

# SYNTHESYS 3

## Synthesis of Systematic Resources

a Seventh Framework Programme funded project

**Project Number** 312253

**Date:** February 23, 2016

**Work package number.** WP3 Innovation, Impact and sustainability

**Objectives 2:** Impact

**Task 2.2:** Improved access to collections and contribution to ERA

**Partner:** MNCN-CSIC

**Reporter:** Celia M. Santos Mazorra, Isabel Rey Fraile, Manuel Sánchez Ruiz, Antonio G<sup>a</sup> Valdecasas, Josefina Barreiro Rodríguez & Marian Ramos



# 1.- Table of Contents

1.- Table of Contents .....	2
2.- Introduction .....	3
3.- Europeana .....	4
4.- Global Biodiversity Information Facility .....	6
5.- The European Bioinformatics Institute .....	8
5.1.- Workflow to submission of sequence information .....	9
6.- ISBER.....	11
4.1.- Workflow to submit biorepositories .....	11

## 2.- Introduction

The SYNTHESYS Networking Activities (NAs) are integrated packages that support and develop the natural history infrastructure in Europe. The greater integration that will result from these activities will enhance the experience of Access users and other visitors.

The NAs aim to raise awareness of best practices in handling and sampling collections by offering guidelines for the care, storage and conservation of collections. They ensure usability of collections that is enhanced by increased adoption of common standards and protocols. It will create an integrated European resource; bringing together the biological and geological collections held by major natural history museums and other collections based institutions.

This document includes guidelines that develop the sub-task 2.2: *“Improved access to collections and contribution to ERA”* within objective 2 named *“Impact”* linked to the WP3 (NA3): *“Innovation, Impact and sustainability”*

*Improve access to virtual and new physical collections by making all data outputs freely available and open access. For peer-to-peer use the key portal for specimen data will be GBIF. For a wider audience including the public and media the newly accessible data will be made available on established open access outlets, e.g. Europeana.*

*Working alongside NA2, NH institutions will be enabled to make their DNA barcodes available via EBI ([www.ebi.ac.uk/](http://www.ebi.ac.uk/)). In addition, DNA, RNA and proteomes will be made available following the best practices of ISBER.*

### 3.- Europeana

(<http://pro.europeana.eu/>)

Europeana is a digital repository for all kind of present known electronic resources: books, films and videos, pictures, photographs, sound archives, maps, manuscripts, etc., and receives contributions from governmental and private cultural institutions from Europe. The main objective of Europeana is to facilitate access to cultural and scientific heritage through Internet.

Cultural and scientific organizations wishing to share their holdings via Europeana should fill in and send the survey that may be found at the bottom of that page (1) to manifest their interest and become a data partner.

Data partners should complete and submit the Data Contribution Form. This provides essential contact and content information for Europeana.

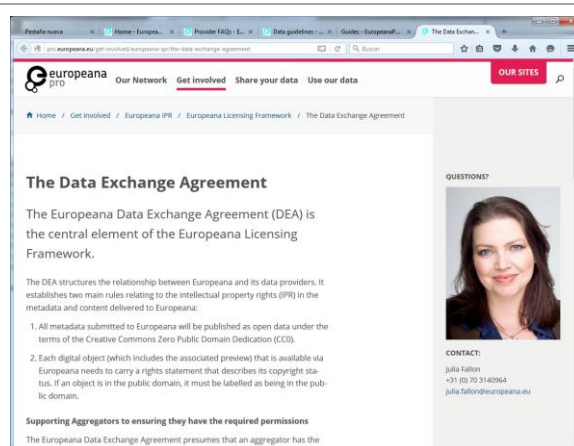
Data partners must sign the Data Exchange Agreement DEA (2) or an accordant agreement with their aggregators, before metadata are published. The provision of metadata under the Creative Commons CC0 1.0 Universal Public Domain Dedication is essential for the functioning of Europeana.

