



LES INITIATIVES POUR LA STANDARDISATION DES DONNÉES D'IMAGERIE MARINE

CATHERINE BORREMANS
REM/BEEP/LEP



ATELIER TECHNIQUE #13 - 9 JUIN 2022

Submersibles
(ROV, HOV, HROV,
AUV, Pagine,...)

OBSERVATORIES

AUTONOMOUS
(video system,
divers,...)

Images

Videos

3D Models

TOOLS

Other
(Images, videos)

Biigle*
(2D Images)

Deep Sea Spy
(2D Images -
citizens)

Adelie
(2D Video)

3D Metrics (3D
Models)

Virtual reality
(3D Models - VR)

AUTOMATIC
ANNOTATION
TOOLS

Standardization
(protocols
development in the
context of the Ifremer
Imagery Working
Group & ALLOHa)

&
Databasing
(to be developed –
operational for Deep
Sea Spy data)

Deep-learning
A.I.

Archives

■ Ifremer Data Centre (SISMER)

■ Personal hard drive (PHD)

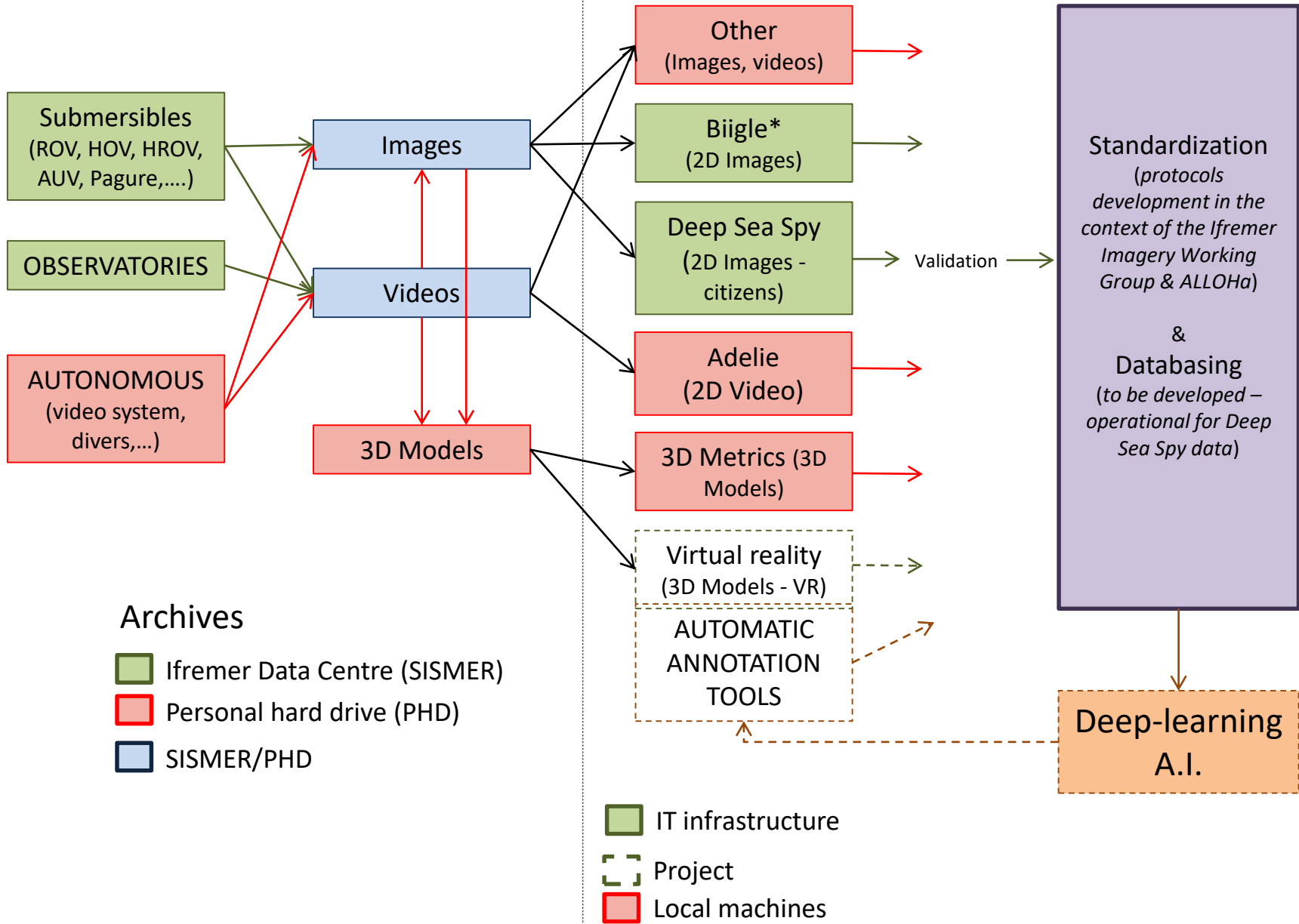
■ SISMER/PHD

■ IT infrastructure

□ Project

■ Local machines

Validation



Challenges

Big data sets to process

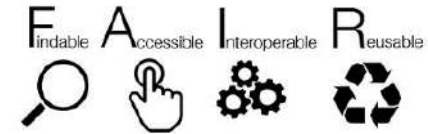
A.I. approaches/machine learning

Variety of annotation tools

Convergence/compatibility

Taxonomy on images

Dedicated identification keys, references and standardization



Towards FAIRness

Develop a set of **common best practices**, build foundation for taxon and morphology resolving methods such as classical microscopy and in situ devices

