), that focus on different taxonomic groups. For the EU BON taxonomic backbone, databases were harmonised and continuously updated. A major achievement in EU BON was the implementation of a web service layer providing unified access to relevant taxonomic data sources, in particular to the sources defined in Appendix 3 of the INSPIRE Directive.

Applicability

The EU BON taxonomic backbone has been implemented as a generic engine for federated searches for taxonomic information (UTIS, Unified Taxonomic Information Service). The EU BON taxonomic backbone is an instance of UTIS which is publicly available at <http://cybertaxonomy.eu/eu-bon/utis> (see Fig. 2).

Potential users

Research, agencies, biodiversity informatics developers.

EU BON taxonomic backbone

EU BON UTIS

The Unified Taxonomic Information Service (UTIS) is the taxonomic backbone for the EU-BON project.

[Contact the developer](#)
[Mozilla Public License 2.0](#)

utis-controller : Utis Controller Show/Hide List Operations Expand Operations

GET	/capabilities	capabilities
GET	/classification/{taxonId}/children	taxonomicChildren
GET	/classification/{taxonId}/parent	higherClassification
GET	/search	search

Figure 2. UTIS service documentation web page.

ARPHA-BIODIV

A toolbox for scholarly publishing and
dissemination of biodiversity data



Lyubomir Penev

✉ penev@pensoft.net

Overview

ARPHA-BioDiv – EU BON’s Toolbox for scholarly publishing and dissemination of biodiversity data (Fig. 1) – is a set of standards, guidelines, recommendations, tools, workflows, services, and journals, based on Pensoft’s ARPHA Journal Publishing Platform, designed to ease scholarly publishing of biodiversity and biodiversity-related data that are of primary interest to the EU BON and GEO BON networks.

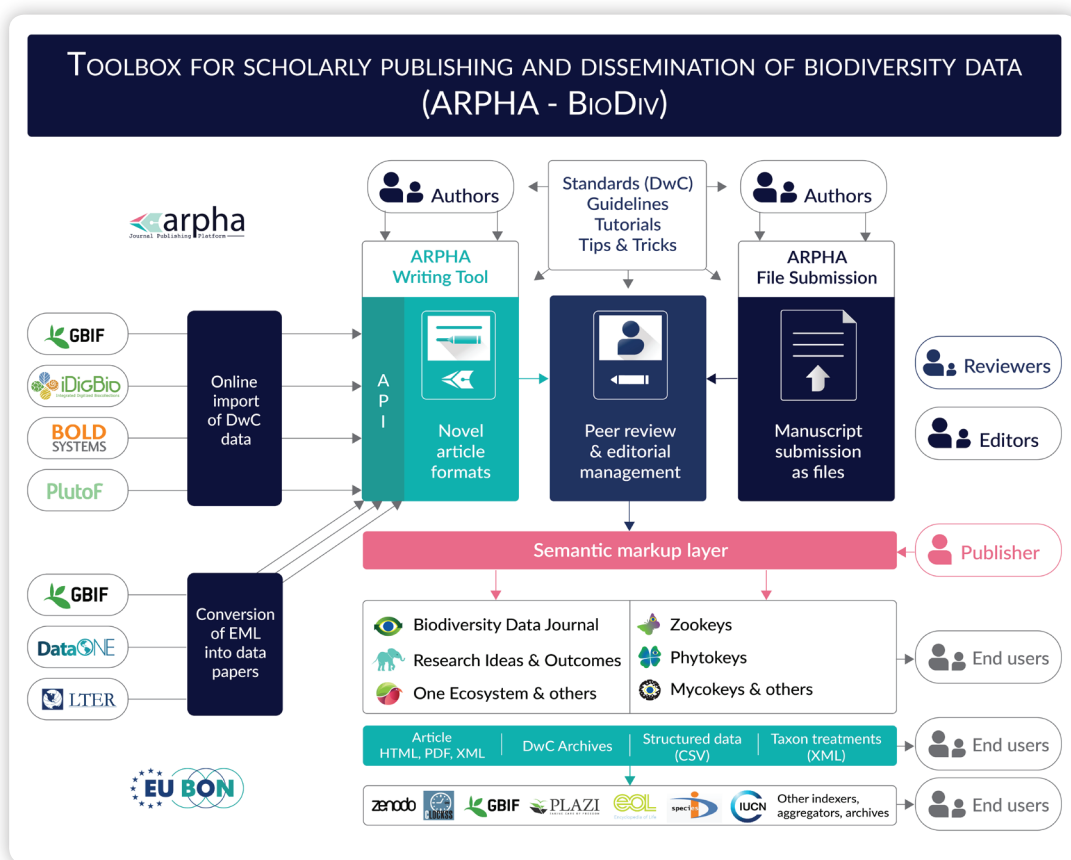


Figure 1. EU BON’s Toolbox for scholarly publishing and dissemination of biodiversity data (ARPHA-BioDiv).

Expected advantages

- Increased efficiency of data publication through novel tools and workflows.
- Mobilisation of non-conventional research outputs via novel article formats.
- Innovative forms of community engagement (data papers, open science collections).

- Machine-readable content available for harvesting and re-use.
- Streamlined creation and import of complex data-rich manuscripts via an Application Programming Interface (API).
- Automated export facilitating data dissemination and re-use.
- Data publishing strategies and guidelines supporting data interoperability.
- Additional incentives for researchers (permanent scientific record, citations) for their effort to collect, maintain and publish biodiversity data.

Applicability

The core element of EU BON’s ARPHA-BioDiv Toolbox is the ARPHA Journal Publishing Platform (<http://arphahub.com/>) developed by Pensoft, along with the associated journals: *Biodiversity Data Journal (BDJ)*, *Research Ideas and Outcomes (RIO)*, and *One Ecosystem*. ARPHA is an innovative publishing solution that supports the full life cycle of a manuscript, from authoring and reviewing to publishing and dissemination. The data publishing strategy of ARPHA aims at increasing the proportion of structured text and data within the article content, so as to allow for both human use and machine readability to the maximum possible extent.

