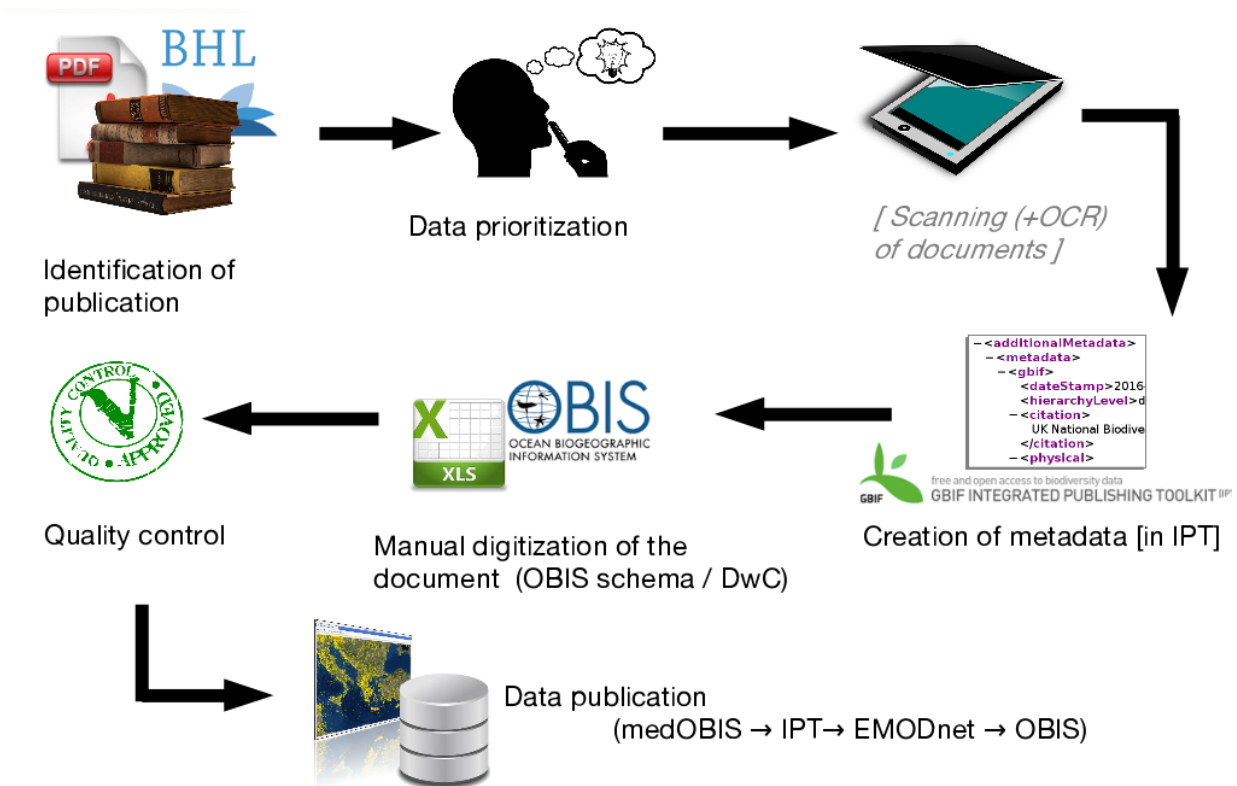


# Biodiversity data import from historical literature assessed in an EMODnet Workshop Report

July 6 2016



Workflow depicting the process of manually extracting data from legacy literature workflow, as currently performed in EMODnet WP4. Credit: Sarah Faulwetter

While biodiversity loss is an undisputable issue concerning everyone on a global scale, data about species distribution and numbers through the centuries is crucial for adopting adequate and timely measures.

However, as abundant as this information currently is, large parts of the actual [data](#) are locked-up as scanned documents, or not digitized at all. Far from the machine-readable knowledge, this information is left effectively inaccessible. In particular, this is the case for data from marine systems.

This is how data managers who implement data archaeology and rescue activities, as well as external experts in data mobilization and data publication, were all brought together in Crete for the European Marine Observation and Data network ([EMODnet](#)) Workshop, which is now reported in the open access journal [Research Ideas and Outcomes](#) (RIO).

"In a time of global change and [biodiversity loss](#), information on species occurrences over time is crucial for the calculation of ecological models and future predictions", explain the authors. "But while data coverage is sufficient for many terrestrial areas and areas with high scientific activity, large gaps exist for other regions, especially concerning the marine systems."

Aiming to fill both spatial and temporal gaps in European marine species occurrence data availability by implementing data archaeology and rescue activities, the workshop took place on 8th and 9th June in 2015 at the [Hellenic Center for Marine Research Crete](#) (HCMR), Heraklion Crete, Greece. There, the participants joined forces to assess possible mechanisms and guidelines to mobilize legacy biodiversity data.

Together, the attendees reviewed the current issues associated with manual extraction of occurrence data. They also used the occasion to test tools and mechanisms that could potentially support a semi-automated

process of data extraction. Long-disputed in the scholarly communities matters surrounding data re-publication, such as openly accessible data and author attribution were also discussed. As a result, at the end of the event, a list of recommendations and conclusions was compiled, also openly available in the Workshop Report publication.

Ahead of the workshop, curators extracted legacy data to compile a list of old faunistic reports, based on certain criteria. While performing the task, they noted the time and the problems they encountered along the way. Thus, they set the starting point for the workshop, where participants would get the chance to practice data extraction themselves at the organised hands-on sessions.

"Legacy biodiversity literature contains a tremendous amount of data that are of high value for many contemporary research directions. This has been recognized by projects and institutions such as the [Biodiversity Heritage Library](#) (BHL), which have initiated mass digitization of century-old books, journals and other publications and are making them available in a digital format over the internet," note the authors.

"However, the information remains locked up even in these scanned files, as they are available only as free text, not in a structured, machine-readable format".

In conclusion, the participants at the [European Marine Observation and Data network](#) Workshop listed practical tips regarding in-house document scanning; suggested a reward scheme for data curators, pointing out that credit needs to be given to the people "who made these valuable data accessible again"; encouraged Data papers publication, for aligning with the "emerging success of open data"; and proposed the establishment of a data encoding schema. They also highlighted the need for academic institutions to increase their number of professional data manager permanent positions, while also providing quality training to

long-term data experts.

**More information:** Sarah Faulwetter et al, EMODnet Workshop on mechanisms and guidelines to mobilise historical data into biogeographic databases, *Research Ideas and Outcomes* (2016). [DOI: 10.3897/rio.2.e9774](https://doi.org/10.3897/rio.2.e9774)

Provided by Pensoft Publishers

Citation: Biodiversity data import from historical literature assessed in an EMODnet Workshop Report (2016, July 6) retrieved 20 April 2024 from <https://phys.org/news/2016-07-biodiversity-import-historical-literature-emodnet.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.