Metadata are accepted for publication after the feedback from Europeana. All requirements on metadata format may be found in the *Europeana Publishing Guide* (3). If the data are not compliant with all aspects of this guide, the dataset will have to be corrected and be re-submitted.

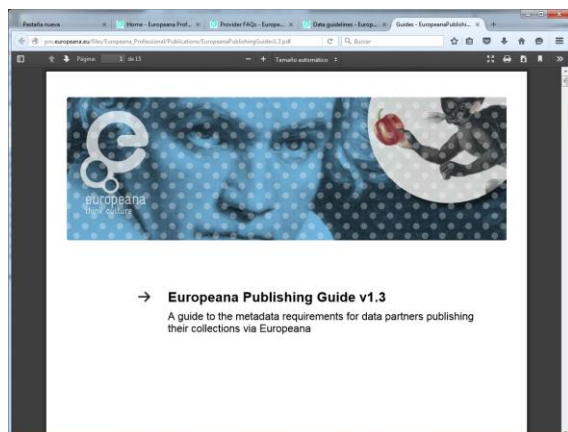
There is a provider FAQs page (4) where answer to common questions and doubts may be found.

There is a webpage (5) with example of data models and useful information on how to share data with Europeana beginning with the EMD, Europeana Data Model.

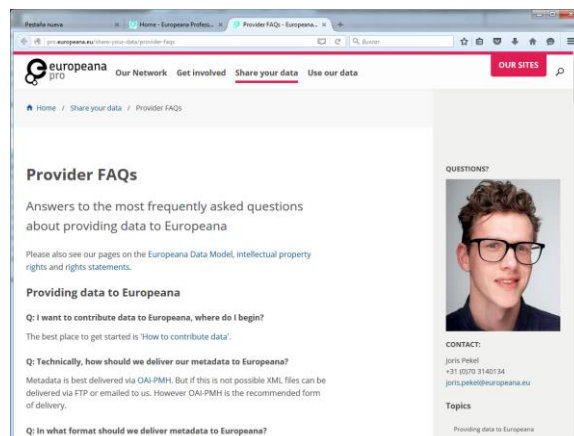
1. <http://pro.europeana.eu/share-your-data/become-a-data-provider>



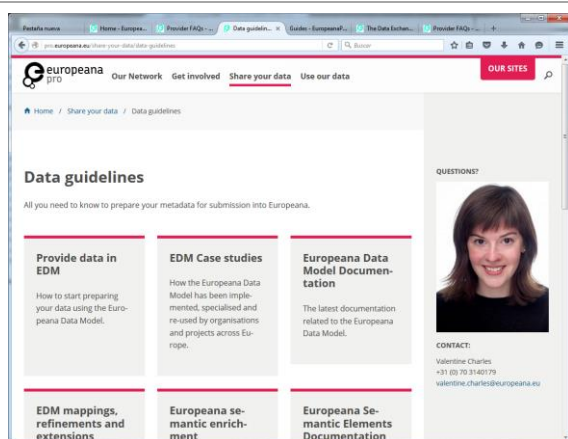
2. <http://pro.europeana.eu/get-involved/europeana-ipr/the-data-exchange-agreement>



3. [http://pro.europeana.eu/files/Europeana\\_Professional/Publications/EuropeanaPublishingGuidev1.3.pdf](http://pro.europeana.eu/files/Europeana_Professional/Publications/EuropeanaPublishingGuidev1.3.pdf)



4. <http://pro.europeana.eu/share-your-data/provider-faqs>



5. <http://pro.europeana.eu/share-your-data/data-guidelines>

## 4.- Global Biodiversity Information Facility

(<http://www.gbif.org>)

The Global Biodiversity Information Facility (GBIF) is an international open data infrastructure, funded by governments. It allows anyone, anywhere to access data about all types of life on Earth, shared across national boundaries via the Internet.