ARPHA stands for Authoring, Reviewing, Publishing, Hosting and Archiving, all in one place. ARPHA consists of two interconnected workflows. A journal can use either of the two or a combination of both (Fig. 2): 1) ARPHA-XML web-based authoring, peer review and publishing, and 2) ARPHA-DOC - Document-based peer review and publishing. The XML-based workflow is currently used by three journals of Pensoft – BDJ, RIO and One Ecosystem. The second, file-based submission workflow, is currently used by 20 other journals published by Pensoft (ZooKeys, PhytoKeys, MycoKeys, and others).

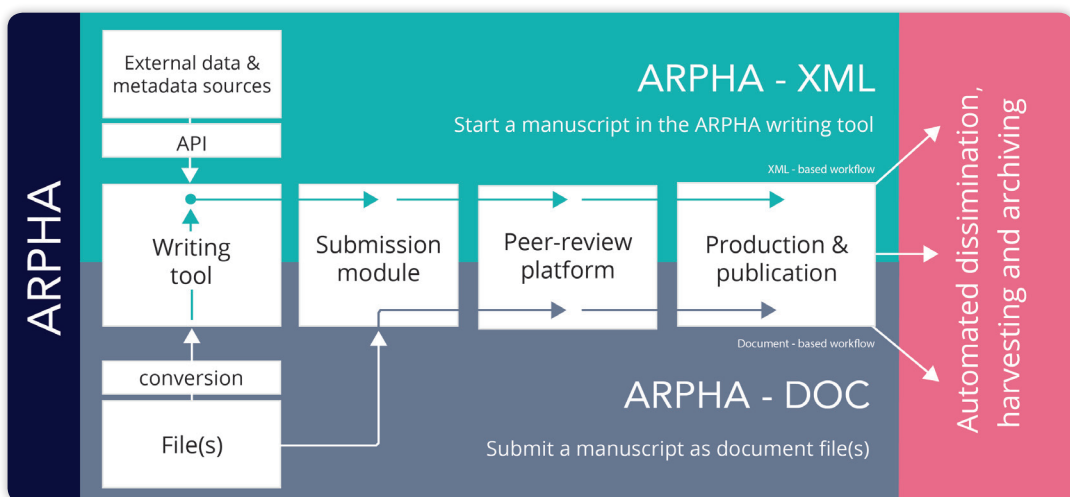


Figure 2. ARPHA consists of two integrated workflows: in ARPHA-XML, the manuscript is written and processed via the ARPHA Writing Tool, and ARPHA-DOC, the manuscript is submitted and processed as document file(s).

ARPHA-BioDiv provides:

- Support through the full life cycle of a manuscript, from writing through submission, revisions and re-submission within a single online collaborative platform.
- Easy online collaborative editing by co-authors and peers.
- Novel, community-based and public, pre-submission, pre-publication and post-publication peer review processes.
- Import of Darwin Core compliant primary biodiversity data from spreadsheet templates or via a manual Darwin Core editor and consequent publication in structured downloadable format.
- Direct online import of Darwin Core compliant primary biodiversity data from GBIF, Barcode of Life, iDigBio, and PlutoF into manuscripts.
- Multiple import of voucher specimen records associated with a particular Barcode Index Number (BIN) from Barcode of Life.
- Automated generation of data paper manuscripts from Ecological Metadata Language (EML) metadata files stored at GBIF Integrated Publishing Toolkit (GBIF IPT), DataONE, and the Long Term Ecological Research Network (LTER).
- Automated export of the occurrence data published in BDJ into Darwin Core Archive (DwC-A) format and its consequent ingestion by GBIF. The DwC-A is freely available for download from each article's webpage that contains occurrence data.
- Automated export of the taxonomic treatments published in BDJ into Darwin Core Archive. The DwC-A is freely available for download from each article's webpage that contains taxonomic treatments data.
- Novel article types in the ARPHA Writing Tool: Monitoring Schema, IUCN Red List compliant Species Conservation Profile, IUCN Global Invasive Species Database (GISD) compliant Alien Species Profile, Ecosystem Inventory, Ecosystem Services Mapping, Policy Brief, Workshop Report, Project Report, and others.
- Nomenclatural acts modelled and developed in BDJ as different types of taxonomic treatments for plant taxonomy.
- Markup and display of biological collection codes against the Global Registry of Biological Repositories (GRBIO) vocabulary.
- Workflow integration with the GBIF Integrated Publishing Toolkit (IPT) for deposition, publication, and permanent linking between data and articles, of primary biodiversity data (species-by-occurrence records), checklists and their associated metadata.
- Workflow integration with the Dryad Data Repository for deposition, publication, and permanent linking between data and articles, of datasets other than primary biodiversity data (e.g., ecological observations, environmental data, genome data and other data types).
- Automated archiving of all articles, published in Pensoft's journals in the Biodiversity Literature Repository (BLR) of ZENODO on the day of publication.

Potential users

Researchers and institutions from all scientific disciplines who want to publish their biodiversity-related data in scholarly articles.

Case studies

Fauna Europaea Special Issue

One of the major data mobilisation initiatives realised by ARPHA and the Biodiversity Data Journal is the publication of data papers on the largest European animal database 'Fauna Europaea'. A new series '[Contributions on Fauna Europaea](#)' was launched in the beginning of 2014. This novel publication model aimed to assemble within a single collection 57 data-papers on different taxonomic groups covered by the Fauna Europaea project, as well as a range of accompanying papers highlighting various aspects of this project (gap-analysis, design, taxonomic assessments, etc.). The first two papers were published on 17 September 2014 and until the end of 2015, 11 articles altogether have been added in BDJ (de Jong et al., 2014, Biodiversity Data Journal, 2: e4034).