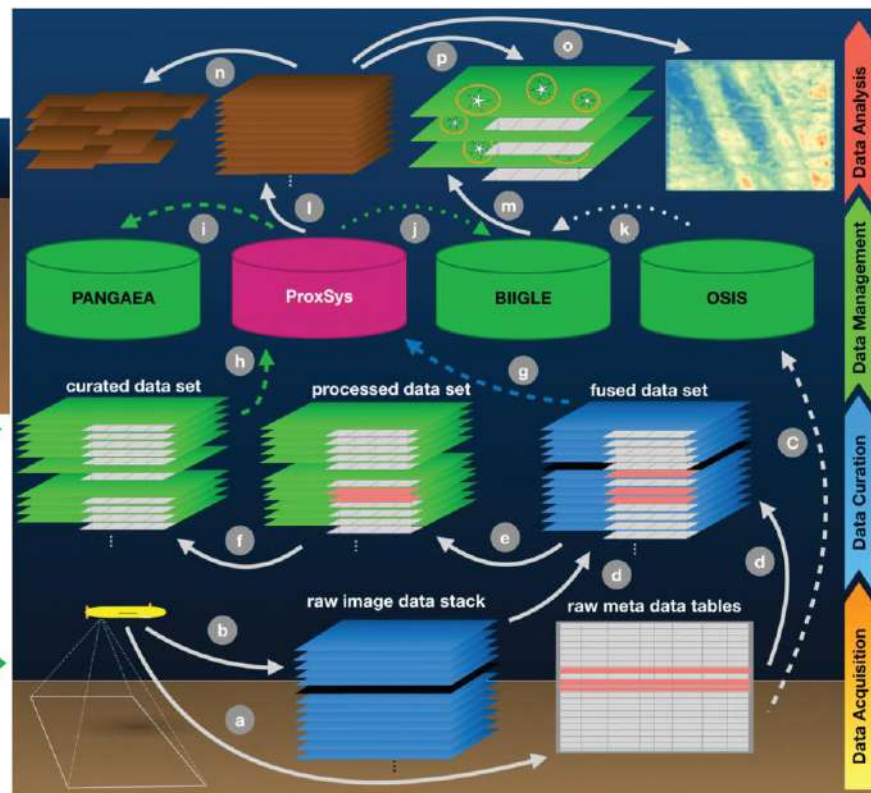
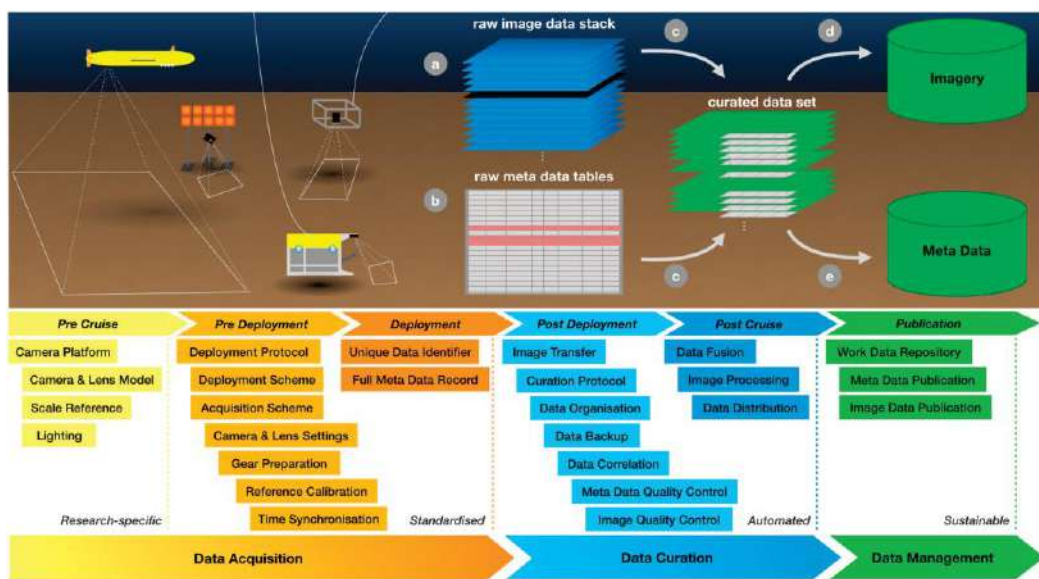
With new observational tools producing datasets of potentially high spatial, temporal, taxonomic and morphological resolution, it is imperative that we develop adaptable **informatics solutions to ensure that these data sets continue to serve the evolving needs of a broad range of users**

Standardisation - Workflow

Schoening, T. *et al.*

An acquisition, curation and management workflow for sustainable, terabyte-scale marine image analysis

Sci. Data 5:180181 doi: 10.1038/sdata.2018.181 (2018)



Applied to the AUV use case

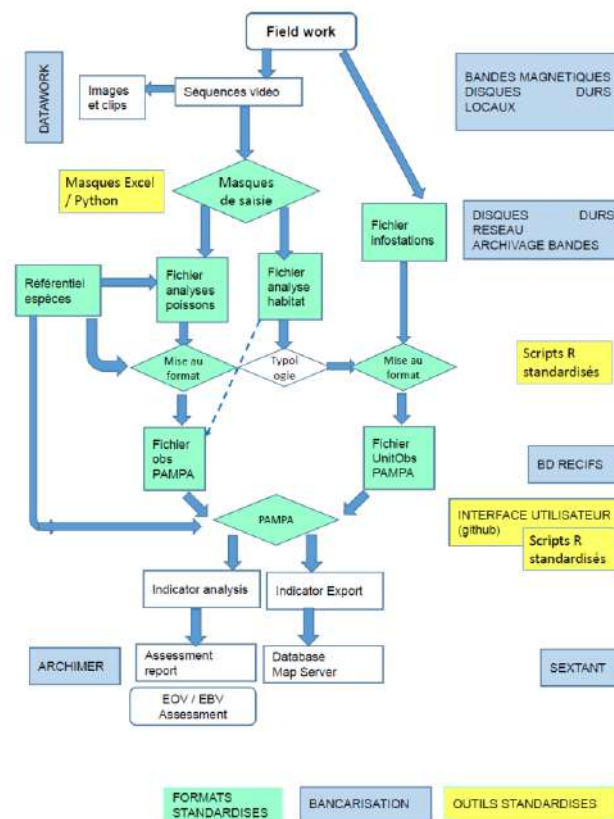
Schematic overview of the proposed image data workflow from acquisition through curation and management