GBIF is in itself a collaboration of many partner institutions and individuals, focused primarily on its network of Participants, including national node institutions and Associate Participant Organizations, and its many data publishers around the world. In addition, for efficient implementation of its work programme through a number of formal partnerships and collaborations at global level, GBIF has partnerships and affiliations with institutions and initiatives involved in common aims and objectives

The GBIF network only publishes data directly from organizations. The Institutions/collections who want to share their catalogues through GBIF, can find contacts for GBIF delegations and nodes staff for their own country on the Participant pages, which can be accessed from (1).

To make biodiversity datasets accessible through GBIF (2) you should:

- Familiarize yourself with publishing tools, workflows and/or partners. The majority of the data now shared with GBIF resides on one of the dozens of installations of our Integrated Publishing Toolkit, or IPT.
- Prepare data for publication. GBIF is built on open standards. GBIF rely primarily on the Darwin Core (DwC) Standard, BioCASE's ABCD standard, the EML standard, and legacy protocols like TAPIR and DiGIR. In any case, to perform some general data quality checking and cleaning is an important step here.
- Request endorsement. GBIF encourage organizations to request endorsement as soon as they agree in principle to share data through GBIF in order to avoid delays in publishing data (3).
- Publish data. When you're using an IPT, simply click the button to 'register' your dataset with GBIF

GBIF has two guides that can cover simple ways of publishing occurrence data:

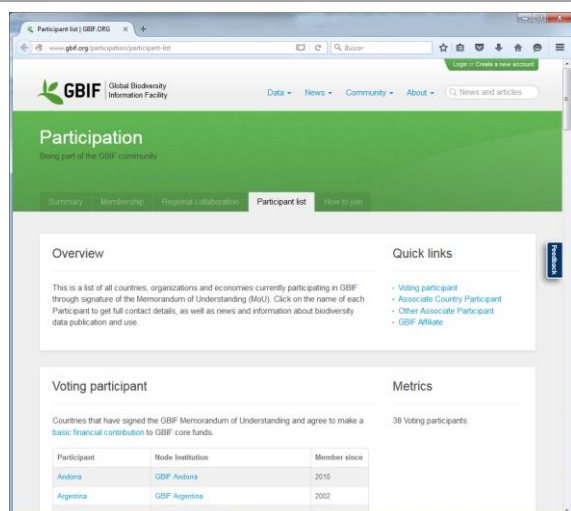
- GBIF (2011). Document map for publishing occurrence data. Contributed by Ko, B. 1 pp. Copenhagen: Global Biodiversity Information Facility. Available online (4).
- GBIF (2011). Document Map for Publishing Metadata. Contributed by Ko, B. 1 pp. Copenhagen: Global Biodiversity Information Facility. Available online (5).

GBIF maintain an extensive knowledge base of tools and documentation, along with detailed manuals for publishers (6).

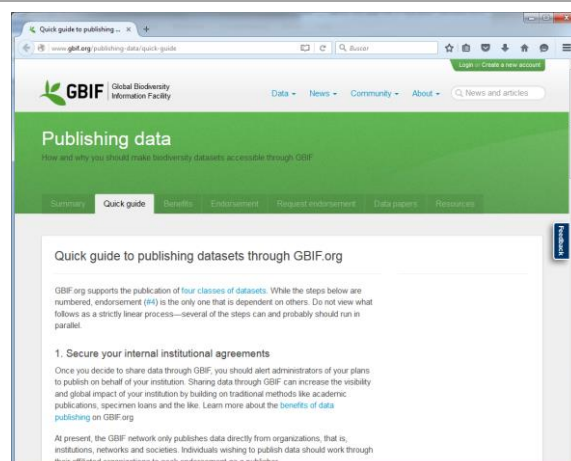
Contact information at several levels can be accessed at <http://www.gbif.org/contact/contact-us>.

Information about Supported classes of datasets can be accessed at <http://www.gbif.org/publishing-data/summary>.

