

DASSH, The Archive for Marine Species and Habitats Data – International Standards for National Data Management Excellence

Anna Luff, The Marine Biological Association (United Kingdom), annluf@mba.ac.uk

Dan Lear, The Marine Biological Association (United Kingdom), dble@mba.ac.uk

Kevin Paxman, The Marine Biological Association (United Kingdom), kevpx@mba.ac.uk

Matthew Arnold, The Marine Biological Association (United Kingdom), matarn@mba.ac.uk

The Archive for Marine Species and Habitats Data (DASSH) is the UK Data Archive Centre for marine biodiversity data, core funded by the UK Department for the Environment, Fisheries and Rural Affairs (DEFRA) and the Scottish Government and hosted by the Marine Biological Association of the United Kingdom. DASSH works in collaboration with, and is accredited by, the Marine and Environmental Data and Information Network (MEDIN), a network of UK organisations committed to sharing and improving access to marine data and information.

On an international scale DASSH is also the UK node of the Ocean Biogeographic Information System (OBIS), and an Associated Data Unit of the International Oceanographic Data and Information Exchange (IODE), giving the Data Archive Centre global recognition.

DASSH is a national and international hub for accessing, safeguarding, managing and disseminating historical, current and future datasets, DASSH also acts as a digital repository for biodiversity images and video.

DASSH provides online, open-source based tools for the description, discovery and download of biodiversity data, including the DASSH data portal, and Global Biodiversity Information Facility Integrated Publishing Toolkit (IPT). Through these tools, data can be downloaded in standard formats for integration into desktop and web GIS and statistical packages including R-based applications and PRIMER.

DASSH also operates a national data and metadata helpdesk and runs workshops to build data management capacity and capability in the UK marine sector.

Within DASSH and the other MEDIN Data Archive Centres, metadata is authored in accordance with the UK GEMINI discovery metadata standard, derived from the INSPIRE regulations, and the international ISO 19115 Geographic Information- metadata standards. The MEDIN community have worked to develop a marine implementation of the INSPIRE discovery metadata, using keywords, terms and definitions from the Vocabulary Service of the British Oceanographic Data Centre, as promoted through SeaDataNet.

As part of implementing the INSPIRE regulations, MEDIN publish data guidelines tailored to suit different methodologies and designed to implement the recording of standardised metadata alongside data, to increase the interoperability, and re-use of marine datasets.

DASSH are committed to applying the FAIR (Findable, Accessible, Interoperable and Reusable, Wilkinson et al, 2016) guiding principles in making marine biological data and associated metadata open and freely available worldwide, to inform decision making, research, education and other public-benefit purposes, whilst contributing to the natural resource of environmental data. Data are made accessible through a range of data systems and aggregators, including the DASSH online portal, the UK National Biodiversity Network Atlas, EMODnet Biology, EurOBIS and GBIF. Metadata are also made available through the MEDIN portal, containing marine datasets from over 400 UK organisations.

Recent developments within DASSH have focussed on the automation of the archiving and publication of standardised datasets. Using the open-source GeoServer application as the engine for delivery,

scripted workflows have been developed to address bottlenecks in data publication. Specifically we have developed an automated ingestion process combining datasets into a shared PostgreSQL database. The harmonisation of data systems drives a dynamic, web front-end for data discovery using geographic and taxonomic based searching.

DASSH have also developed a validation tool for MEDIN data which has vastly increased the accuracy of our QA procedure and decreased the human time required to check incoming datasets. Future developments include the creation of an online version of this tool, allowing users to validate their data before submission.

We have also been able to make progress in conversion between standards, for example from the Marine Recorder marine survey database to MEDIN Data Guideline and from MEDIN to Darwin Core. The latter is a vital component of our role as an OBIS node, allowing the publishing of datasets via IPT.

The presentation will focus on how the application of automated Quality Assurance and ingestion tools can streamline data flow at the national and global scale, and how the UK marine biological data infrastructure is contributing to the development of a global marine biodiversity observation network.