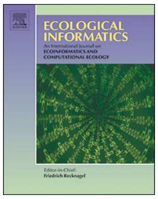
Ecological Informatics 59 (2020) 101115

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



An open-source database model and collections management system for fish scale and otolith archives

Elizabeth Tray[a,*], Adam Leadbetter[b], Will Meaney[b], Andrew Conway[b], Caoimhín Kelly[b], Niall Ó Maoiléidigh[c,**], Elvira de Eyto[c], Siobhan Moran[b], Deirdre Brophy[a]

[a] Marine and Freshwater Research Centre, Galway-Mayo Institute of Technology, Co. Galway, Ireland
[b] Marine Institute, Rinville, Oranmore, Co. Galway, Ireland
[c] Marine Institute, Furnace, Newport, Co. Mayo, Ireland

ARTICLE INFO

Keywords:
Database
Fish scale
Otolith
Archive
FAIR data principles

ABSTRACT

Scales and otoliths (ear stones) from fish are routinely sampled for age estimation and fisheries management purposes. Growth records from scales and otoliths can be used to generate long-term time series data, and in combination with environmental data, can reveal species specific population responses to a changing climate. Additionally, scale and otolith microchemical data can be utilized to investigate fish habitat usage. A common problem associated with biological collections, is that while sample intake grows, long-term physical storage is rarely a priority, and much of the sampling took place before the advent of open-access digital infrastructure. Material is often collected to meet short-term objectives and resources are seldom committed to maintaining and archiving collections. As a consequence, precious samples are frequently stored in many different and unsuitable locations, and may become lost or separated from associated metadata. The Marine Institute's ecological research station in in Newport, Co. Mayo, Ireland, holds a multi-decadal (1928–2020) collection of scales and otoliths from various fish species, gathered from many geographic locations. Here we present an open-source database model and archiving system to consolidate and digitize this collection, and show how this case study infrastructure could be used for other biological sample collections. The system utilizes the FAIR (Findable Accessible Interoperable and Reusable) open-data principles, and includes a physical repository, sample metadata catalogue, and image library.

1. Introduction

Calcified biomineral structures from fish, such as otoliths (ear stones), scales, opercula (gill covers) and vertebrae contain growth marks which form incrementally and be used to estimate fish age (Panfili et al., 2002). For example, in otoliths, which assist the fish in hearing and maintaining balance, annual bands are formed due to a slowing of growth rates in winter (Campana and Thorrold, 2001; Popper et al., 2005). Furthermore, growth marks can be deposited at a rate that is proportional to body growth, and can be used to reconstruct individual growth histories (Farley et al., 2007; Friedland et al., 2000; Pannella, 1971). Estimates of population growth rates and age structure are important parameters in the stock assessment of many commercially exploited species (Begg et al., 1999; Mapp et al., 2017; Marine Institute, 2019; Wells et al., 2013). Therefore, calcified structures have been routinely collected by fisheries institutes worldwide for the purpose of age estimation for the last 100 years (Rivers and Ardren, 1998). Due to their largely inorganic composition, these structures can be dried and stored easily for extended periods of time, and subsequently enabled many institutes to have amassed large multi-decadal collections.

Biomineral samples from fish can be used to answer a diverse range of ecological research questions using both established and advanced analytical techniques. Measurements of growth marks in calcified structures provide a proxy for fish growth, and can be used to assemble extended biochronology time series and to examine relationships with climatic, population and fishing related drivers (Matta et al., 2010; Morrongiello et al., 2019; Peyronnet et al., 2007; Smoliński and Mirny, 2017). Many studies have used trace elements and isotopes in the inorganic portion of calcified structures to infer fish origin (Adey et al., 2009; Flem et al., 2017; Zeigler and Whitledge, 2011) and habitat usage (Brennan et al., 2019; Fraile et al., 2016) or to detect exposure to

---

* Corresponding author.
** Corresponding author.
E-mail addresses: elizabeth.tray@gmit.ie (E. Tray), niall.omaoileidigh@marine.ie (N.Ó. Maoiléidigh).

https://doi.org/10.1016/j.ecoinf.2020.101115
Received 13 February 2020; Received in revised form 2 April 2020
Available online 08 June 2020

pollution (Limburg et al., 2015; Limburg and Casini, 2019; López-Duarte et al., 2016). The extraction of DNA from scales and otoliths facilitates the study of temporal change in genetic diversity (Hansen et al., 2002; Hutchinson et al., 2003), while measurements of stable isotopes in the organic component of scales (MacKenzie et al., 2012; Nonogaki et al., 2007; Trueman et al., 2012) and more recently otoliths (Grønkjær et al., 2013; Sirot et al., 2017) support investigations of diet, metabolism and trophic structure. As sclerochronology techniques develop so too does the potential for scale and otolith archives (or other biomineral collections) to provide insight into marine ecosystem processes (Hunter et al., 2018; Morrongiello et al., 2012).

