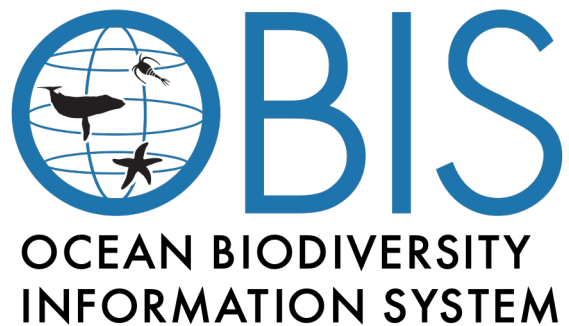




International
Oceanographic
Data and Information
Exchange

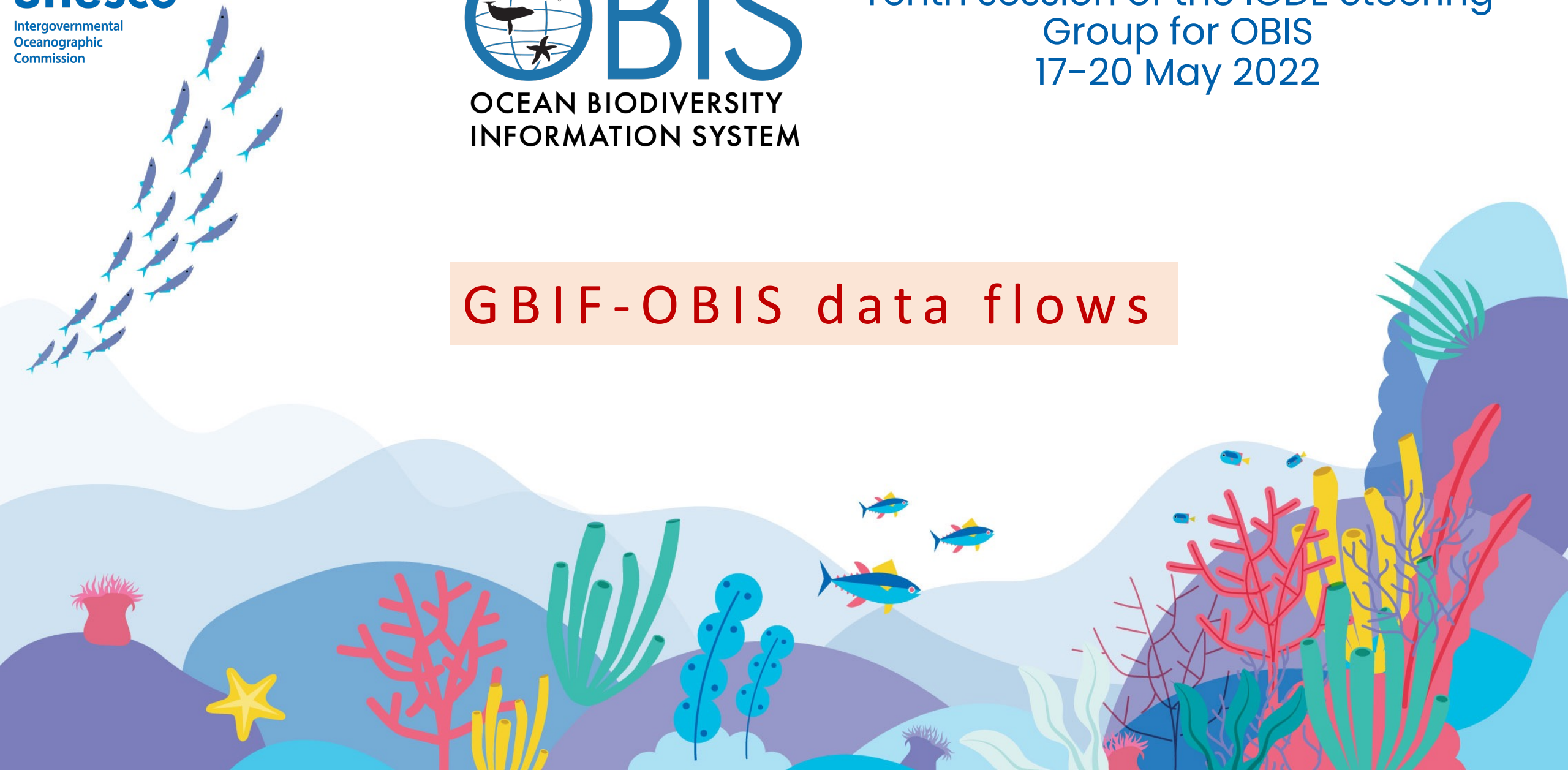
unesco

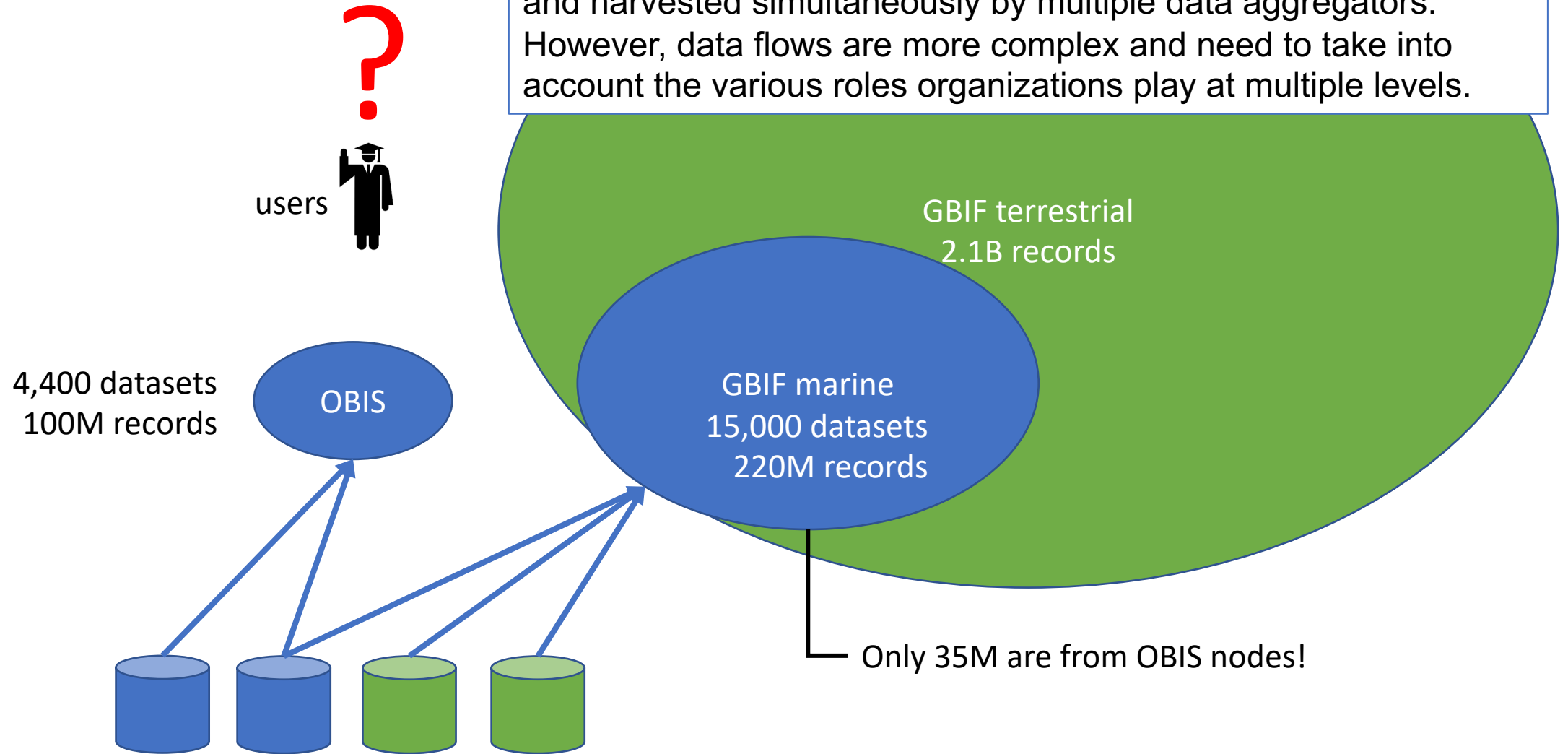
Intergovernmental
Oceanographic
Commission