EU BON Open Science Publication Pilot

The journal Research Ideas and Outcomes (RIO) was designed to publish all outputs of the research cycle, from research ideas and grant proposals to data, software, research articles and research collateral, such as workshop and project reports, guidelines, policy briefs, and Wikipedia articles, to name just a few. Within RIO Journal, EU BON realised one of the first ever open science collections of project publications entitled: [Building the European Biodiversity Observation Network \(EU BON\) Project Outcomes](#). Up to date, the collection contains 13 publications.

AQUAMAPS

Possible climate change impact on bony fish diversity in Large Marine Ecosystems



Kathleen Kesner-Reyes

✉ k.reyes@fin.ph

Cristina Garilao

✉ cgarilao@geomar.de

Overview

AquaMaps is an approach to generating model-based, large-scale predictions of where aquatic species occur naturally (Fig. 1). The modelling approach uses species occurrence data coupled with expert knowledge to estimate species tolerances with respect to depth, sea temperature, salinity, primary production, sea ice concentration and distance to land, as predictors of species presence. Predictions, expressed as probabilities of species occurrence, are generated by matching species tolerances against observed local environmental conditions to determine the relative suitability of an area for a species. These probabilities are illustrated through color-coded species range maps in half-degree latitude and longitude cell resolution. And now, with future climate data sets, AquaMaps also models the impact of climate change on global distribution of marine species using environmental conditions based on the IPCC A2 emissions scenario for the years 2050 and 2100.

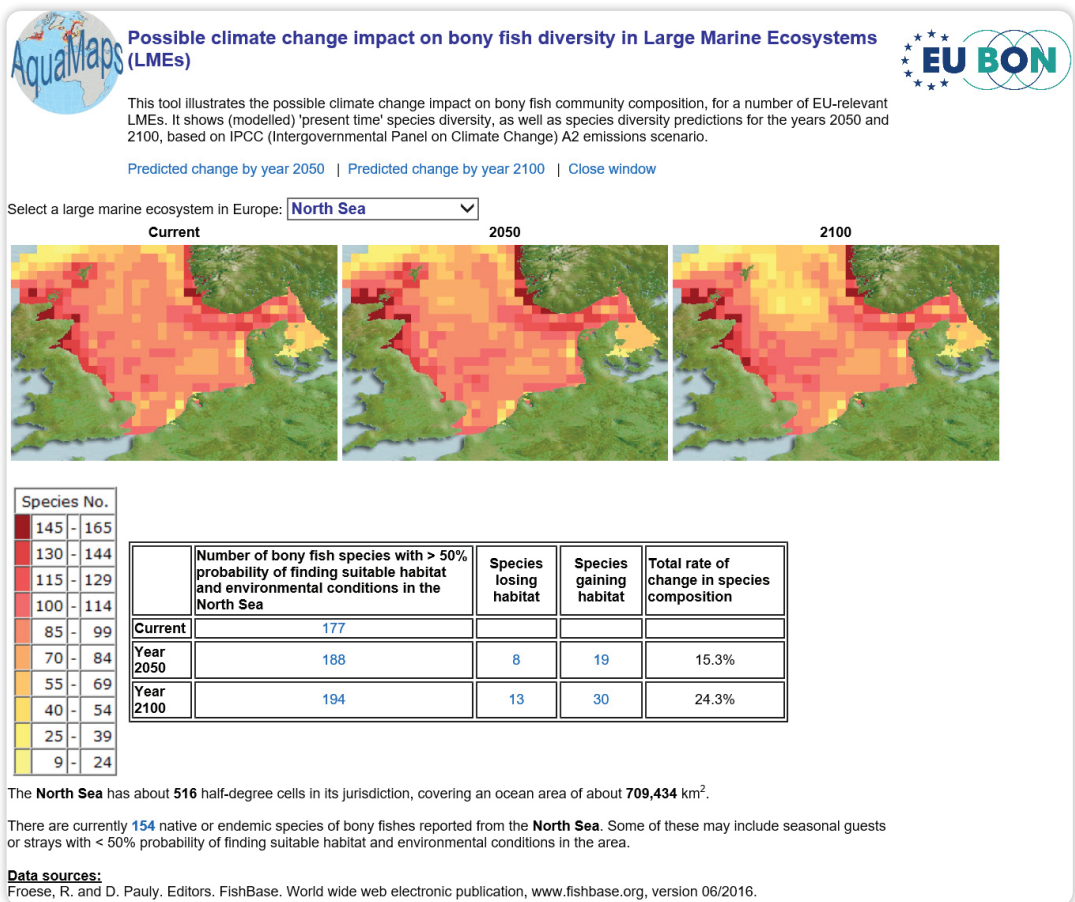


Figure 1. Tool on possible climate change impacts on community composition using the example for bony fishes in the North Sea. Shown are the modelled 'present time' species richness and species richness predictions for the years 2050 and 2100, based on IPCC A2 scenario. Species counts only consider species with >50% probability of occurrence in the North Sea.

This tool shows predictive species richness maps for selected Large Marine Ecosystems (LMEs) in Europe for the current period, 2050 and 2100, and includes links to additional maps showing predicted change in species richness in by the years 2050 and 2100.

Expected advantages

1. **Easy ecosystem search:** Select from the drop-down list of ecosystems; the list is sorted alphabetically by Large Marine Ecosystem (LME) (Fig. 1);
2. **Quick stats:** Includes species richness from literature and predicted periods, estimates of species lost or gained, and the rate of change in species composition (Fig. 1). Figures are clickable and allow access to species lists that deep-link to corresponding FishBase, SeaLifeBase, AquaMaps species page;
3. **Derived maps:** Includes predicted change in species richness per half-degree cell in an LME, and percent change in species composition relative to predicted current species composition illustrate geographical patterns of changes in species diversity in an LME (Fig. 2);
4. **Implications to conservation and policy:** Mapped projections of climate change-induced spatial shifts in species diversity provide insights on possible implications on conservation measures and policy (e.g., establishment and location of marine protected areas in an LME).