Standardisation - Workflow

STAVIRO and MICADO standardized protocols and workflow - **PAMPA** Format (D. Pelletier)

Standardized protocols and workflow from image collection to Essential Biological Variables/Essential Ocean Variables production (Working Groups **MBON**, D. Pelletier)

Manual **annotation guidelines** for the purpose of machine-learning algorithms development towards automatic identification on submarine images (**ALLOHa**)

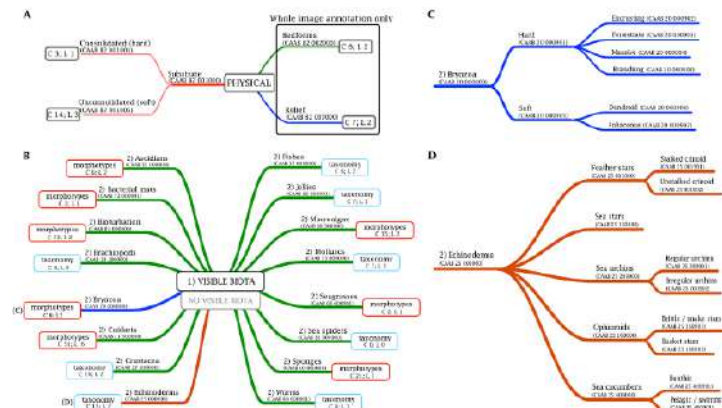


Standardisation - References

Althaus, F. *et al.*

A Standardised Vocabulary for Identifying Benthic Biota and Substrata from Underwater Imagery: The CATAMI Classification Scheme

PLoS ONE 10(10): e0141039. doi:10.1371/journal.pone.0141039 (2015)



Howell, KL. *et al.*

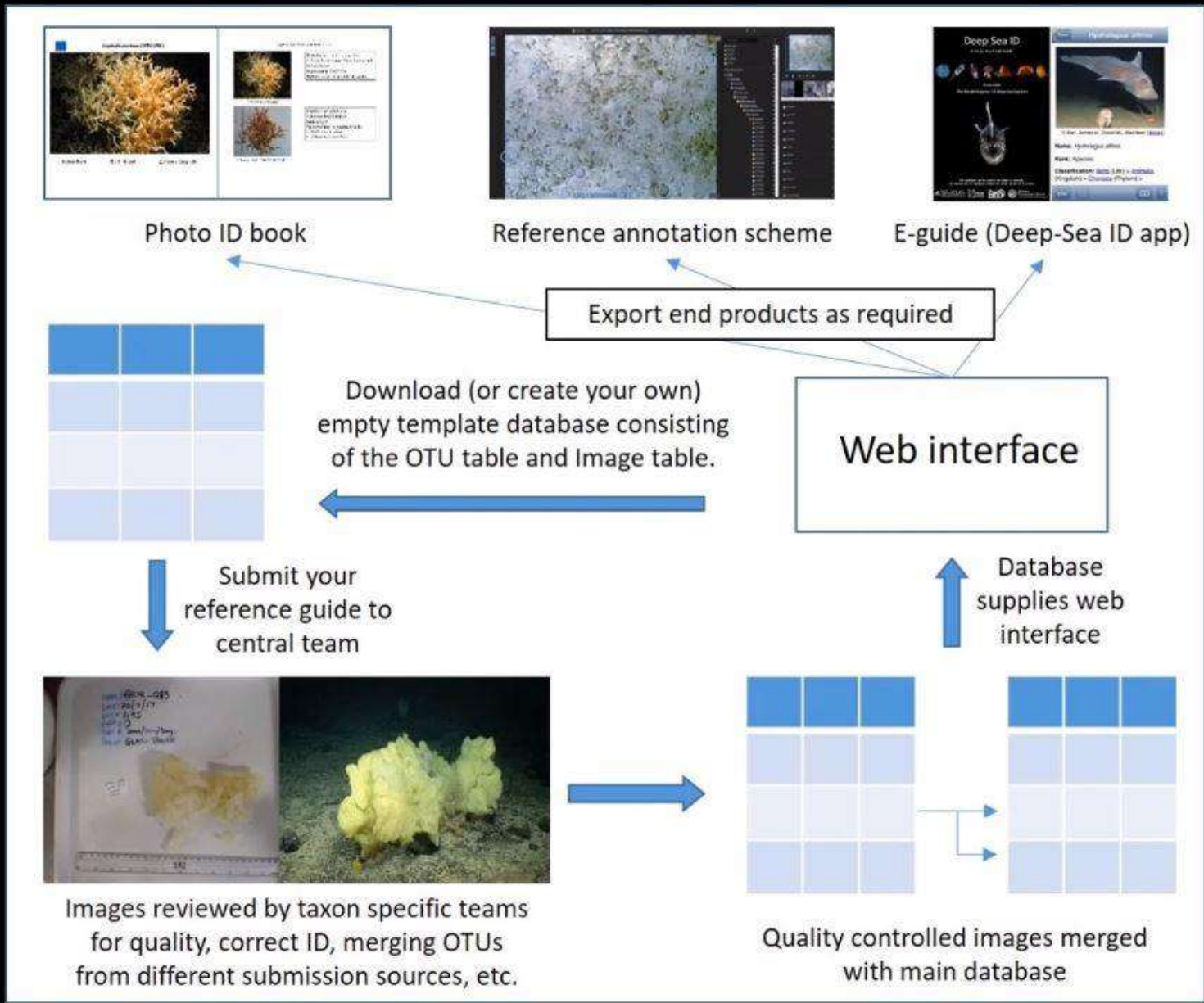
A framework for the development of a global standardised marine taxon reference image database (SMarTaR-ID) to support imagebased analyses

PLoS ONE 14(12): e0218904. <https://doi.org/10.1371/journal.pone.0218904> (2019)