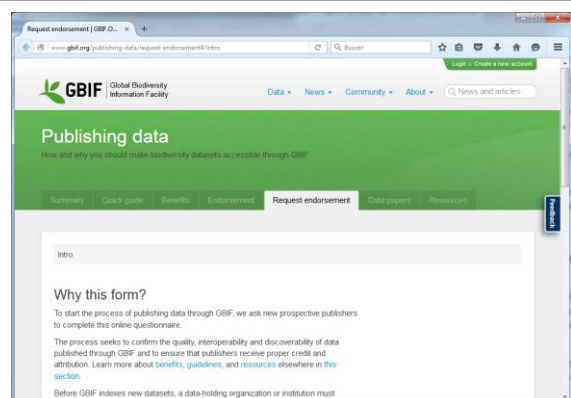
Key information about GBIF can be accessed at <http://www.gbif.org/resources/key-information>.



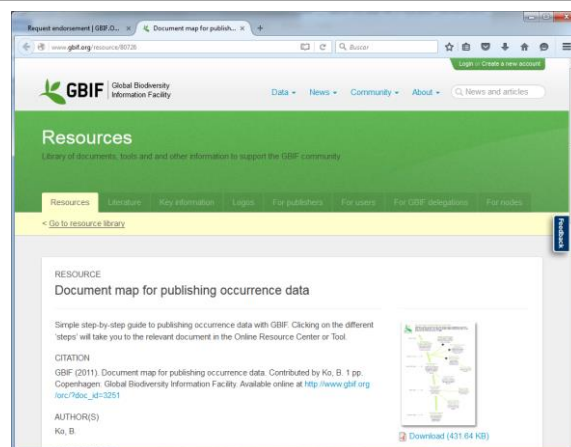
1. <http://www.gbif.org/participation/participant-list>



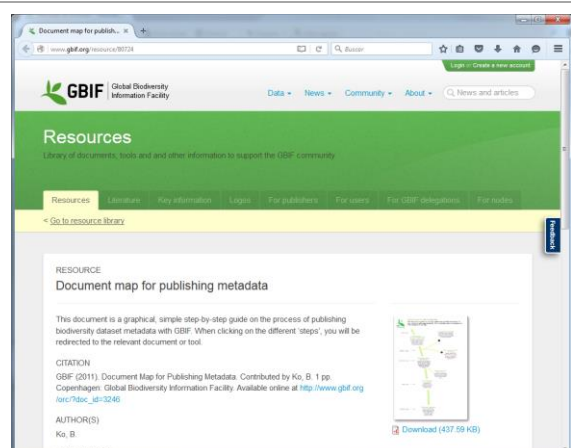
2. <http://www.gbif.org/publishing-data/quick-guide>



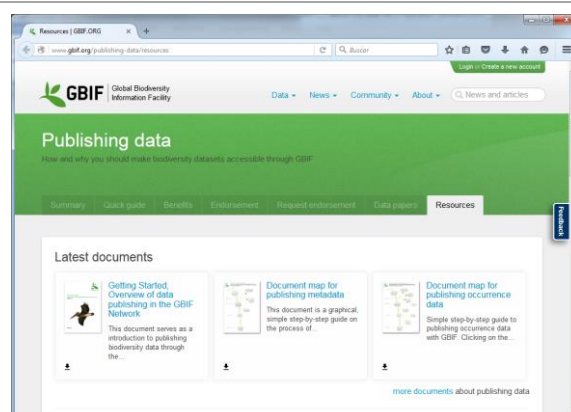
3. <http://www.gbif.org/publishing-data/request-endorsement>



4. [http://www.gbif.org/orc/?doc\\_id=3251](http://www.gbif.org/orc/?doc_id=3251)



5. [http://www.gbif.org/orc/?doc\\_id=3246](http://www.gbif.org/orc/?doc_id=3246)



6. <http://www.gbif.org/publishing-data/resources>

## 5.- The European Bioinformatics Institute

([www.ebi.ac.uk](http://www.ebi.ac.uk))

The European Bioinformatics Institute (EMBL-EBI)<sup>(i)</sup> is a Centre for research and services in bioinformatics, and is part of the European Molecular Biology Laboratory<sup>(ii)</sup>. According to their missions, EMBL-EBI is responsible for providing a range of Bioinformatics services (<http://www.ebi.ac.uk/services>) with molecular data and tools freely available.