Applicability

The Possible Climate Change Impact on Bony Fish Diversity in Large Marine Ecosystems tool (Fig. 2) is publicly accessible through the European Biodiversity Portal (<http://biodiversity.eubon.eu/aquamaps>) and can also be found in the tools section of the Marine AquaMaps for EU BON interface at http://aquamaps.org/am_eubon. Data has been updated based on the 2016 AquaMaps production run. Yearly updates are planned for the last quarter of the year.

Searching, exploring and downloading of maps and the georeferenced data are the expected common usage for this tool. Downloading of georeferenced map data on predicted changes in species richness for an LME is planned to be made available for this tool. The production of infographics showing possible impact in the different LMEs is also foreseen (e.g. <http://wcmc.io/North-Sea>).

Potential users

The target users of this online tools are the general public and policy makers.

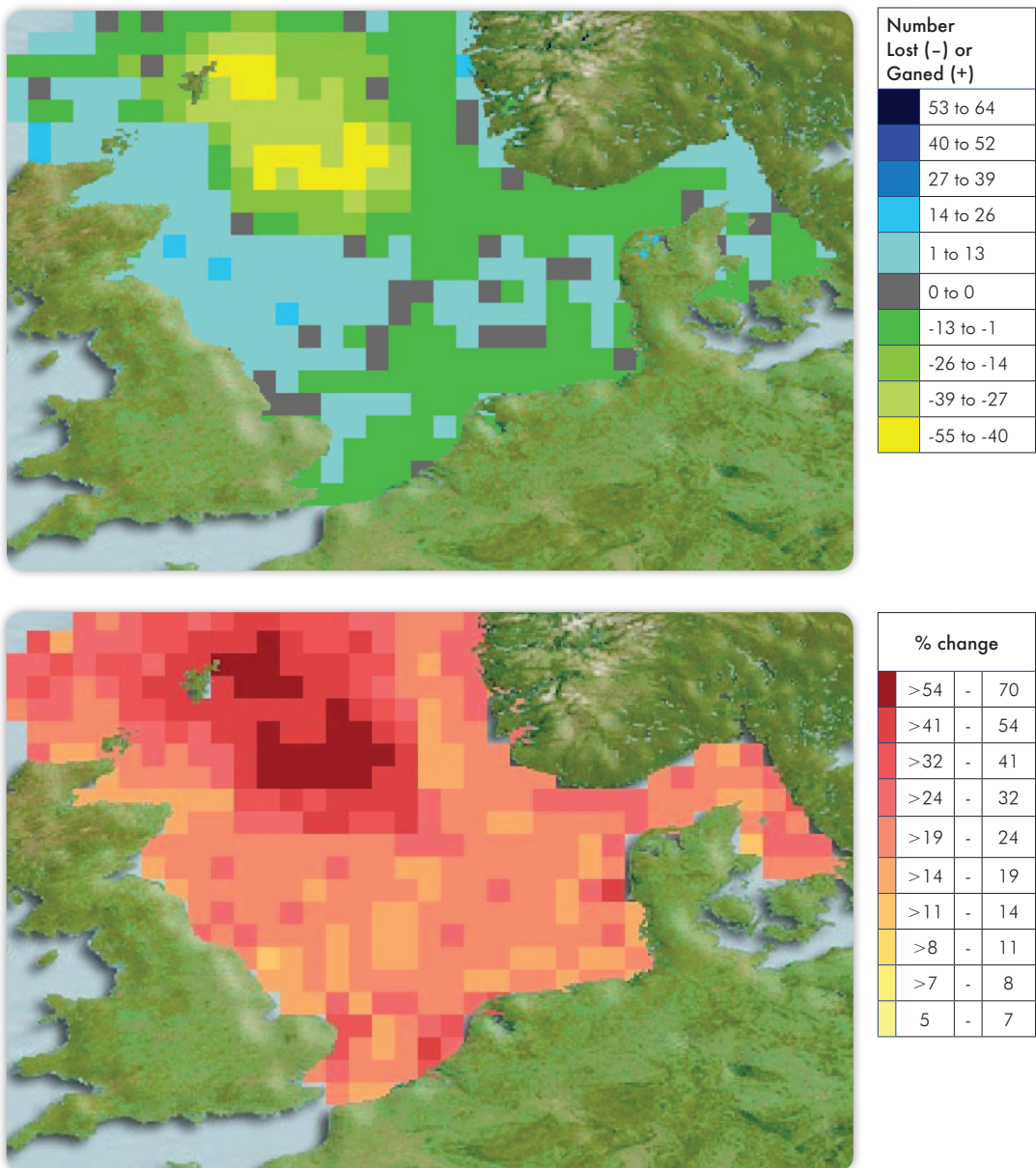


Figure 2. Projected change in species richness of bony fishes in the North Sea by the year 2100 (IPCC A2 scenario). Species included here are limited to those with > 50% probability of finding suitable habitat and environmental conditions in the area. (A) Shows predicted changes in species counts for each half-degree cell. (B) Shows percent change in species composition relative to modelled current species composition.



GEOCAT

Geospatial Conservation Assessment Tool



Steven Philip Bachman

✉ s.bachman@kew.org

Justin Moat

✉ J.Moat@kew.org

Overview

GeoCAT (<http://geocat.kew.org>) is an open source, browser based tool that performs rapid geospatial analysis to ease the process of Red Listing taxa. Developed to utilise spatially referenced primary occurrence data, the analysis focuses on two aspects of the geographic range of a taxon: the extent of occurrence (EOO) and the area of occupancy (AOO). The next version of GeoCAT will also be able to handle multiple species search (Fig. 1).