SMarTaR-ID conceptual model

Alt
A St
from
PLoS

emer
ata



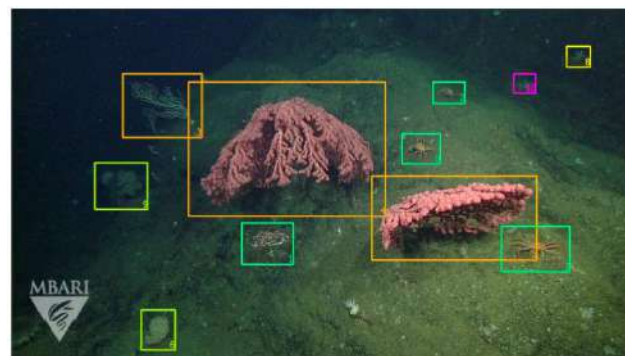
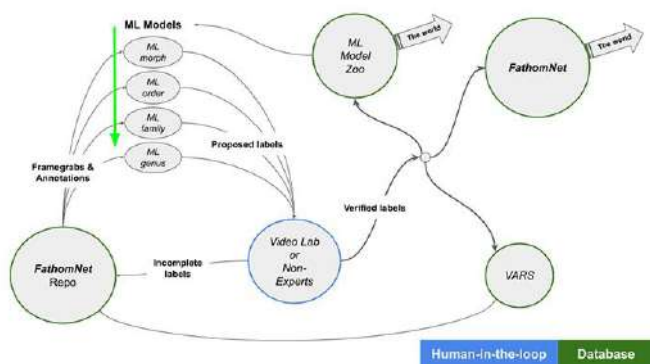
How
A fr
taxo
ana
PLoS

sed

Standardisation - Data

FathomNet is a publicly available database that makes use of existing (and future), expertly curated data from a number of sources (Monterey Bay Aquarium Research Institute (MBARI))

FathomNet will provide much-needed training data (e.g., annotated and localized imagery) for developing machine learning algorithms that will enable fast, sophisticated analysis of visual data.



Source: <https://www.mbari.org/fathomnet/>

Standardisation - Data

iFDO - image FAIR Digital Object

Marine image data collections need a set of standardized metadata to achieve FAIRness of the data for open publication. An entire image set (e.g. deployment, station, dive, mission) requires information on the ownership and allowed usage of the collection. Numerical metadata is required for each image on its acquisition position. It is recommended to provide further optional metadata based on the imaging use case.

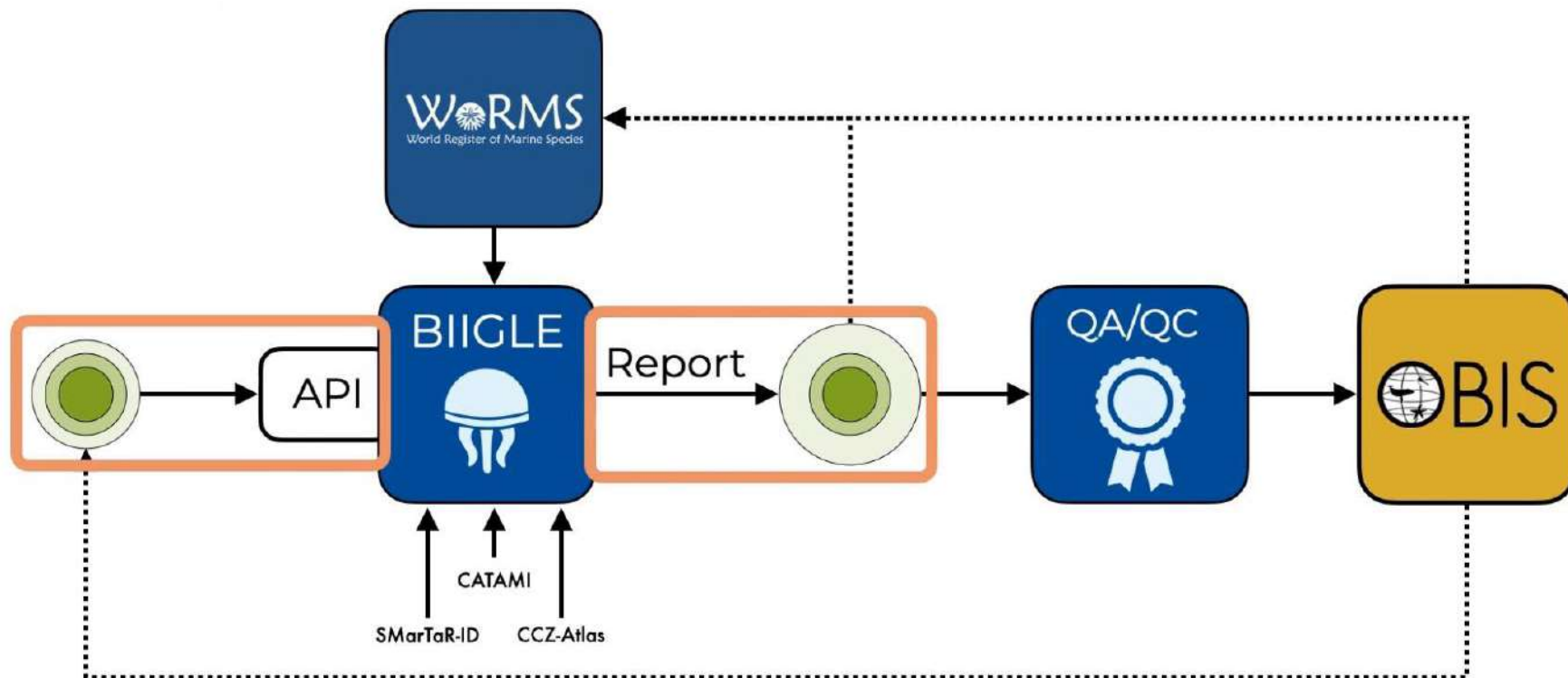
The iFDO standard defines a format to structure such metadata.