EMBL-EBI has developed a platform that works as repository of sequences and genomes (virtual collection), generated by SYNTHESYS institutions and obtained from hundreds of specimens preserved in them.

The European Nucleotide Archive (ENA) <http://www.ebi.ac.uk/ena> provides a mayor comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation, together with the National Center for Biotechnology Information (NCBI)<sup>(iii)</sup> and the DNA Data Bank of Japan, (DDBJ).

The three ENA data tiers are (<http://www.ebi.ac.uk/ena/submit/data-formats>):

- Reads: sequencing machine output including base and colour calls, call qualities and signals.
- Assembly: information relating overlapping fragmented sequence reads to contigs and higher order structures representing complete biological molecules, such as chromosomes.
- Annotation: interpretations of biological function projected onto coordinate-defined regions of assembled sequence in the form of annotation.

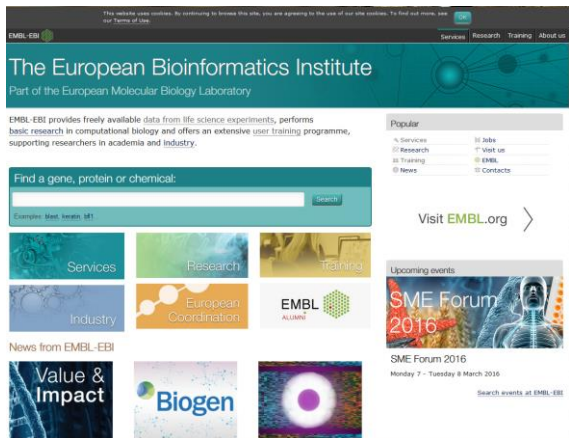
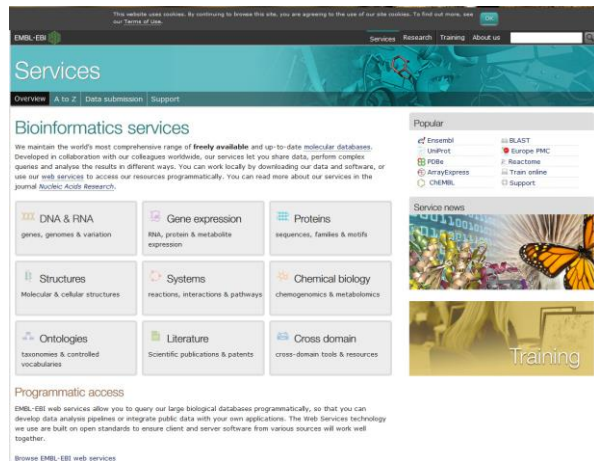
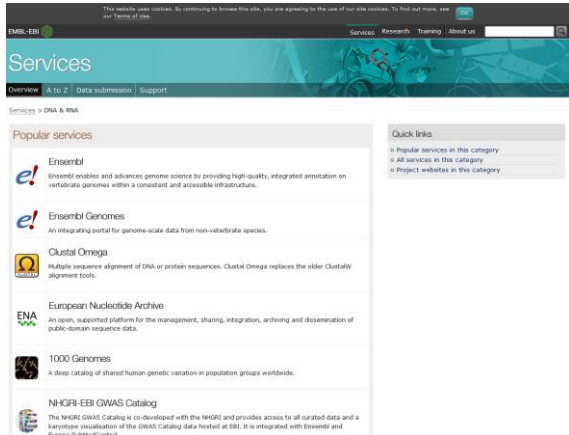
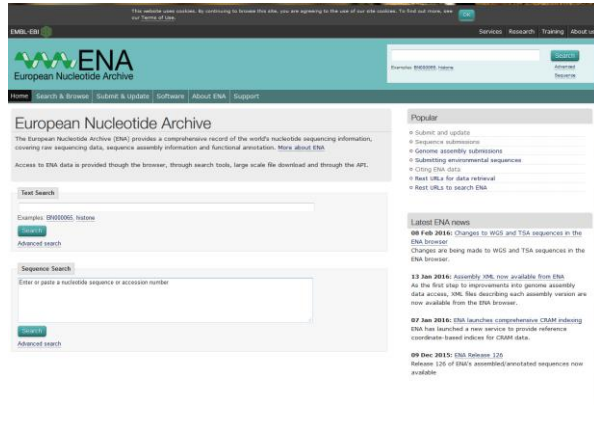
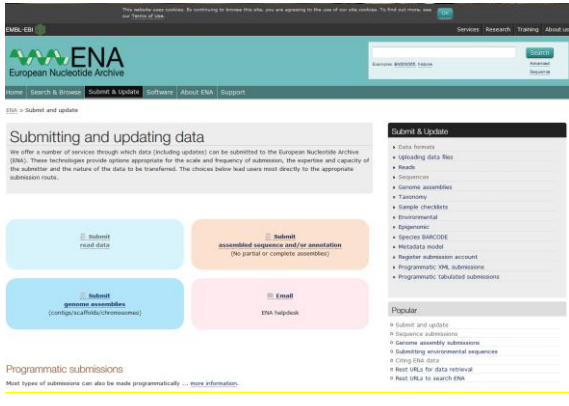
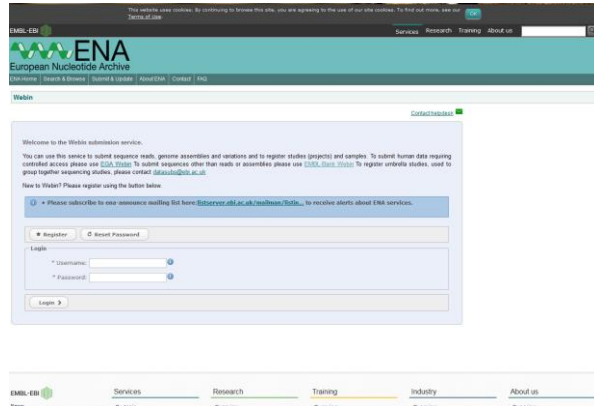
Recommended references:

- (i) Cook, C. E., Bergman, M. T., Finn, R. D., Cochrane, G., Birney, E., & Apweiler, R. (2016). The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic acids research*, 44(D1), D20-D26. [DOI:10.1093/nar/gkv1352](https://doi.org/10.1093/nar/gkv1352)
- (ii) Kanz, C., *et al.* (2005). The EMBL nucleotide sequence database. *Nucleic Acids Research*, 33 (suppl 1), D29-D33. [DOI:10.1093/nar/gki098](https://doi.org/10.1093/nar/gki098)
- (iii) NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. *Nucleic acids research*, 44 (Database issue), D7. [DOI:10.1093/nar/gkv1290](https://doi.org/10.1093/nar/gkv1290)



## 5.1.- Workflow to submission of sequence information

To make raw sequencing data accessible through ENA, you can follow this path

	
<p>1. <a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a></p>	<p>2. <a href="http://www.ebi.ac.uk/services">http://www.ebi.ac.uk/services</a></p>
	
<p>3. <a href="http://www.ebi.ac.uk/services/dna-rna">http://www.ebi.ac.uk/services/dna-rna</a></p>	<p>4. <a href="http://www.ebi.ac.uk/ena">http://www.ebi.ac.uk/ena</a></p>
	
<p>5. <a href="http://www.ebi.ac.uk/ena/submit">http://www.ebi.ac.uk/ena/submit</a></p>	<p>6. <a href="https://www.ebi.ac.uk/ena/submit/sra/#home">https://www.ebi.ac.uk/ena/submit/sra/#home</a></p>

7. <https://www.ebi.ac.uk/ena/submit/sra/#registration>

The following screens are visible only when “username” and “password” fields have been filled:

8. <https://www.ebi.ac.uk/ena/submit/sra/#home>

9. <https://www.ebi.ac.uk/ena/submit/sra/#newSubmission-sequenceChoice-start>

## 6.- ISBER

(<http://www.isber.org/>)

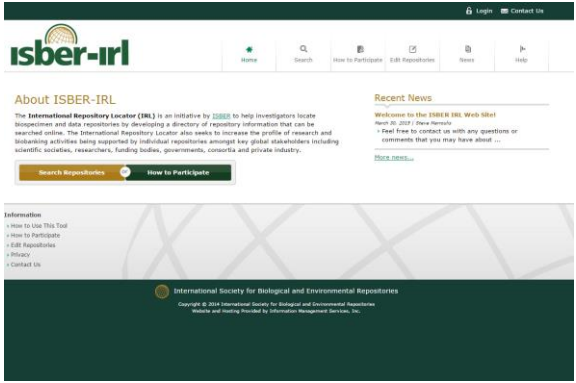
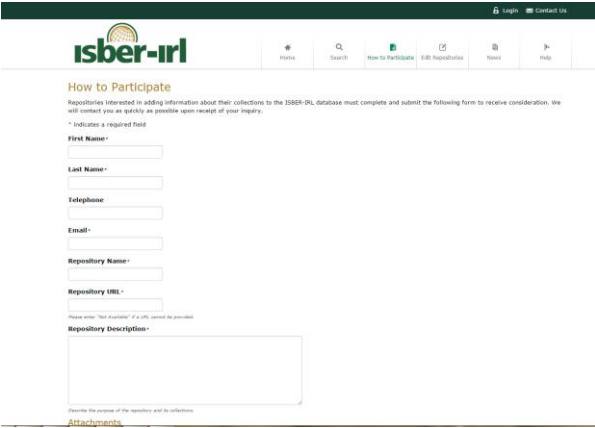
The International Society for Biological and Environmental Repositories (ISBER) is a global forum that addresses harmonization of scientific, technical, legal, and ethical issues relevant to repositories of biological and environmental specimens. The ISBER Best Practices for Repositories presents effective practices for the management of specimen collections and repositories. The International Repository Locator (ISBER-IRL) (1) is an initiative by ISBER. ISBER-IRL helps investigators locate biospecimen and data repositories by developing a directory of repository information that can be searched online. The IRL can be considered an open window to reach new potential users of Natural History Collections from SYNTHESYS partners.

Up to now ISBER-IRL includes the following types of repositories: animal specimen, environmental, human (adult or pediatric), microorganism culture, museum and plant/seed. SYNTHESYS partners interested in adding information about their collections to the ISBER-IRL database must complete and submit a form through the ISBER-IRL web (2); the collections to add (Blood, Buffy Coat, Cell lines, DNA, Plasma, RNA, Serum, Stem cells, others).

ISBER-IRL will send the contact person by mail an account to finalize registration (username and password). After login, repository records can be added to the ISBER-IRL database by clicking the "Add Repository" button on the main page, or by using the "Add Repository" links on the top and bottom navigation menus. Records are created using a simple text box form. To finalize a repository record, click the "Submit" button at the bottom of the screen.

A "Contact us" button may be used to send any message to the system administrator.

### 4.1.- Workflow to submit biorepositories

	
<p>1. <a href="https://www.irlocator.isber.org/">https://www.irlocator.isber.org/</a></p>	<p>2. <a href="https://www.irlocator.isber.org/repositories/inquiry/">https://www.irlocator.isber.org/repositories/inquiry/</a></p>