GeoCAT is built using the latest web-technologies based in JavaScript and HTML5. The result is a responsive and intuitive environment for web-based GIS and conservation analysis algorithms. The tool was built to combine private data provided by the user, public resources such as Flickr, and scientific resources such as GBIF. GeoCAT makes importing geospatial species data simple, by either searching and loading data from the online sources or importing and mapping CSV files.

The Google Maps API and the custom user interface provide a high quality map environment to perform geographic analysis of data location and its quality; the user can delete or move data individually or through filters (e.g. drawing bounding boxes) also defining thresholds for common components of the data such as coordinate precision (Fig. 1). Algorithms for measuring species threat are implemented directly in the browser, avoiding any need to move data to desktop applications or to send the data for server-side processing. The GeoCAT file format streamlines the process of restarting a project by encoding all data, including algorithm parameters, outputs, and application state, into a web syntax called JSON. The file can then be stored by a user for sharing or later use.

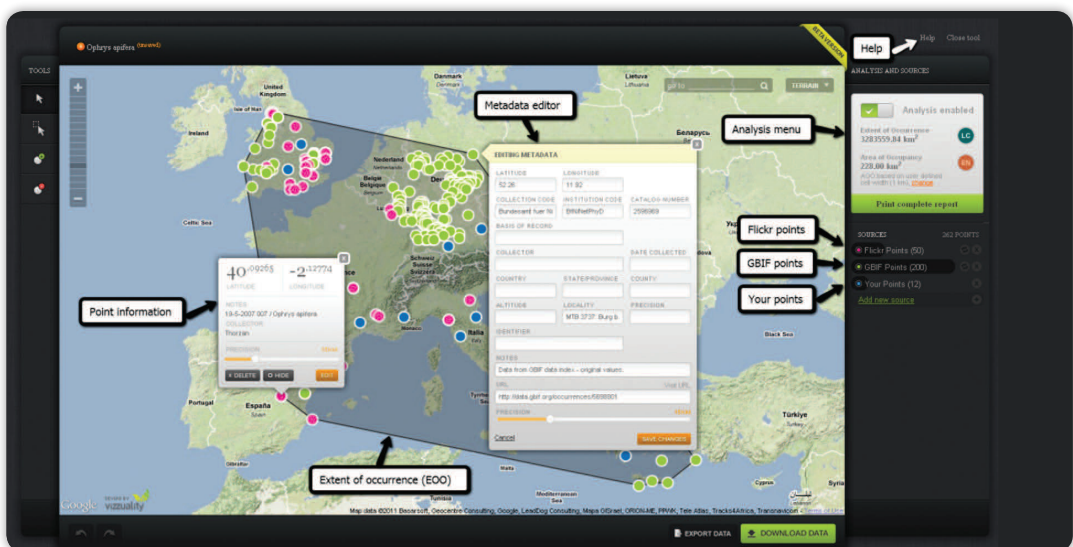


Figure 1. Within a familiar Google Maps environment, GeoCAT users can quickly and easily combine data from multiple sources.

Expected advantages

- A first step toward automating the data handling process of Red List assessing
- Users can quickly and easily combine data from multiple sources such as GBIF, Flickr and Scratchpads as well as user generated occurrence data.
- Analysis of data at the click of a button.
- Instant visualisation of analysis, providing an indication of the Red List threat rating.
- Outputs including the results, data and parameters used for analysis are stored in a GeoCAT file that can be easily reloaded or shared with collaborators.
- Applicability.

Recent estimates suggest there could be 8.7 million (\pm 1.3 million) species on the planet. Even at the lowest estimate, less than 1% (61,914, IUCN 2011) of those species have been formally assessed using the Red List system to determine their conservation status i.e. an assessment of the risk that they will become extinct.

A key factor in the lack of progress in the production of species conservation assessments is the scarcity of user friendly, but powerful, analytical tools which are readily available to scientists and communities to carry out these assessments. Furthermore, large amounts of primary biodiversity data are now available via services such as the Global Biodiversity Information Facility (GBIF), but have yet to be fully utilised for conservation action. With the trend in biodiversity loss increasing across the globe (Secretariat of the Convention on Biological Diversity 2010) it is essential that we speed up the production of assessments. This will enable us to more quickly identify species and regions at greatest risk so that it may guide conservation action. To scale up the production of conservation assessments to the level of mega-diverse groups such as plants and insects, there needs to be significant progress in the development of automated and semi-automated techniques that scientists and other experts can harness.

Benefits

GeoCAT provides a mechanism for data driven conservation assessments in a transparent, repeatable and rapid way through a user friendly environment (Fig. 2). The benefits can be summarised as the following:

- Data driven assessments, giving an auditable data trail i.e. complete transparency of data used for assessments.
- A simple, modern and easy to use interface.
- Accessible - opening up to assessors across the world - only an Internet connection is needed.
- Standardised, automated and repeatable analysis.
- Single-click analysis of Extent of Occurrence (EOO) and Area of Occupancy (AOO).

