SMHI

MAKING VIDEOS AND IMAGES FROM MARINE ENVIRONMENTAL MONITORING AVAILABLE TO ALL

A PICTURE IS WORTH A THOUSAND DATA POINTS

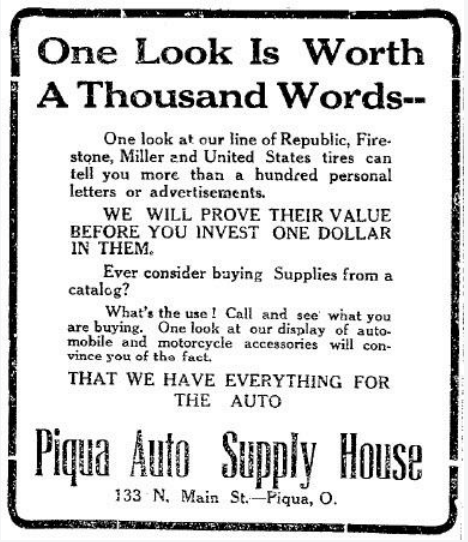
Markus Lindh

IMDIS 2021

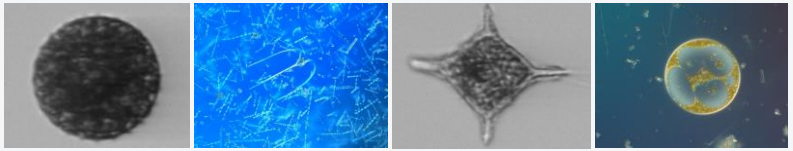
2021-03-12

Outline

- What is imagery data and why do we need to make it publicly available?
- Case study: Data from Imaging Flow Cytobot (IFCB)
- Potential issues and future development



Imagery data



- “Imagery data” can be referred to as qualitative and quantitative information from a collection of images.
- Imaging systems are used more and more frequently in the marine domain to generate huge amounts of imagery data.
- For example, automatic image classification is used to determine the abundance, size and biomass of plankton communities.

Why we need to make imagery data publicly available

- Valuable raw data! Need to connect between analyzed data and the raw imagery data. **Needed for reproducibility!**

- Stored locally, universites, regional governments, private sector. **National data coordination needed!**

- Strict demands from **INSPIRE directive**, demands also likely to be picked up from the **FAIR principles**.

The SMHI system

Deliverer

Video/image files delivered

Metadata* file delivered

QC Screening

Zip archiving

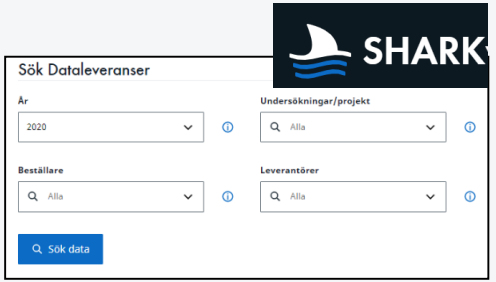
Upload to datacite

Publish in database
SHARKmedia

Generate DOI

External database

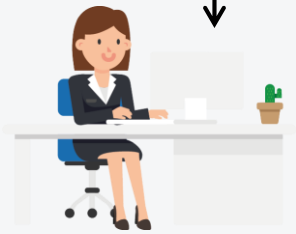
Couple DOI



Searchable through metadata fields, including DOI

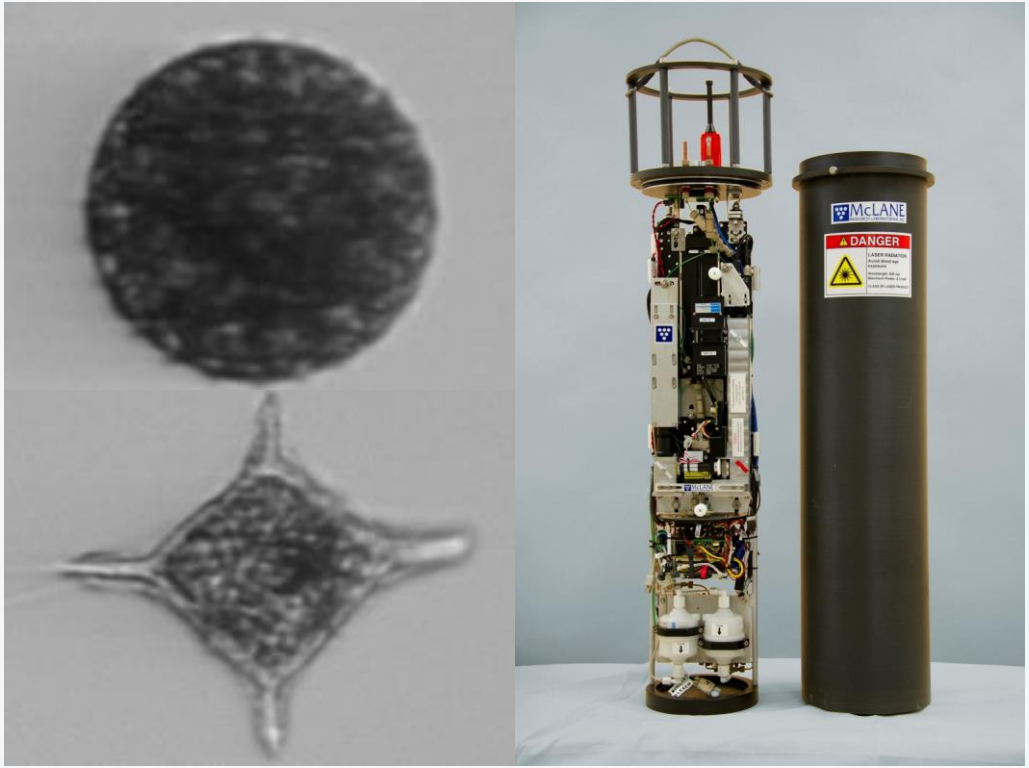
User directed to ftp

User



* We are using a metadata description applied from the INSPIRE directive and ICES/BODC vocabulary and include the use of digital object identifiers (DOI) for published data packages so that data is findable

Case study: IFCB data



1 sample = on average 20 MB data

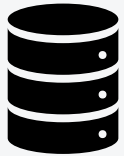
Number of samples per day (sampling every 20 minutes) = 24 x 3 = 72 samples

Number of days per expedition = 7

Number of expeditions = 12 per year

Total amount of data: 20 MB x 72 samples x 7 days x 12 expeditions = 120960 MB = **120 GB**

Potential issues

- Storage-related issues, both cost and IT infrastructure 
- Geographical information protected by law, special permit for publishing and secure infrastructure for archiving data required §
- GDPR issues §

Future development

- Ocean Best Practices within the JERICO-RI project and contribute to the Ocean Best Practices System Repository (OBPS-R). For more on the JERICO project see: https://biss.pensoft.net/article/58932/

- Workflow of plankton images through https://ecotaxa.obs-vlfr.fr/ Connection with Darwin Core formatting (DwC-A)

- API development instead of ftp

SMHI

Please get in touch with us:

sharkmedia@smhi.se

Thank you for your attention

Funding by:

Swedish Agency for Marine and Water Management