Quick facts

- iFDOs are made for photos (still images) and videos (moving images)
- iFDOs consist of an image-set-header and an image-set-items part
- iFDOs group metadata fields in three sections: iFDO core, iFDO capture and iFDO content
- iFDO core fields are mandatory
- iFDO capture and content fields are optional but recommended
- iFDO fields can be mapped to many other metadata standards
- iFDOs make image data FAIR without requiring them to be open, access-restriction remains possible

The iFDO documentation is written with Python implementation in mind

iFDO infrastructure environment



Standardisation Initiatives

ISA Workshop on Enhancing Image-based Biodiversity Assessments To Advance Deep-sea Taxonomy (October 2021) – Kerry Howell (Plymouth Univ.)

Acquisition, data archiving and exchange, reference catalogue

JNCC Annotation Standard Workshop (January 2022) – Kerry Howell et al.

FathomNet Workshop (March-April 2022) – Kakani Katija (MBARI)

i-Atlantic Imaging the Ocean Webinar (May 2022)

Challenger 150 megafaunal image-based technical working group - standards in image analysis (May 2022) – Kerry Howell (Plymouth Univ.)

Image annotations inventory

Ifremer annotations projects:

Biigle

DeepSeaSpy

ImageJ

...

Machine Learning algorithms :

Mask RCNN

Yolov4

Yolov5

...

Needs :

Promotion

Collaboration

ML algorithm integration

Big Data models

But :

No interoperability

ML algorithm are not friendly to use

No collaboration for big data

Proposed solution:

Data structuration

Images generation and cleaning of results

ML algorithm integration

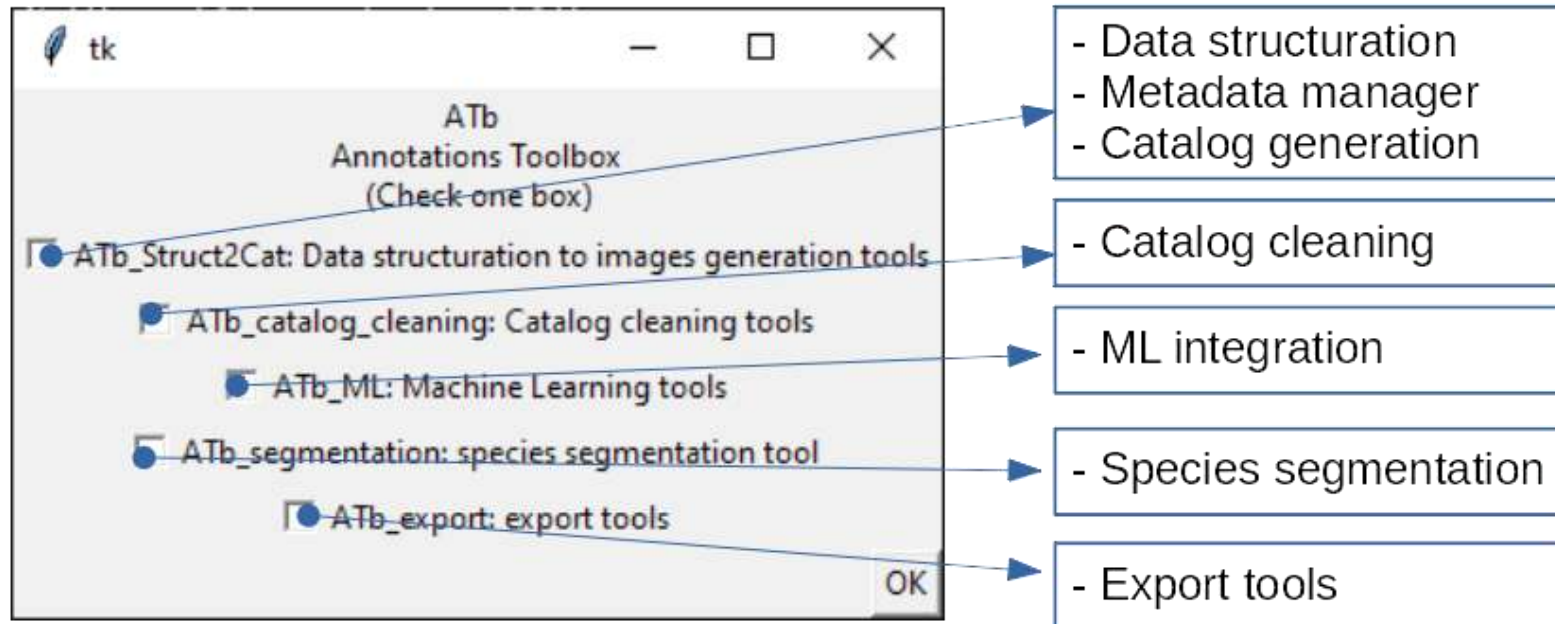
Export for big data based collaboration

Graphical interface

=> Annotations Toolbox : ATb

Annotations toolbox - ATb

Proposed solution:
Python graphical interface



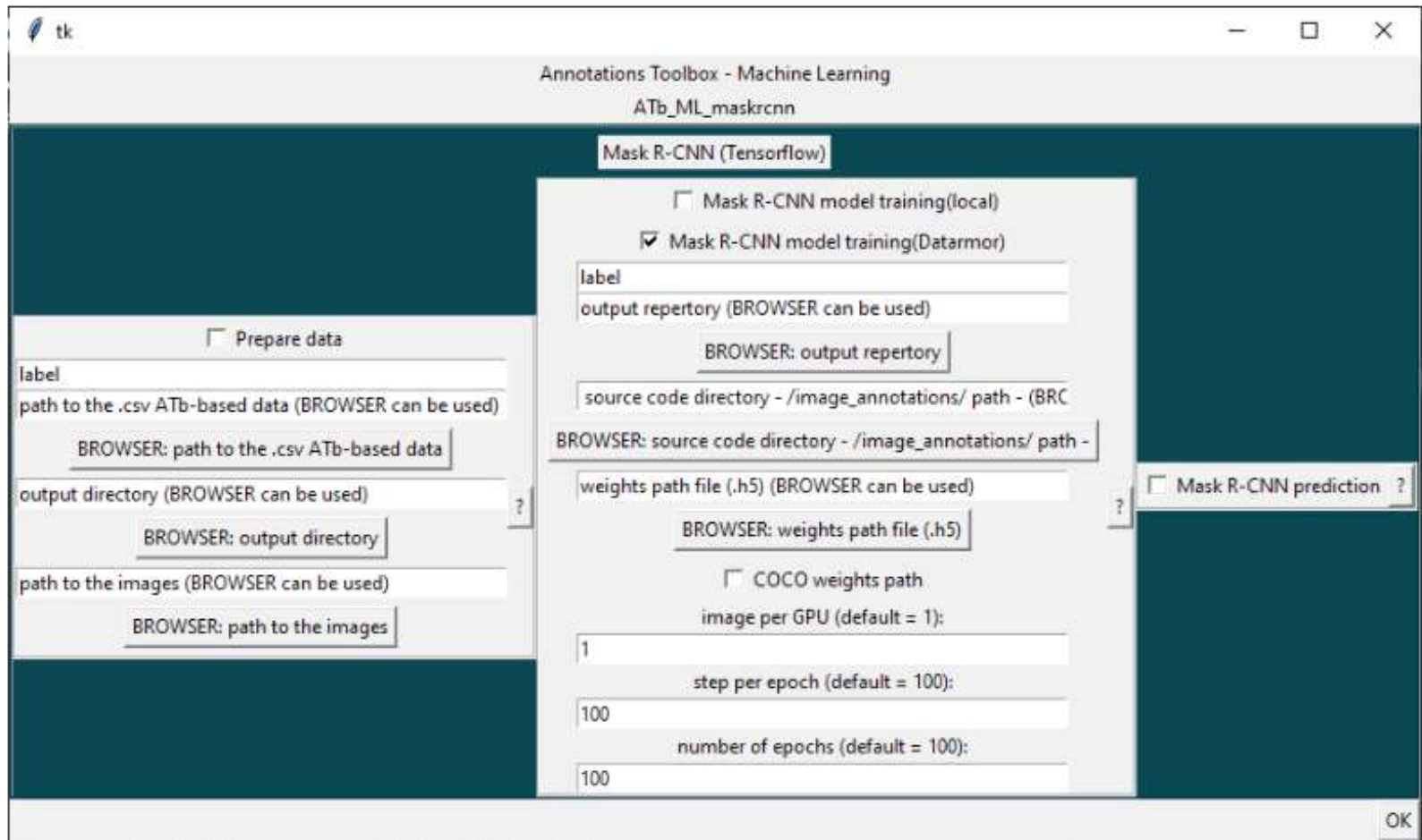
The screenshot shows a Tkinter window titled "tk" with a title bar containing standard window controls. The main content area is titled "ATb Annotations Toolbox (Check one box)". It contains a list of radio buttons, each with a label and a description:

- ATb_Struct2Cat: Data structuration to images generation tools
- ATb_catalog_cleaning: Catalog cleaning tools
- ATb_ML: Machine Learning tools
- ATb_segmentation: species segmentation tool
- ATb_export: export tools

An "OK" button is located at the bottom right of the window. Blue arrows point from each radio button to a corresponding box on the right side of the image, which lists the features associated with that tool:

- ATb_Struct2Cat: - Data structuration, - Metadata manager, - Catalog generation
- ATb_catalog_cleaning: - Catalog cleaning
- ATb_ML: - ML integration
- ATb_segmentation: - Species segmentation
- ATb_export: - Export tools

ATb via Datarmor → to train a model



The screenshot shows the 'Annotations Toolbox - Machine Learning' window with the 'ATb_ML_maskrcnn' configuration. The 'Mask R-CNN (Tensorflow)' section is active, and the 'Mask R-CNN model training(Datarmor)' option is selected. The configuration includes fields for 'label', 'output repertory', 'source code directory', 'weights path file', 'image per GPU', 'step per epoch', and 'number of epochs'. The 'Mask R-CNN prediction' option is also visible.

Annotations Toolbox - Machine Learning
ATb_ML_maskrcnn

Mask R-CNN (Tensorflow)

Mask R-CNN model training(local)
 Mask R-CNN model training(Datarmor)

label
output repertory (BROWSER can be used)
BROWSER: output repertory

source code directory - /image_annotations/ path - (BRC
BROWSER: source code directory - /image_annotations/ path -

weights path file (.h5) (BROWSER can be used) ?
BROWSER: weights path file (.h5) ?

Mask R-CNN prediction ?