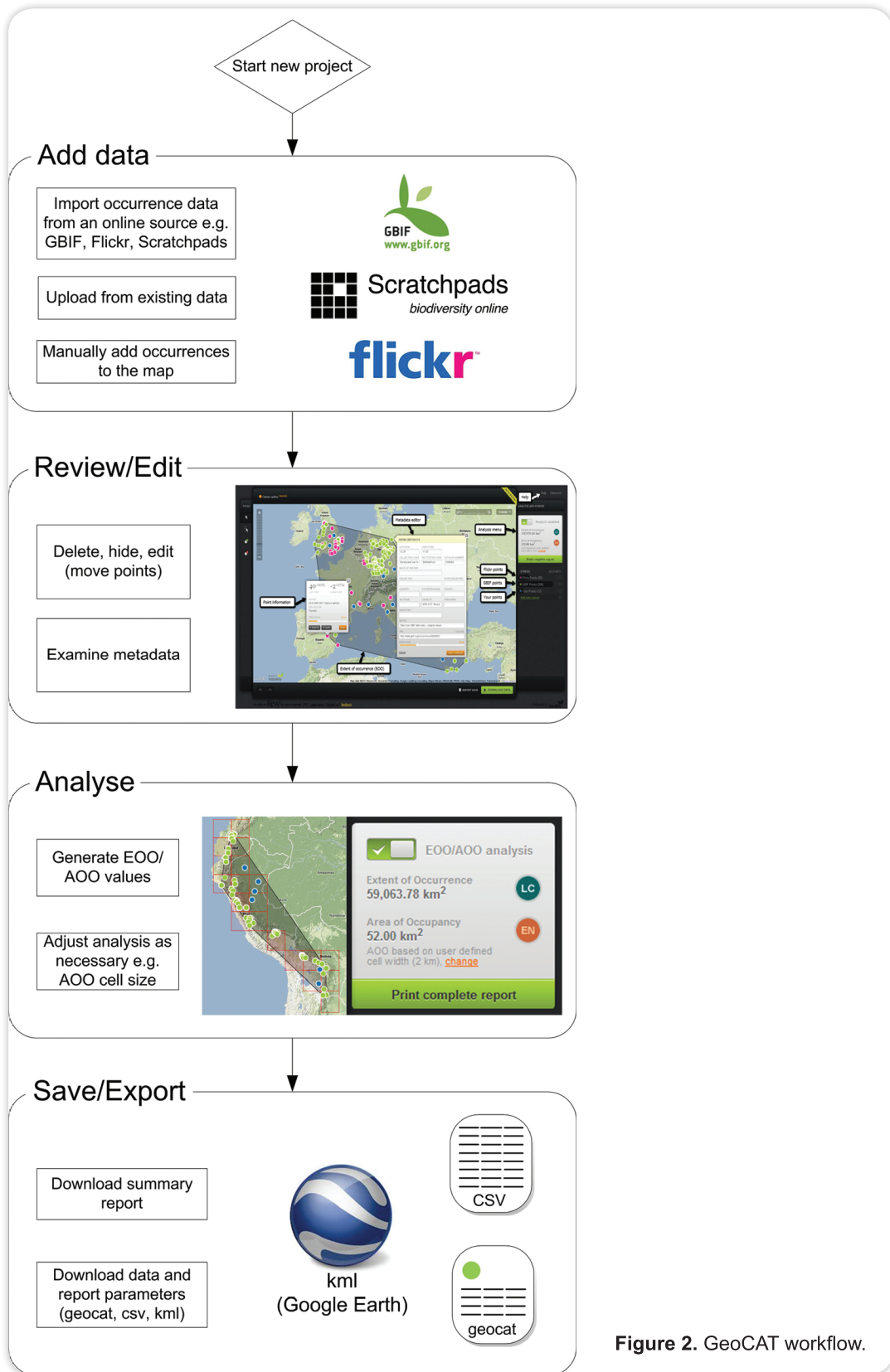


Figure 2. GeoCAT workflow.

- Ability to import occurrence data from online sources such as GBIF or Flickr and other systems such as Brahms and Scratchpads. GeoCAT also allows export and reporting to other formats for further analysis or storage.
- Quick to use and easy to distribute data which can only accelerate the production of Red List assessments.
- Code is open source and development of algorithms are encouraged so the tool can develop towards a powerful automated assessment tool and for other geographic analysis.

GeoCAT responds directly to the growing need for more data driven analytical tools to aid the process of assessing species against the Red List criteria.

Potential users

Researchers and institutions with biodiversity and conservation background.

TREATMENTBANK

Access to taxonomic treatments and
observation records through text and data
mining of taxonomic publications



Donat Agosti

✉ agosti@plazi.org

Overview

Scientific literature contains not only the descriptions and illustrations of biodiversity as we know it, but also citations of the specimens that form the basis for this primary taxonomic study. Because of their contribution to revisions, monographs, descriptions, and other primary taxonomic literature, these specimen records are among the highest quality biodiversity data available. In an age when we have the tools to effectively manage and analyse large quantities of data, and when environmental changes call for data-driven decision-making, the digitization, structuring and extraction of content from scientific literature is needed to provide a more comprehensive supply of information - specially for the many species not well covered in biodiversity databases.

TreatmentBank (<http://treatmentbank.org>) offers the tools to find, convert, store and disseminate these data globally (see Fig. 1). They can be chained together and configured to process automatically the input of new articles or convert entire journal runs. **Golden-Gate Imagine (GGI)** is the tool that allows semi- to fully automatic, interactive extraction of taxonomic treatments, scientific names, named entities, illustrations, bibliographic references, and observation data.