Tenth session of the IODE Steering
Group for OBIS
17-20 May 2022

GBIF-OBIS data flows





OBIS nodes survey

25 OBIS nodes responded and there was general agreement that:

1. marine datasets published to GBIF should also flow to OBIS (note that most OBIS nodes are also publishing directly to GBIF, see overview here: <https://r.obis.org/gbif>),
2. that datasets should be harvested directly from the source IPTs, and
3. that we need to provide training to GBIF nodes/publishers in following OBIS best practices (e.g., WoRMS LSIDs and eMoF extension).

Three OBIS nodes said that these (marine) GBIF publishers should become OBIS nodes first and seven OBIS nodes said that marine datasets in GBIF should be endorsed by an OBIS node before ingesting these data into OBIS.

One of the issues was raised regarding potential duplication of data in OBIS and GBIF.
-> To mitigate this, the OBIS data manager developed a script (<https://iobis.github.io/notebook-duplicates/>), which should help the OBIS nodes to identify data duplication.

Upon request from several OBIS nodes to better represent the OBIS network (<https://www.gbif.org/network/2b7c7b4f-4d4f-40d3-94de-c28b6fa054a6>) on the GBIF portal , GBIF has released IPT version 2.5.2 (on 30 November 2021) which includes the functionality to link your datasets to networks like OBIS. These datasets are then automatically added to our network page on the GBIF website. Currently 2,074 datasets representing 38 million records are listed as belonging to the OBIS network. This is less than half of the data currently in OBIS.

The new functionality to indicate OBIS as the network in the IPT replaces the recommendation to add "marine, harvested by iOBIS" in the EML field "additional information". This was used by OBIS nodes who also manage non-marine datasets.

DRAFT

=> The SG-OBIS requests all OBIS nodes to upgrade their IPT to the latest version and indicate the OBIS network in all their datasets

Data flow from GBIF to OBIS

Based on the GBIF API, the OBIS secretariat has developed a script to list marine datasets that have the OBIS network tag, but are not yet in OBIS, as issues to this GitHub repository (<https://github.com/iobis/obis-network-datasets>)

DRAFT

The **SG-OBIS agrees** that the OBIS secretariat harvests marine datasets directly from the GBIF source IPTs following endorsement of the dataset by an OBIS node. The OBIS secretariat will then add the dataset to the endorsing OBIS node page. The OBIS node is in close contact with the GBIF publisher in case quality issues needs to be resolved. *The OBIS secretariat identifies which OBIS node would be best placed to endorse the dataset. If no response is received after one month, another node will be identified.*

In order to strive to have the same and highest quality of data in both GBIF and OBIS, the **SG-OBIS recommends** that the same and best version of each dataset (i.e. the master copy) should flow to GBIF and OBIS. This means that good communication is needed between OBIS nodes and GBIF publishers. For example, if a marine dataset is published directly to GBIF, but an OBIS node holds a better version (richer data, higher quality), then there are two options: (i) the GBIF publisher updates the dataset and publishes to GBIF and OBIS, or (ii) the OBIS node becomes the primary publisher and publishes the dataset to OBIS and GBIF. In the latter case, the GBIF publisher should not register the dataset to GBIF.