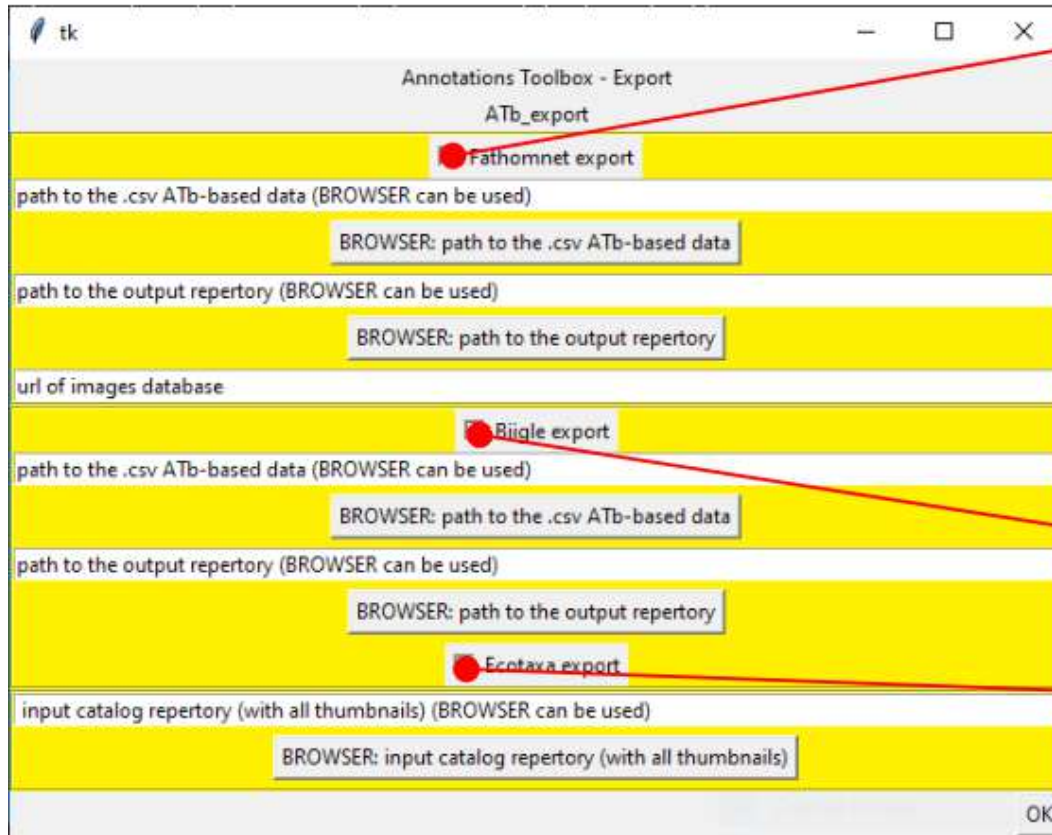
COCO weights path
image per GPU (default = 1):
1

step per epoch (default = 100):
100

number of epochs (default = 100):
100

OK

ATb_export



tk

Annotations Toolbox - Export
ATb_export

Fathomnet export

path to the .csv ATb-based data (BROWSER can be used)

BROWSER: path to the .csv ATb-based data

path to the output repertory (BROWSER can be used)

BROWSER: path to the output repertory

url of images database

Biigle export

path to the .csv ATb-based data (BROWSER can be used)

BROWSER: path to the .csv ATb-based data

path to the output repertory (BROWSER can be used)

BROWSER: path to the output repertory

Ecotaxa export

input catalog repertory (with all thumbnails) (BROWSER can be used)

BROWSER: input catalog repertory (with all thumbnails)

OK

Export to Fathomnet

Input :

- path to the .csv ATb-structured data
- path to the output repertory
- url of images database (work in progress)

Output :

- Fathomnet-structured .csv file

Export to Biigle

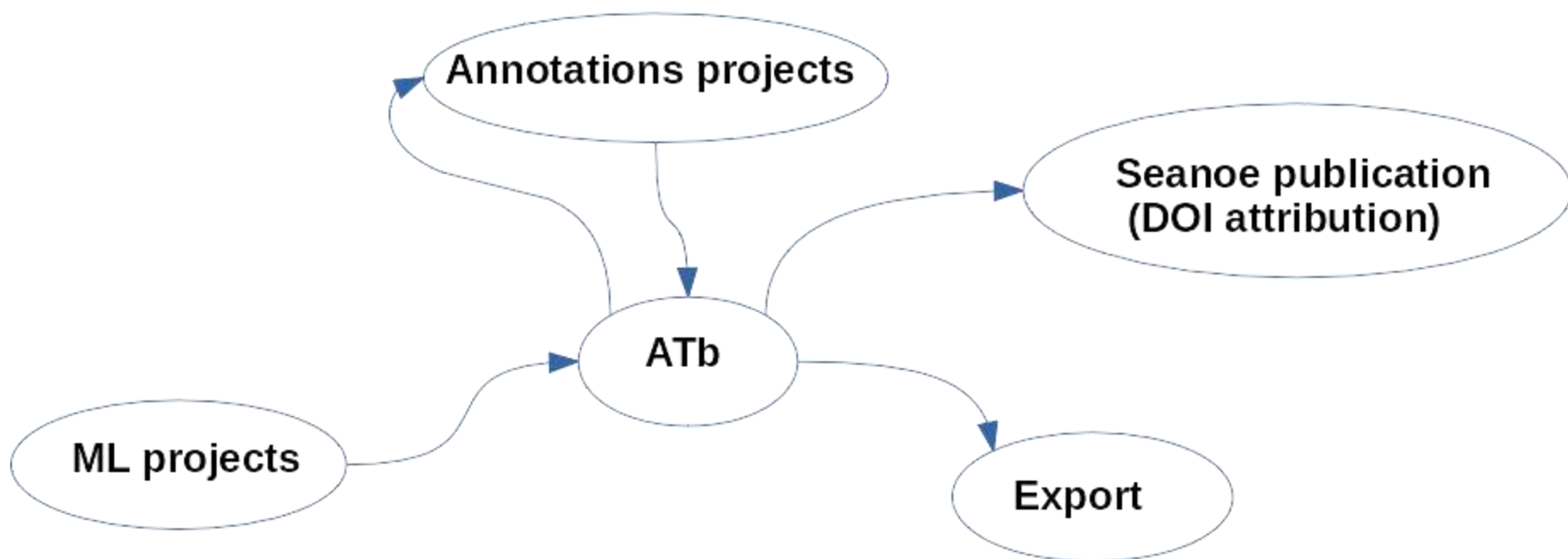
→ work in progress

Export to Ecotaxa

Input :

- input catalog repertory

Conclusion flowchart



Perspectives

- **iMagine Project**
- HEU [call](#) "**Integration of biodiversity monitoring data into the Digital Twin Ocean**"
to address the challenge of integrating image data (and information derived from it) into biodiversity assessment (including making image data FAIR and public, accessible to crowd annotation etc., including AI application, providing a link from annotations to OBIS, automated habitat mapping from annotation data, building a digital twin around annotations and habitat maps, etc.).
- **4th Marine Imaging Workshop**