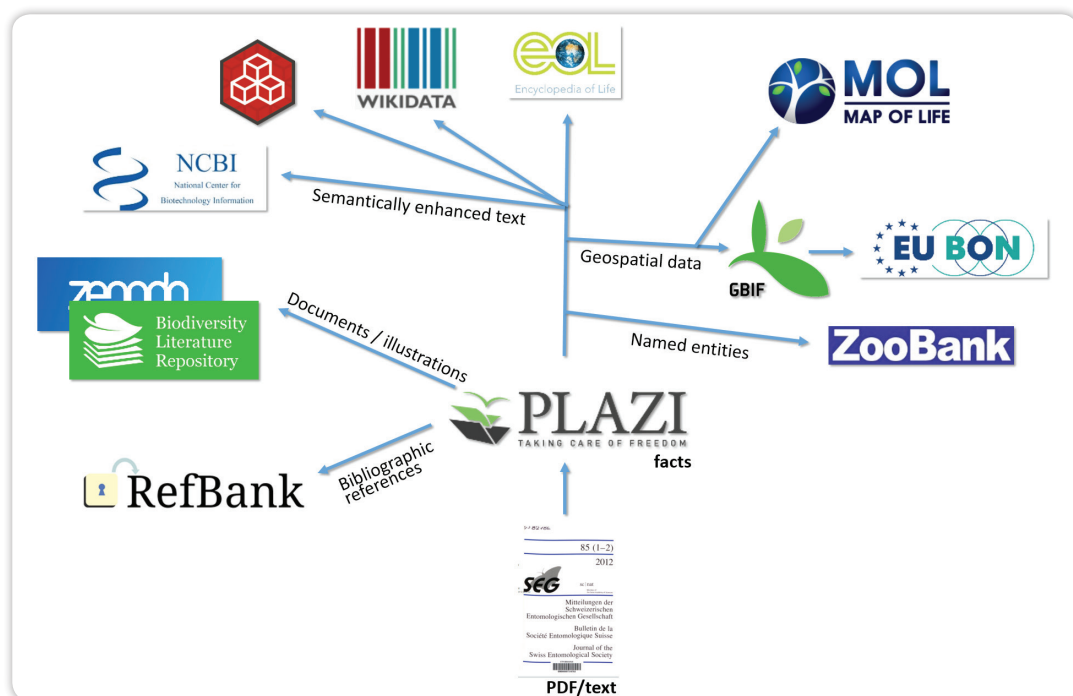


Figure 1. The increased value of articles that are converted into machine-readable documents is that the extracted data are made accessible via TreatmentBank (Plazi).

Expected advantages

1. **Immediate Access** to and wide distribution of biodiversity relevant **data** contained in ongoing published scientific as well as legacy **taxonomic publications**.
2. **Visualization** of content.
3. **Citation: Linking of observation data** to source treatment, illustration and publication, and other external resources such as cited collections and gene sequences.
4. **Update** for **name services** with new scientific names in a timely fashion
5. **Citability of sub-article elements** enabling linking to and from data elements in publications.

Applicability

TreatmentBank is a free and open service. **GoldenGate Imagine** is an open source program that is available and maintained at <http://plazi.org>. It is currently a desktop application. GGI is a generic tool that can be highly customized for journals to allow automatic text and data mining and linking to external resources with a high degree of accuracy and granularity. Currently new articles from 16 scientific journals are automatically processed and made accessible daily, including the submission of extracted illustrations to the Biodiversity Literature Repository at Zenodo/CERN.

Services with taxonomic back bones, such as the Global Biodiversity Information Facility (GBIF), National Centre for Biotechnology Informatics (NCBI), Catalogue of Life (COL) or the EU BON taxonomic backbone are currently updated with names linked directly to the source publication.

GGI can be used and extended by users. PDF, HTML or XML can be used as input formats.

Potential users

Users of taxonomic data access: Anybody with an interest in up-to-date data from scientific publications, such as treatments, observation records, illustrations, scientific names of recently discovered, taxonomically revised or rare species linked to the source taxonomic treatment and publication.

Case study

For EU BON, Plazi is supplying the tools to data mine and extract observation records from the published literature (Fig. 2). In its most advanced version, the workflows start with the discovery of new articles when they are published, and later on are imported, processed automatically, made accessible on TreatmentBank, and then sent to GBIF when new treatments are documented, which they will harvest and make accessible for use in EU BON. The respective treatments can be opened in GGI and further processed using a specific tool to discover and parse the observation records. They then will be resubmitted as DarwinCore Archives to GBIF or anybody who would like to receive this data.

Currently, 19,456 articles, 188,044 treatments and 78,782 observation records are available on TreatmentBank, with a daily increase of up to 100 new treatments. 103,768 illustrations have been extracted and made open access at the **Biodiversity Literature Repository** (<http://biolitrepo.org>).

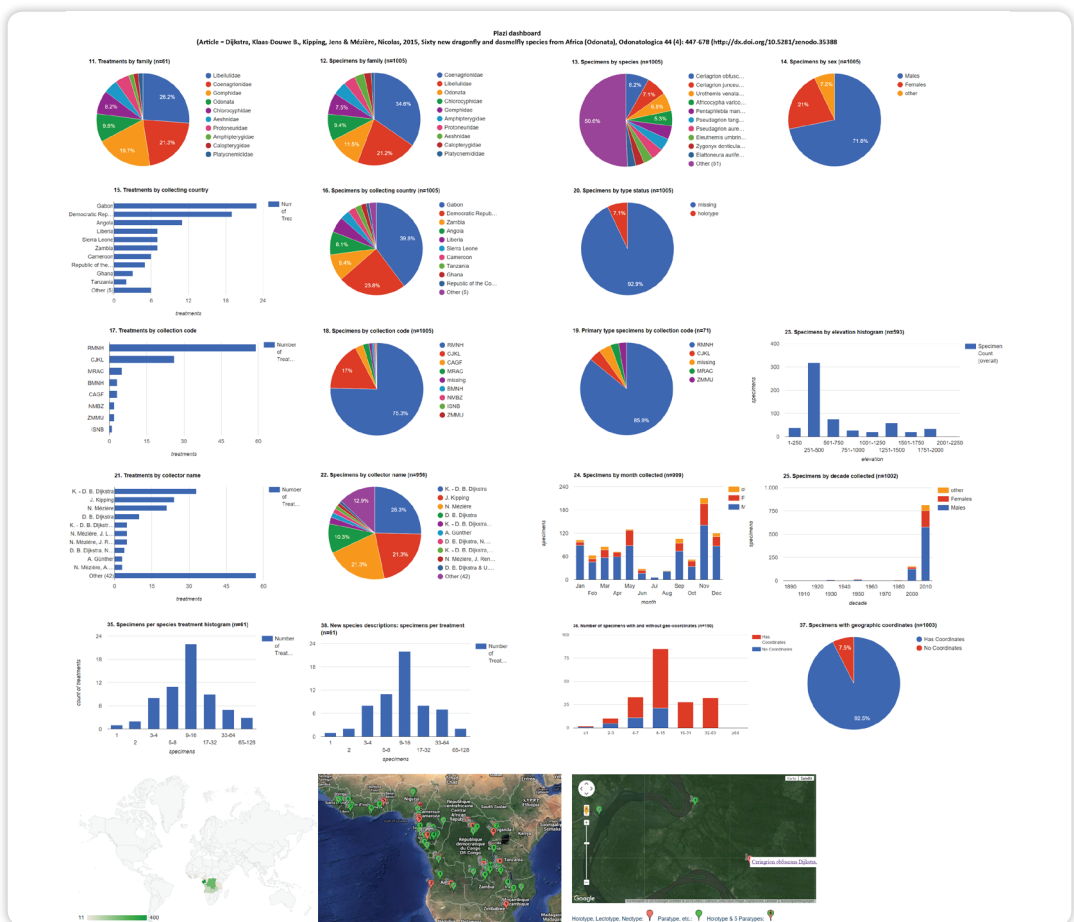


Figure 2. The visualization of the content of a taxonomic article (Odonatologica 44(4), 2015) illustrates the rich data contained as unstructured data within a standard scientific article.

DINA-WEB IN EU BON



Fredrik Ronquist

✉ fredrik.ronquist@nrm.se

Markus Skyttner

✉ Markus.Skyttner@nrm.se

Overview

The DINA-Web project (<http://dina-project.net/>) develops an open-source web-based **D**igital information management system for **N**atural history collections. It supports collection management – the assembling, managing and sharing data associated with natural history collections and their curation (Fig.1). Targeted natural history collections include zoological, botanical, geological and paleontological collections, living collections, biodiversity inventories, observation records and molecular data.

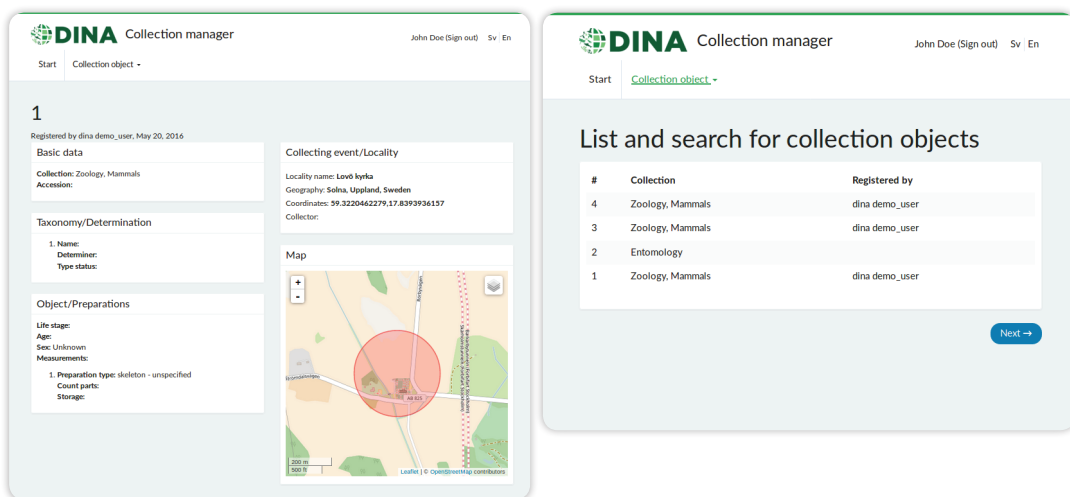


Figure 1. Screenshots of the DINA-Web Collections Management user interface.

Applicability

The DINA-Web system is free and open-source software provided under a license that allows modification, reuse and customizations. It provides a platform for digital natural history collections providing components and datasets to support biological research, nature conservation, collection management and biodiversity monitoring.

The system is developed by the DINA consortium, an unincorporated international partnership among institutions for collaborative open-source development. It is primarily intended for deployments servicing the collection management needs of large institutions or institutional networks within countries or other geopolitical regions.

Expected advantages

The design is modular and based on a micro-services architecture that integrates a set of independently-developed components. A modular design approach facilitates collabor-

orative open-source development of system components by enabling contributions from teams within a distributed consortium of institutions, using modern software development best practices and open-source tools, such as Git and Docker.

Potential users

The primary end-user group includes curators of natural history museum collections, but DINA-Web also accommodates the user requirements from other stakeholders, such as the research community and institutions running the system. The design allows external teams and commercial providers to easily develop plug-and-play modules to extend functionality from the core DINA-Web components.

A modular design requires well-defined modules, each with a robust and well-documented API based on proven technology, and reliable information linking mechanisms extending across modules through persistent unique identifiers.

Examples of tools

More than twenty core and non-core modules – including front-end user interface components, back-end web services and various command line interface tools – are currently available (Fig. 2). Source code is available at <https://github.com/DINA-Web> and binaries are provided at Docker Hub.

Figure 2. Tools and components in the DINA-Web Collections Management system – with source code available at GitHub and with binaries at Docker Hub.

Contact

Dr. Christoph Häuser

Project Coordinator

Museum für Naturkunde

Leibniz Institute for Research on Evolution
and Biodiversity
Invalidenstrasse 43
10115 Berlin
Germany
Tel.: +49 (0)30 2093 8479
Fax: +49 (0)30 2093 8561
✉ christoph.haeuser [at] mfn-berlin.de

Dr. Anke Hoffmann

Project Manager

Museum für Naturkunde

Leibniz Institute for Research on Evolution
and Biodiversity
Invalidenstrasse 43
10115 Berlin
Germany
Tel.: +49 (0)30 2093 8950
Fax: +49 (0)30 2093 8323
✉ anke.hoffmann [at] mfn-berlin.de

www.eubon.eu