Merci pour votre attention

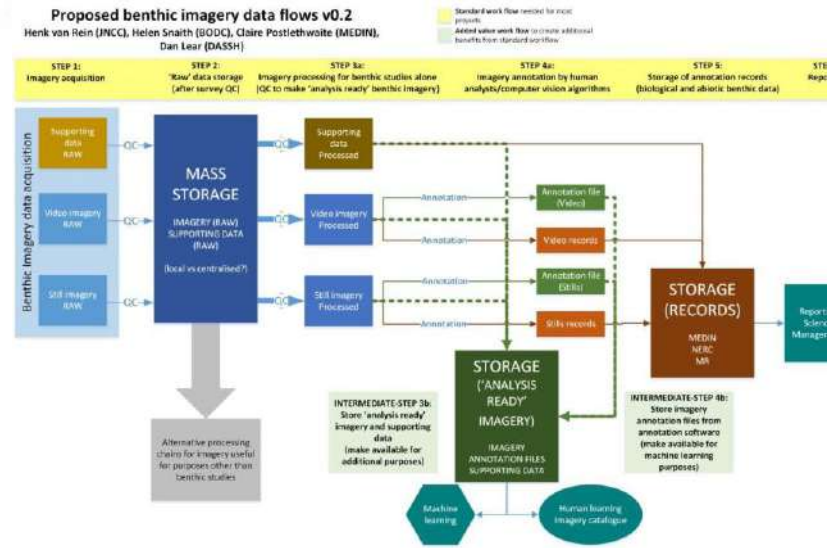
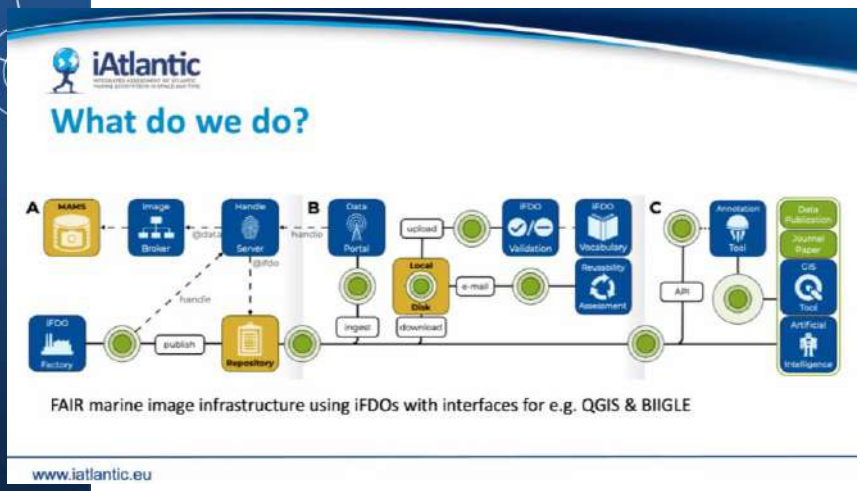


Standardisation Initiatives

FathomNet has many potential uses

Four key communities that FathomNet can benefit:

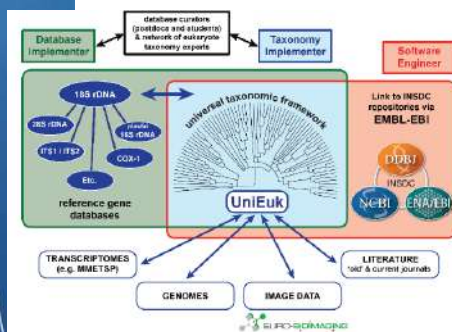
1. **Taxonomists:** Provide collaborative, global marine life key with flexible hierarchies that can aggregate observations of known and unknown biota.
2. **Programmers:** Deliver access to a novel data set that can be used to develop and evaluate state-of-the-art data science and computer vision algorithms.
3. **Enthusiasts:** Enable direct contributions to marine science and participate in the ocean life discovery process.
4. **Educators:** Provide resources and access to ocean visual data that enables experiential learning opportunities.



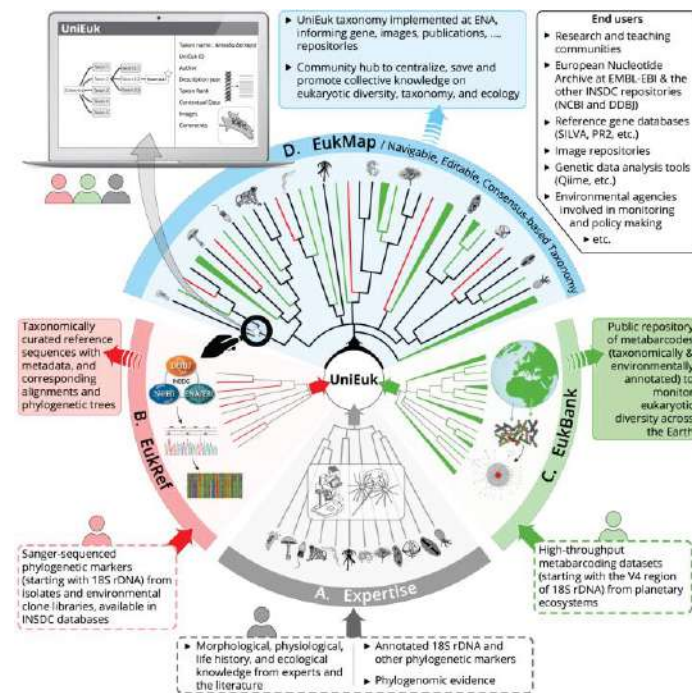
Standardisation - References

UniEuk

Universal taxonomic framework and integrated reference gene databases for Eukaryotic biology, ecology, and evolution



integrates information from relevant genetic markers and classical morphology-based data, and validated by a comprehensive network of taxonomy experts



... and interoperability