The maintenance of scale and otolith collections can change depending on management, ongoing monitoring priorities, and project-specific research needs (Copp et al., 2014). The material is usually collected to serve short-term management objectives and its long-term storage is often not maintained according to clearly defined standard operating procedures (Rivers and Ardren, 1998). As a consequence, procedures change with employee turnover and over time a sample's provenance may become ambiguous. For example, short-hand used for species names may be uninterpretable after two decades, leaving the sample unusable. Continuous sampling programs and opportunistic sampling of fish scales and otoliths can create invaluable, albeit irregular, archive collections. Essential metadata is often recorded in a field notebook or a digital spreadsheet, which is rarely stored with the associated physical samples, and if it is, it can become separated after time. Furthermore, if samples are not appropriately stored, pests and moisture can degrade their integrity after long periods of time (Vollmar et al., 2010). Propriety software is currently available for implementing collection management (CM) systems, but it can be expensive and dependent on ongoing investment (Peluso et al., 2016; Short and Anderl, 2012) and limited in adaptability and scalability. As a result, CM systems can become inaccessible after long periods of time. A viable alternative is to develop tailored open-access inventory software (Arriaga et al., 2015; Gries et al., 2016).

Global legislative efforts to ensure that research data remains accessible have contributed to significant improvements in data quality, interoperability, and visibility (Wieczorek et al., 2012; Clarke and Margetts, 2014). At the European regional level, the INSPIRE Directive is a legislative driver for governments to construct standardized spatial data infrastructure, specifically for environmental data (Craglia and Annoni, 2007). The INSPIRE regulations aim to enable spatial data sharing between public sectors, improve public access, and support cross-boundary policy formation. To comply with INSPIRE regulations, data infrastructures must contain the flexibility to align within the widely used Observations and Measurements (O&M) model, (Cox, 2016). Furthermore, other standards structures have gained widespread community usage, such as biodiversity data standards, like Darwin Core (Wieczorek et al., 2012). Recently, the FAIR (Findable Accessible Interoperable and Reusable) open-data principles (Wilkinson et al., 2016), were created through a bottom-up, community based initiative. A component of the FAIR principals requires the usage of existing standards (e.g. INSPIRE, O&M, Darwin Core). Consequently, scientists have implemented FAIR data repositories specifically intended for sample and data management (Adam et al., 2019; Conze et al., 2017; Dassié et al., 2017; Lehnert et al., 2006; McNutt et al., 2016). However, a FAIR compliant system and model for local repositories, purposefully designed for physical ecological sample archives and their derived data, has yet to materialize.

The Marine Institute (MI) in Ireland, recently constructed an INSPIRE compliant Data Catalogue for its environmental data (Leadbetter et al., 2019; Leadbetter et al., 2020). The CM system presented in this paper is an extended feature of the Data Catalogue described by Leadbetter et al., 2020. The CM system was explicitly designed to house a fish scale and otolith collection, which consists of thousands of samples, dating back to 1928, from a range of geographic locations. The system was developed with two primary objectives; to 1) produce a scalable and interoperable CM archive database and physical repository for scale and otolith samples, and 2) document a work-flow to archive past, present, and future samples, and their data progeny. The aim of this study is to lay a road map, and provide a toolbox, for other agencies who wish to create an 'open' digital and physical infrastructure for their own biomineral archives. While this system was specifically designed for fish scale and otolith archives, and their resulting images and data, it could certainly be adapted for biological sample archives from a variety of species, locations, or time periods.

2. MI research station & history

A field research station has been operating in the Burrishoole catchment in Newport, Ireland since 1955 with a primary focus on the monitoring of native fish species including Atlantic salmon (*Salmo salar* L.), brown trout (*Salmo trutta* L.) and European eel (*Anguilla anguilla* L.). The diadromous migration of these populations has been monitored for many decades, as they move between marine and freshwater habitats, through permanent fish traps constructed at the base of the Burrishoole catchment. The Burrishoole catchment is an internationally important index system for the long term monitoring and stock assessment of these species, and biological samples have been collected and stored at the Burrishoole research station throughout its existence. The ownership and management of the research station has changed over the decades, from the privately owned Salmon Research Trust (SRT), to the publicly owned Salmon Research Agency of Ireland (SRAI). Since 1999, it has been owned and managed by the MI, an Irish government state agency, primarily responsible for providing aquatic scientific and technical advice to the government. Commensurate with its complicated history, the biological samples housed at the research station derive from many sources and locations. There is a multi-decadal collection of thousands of salmonid and trout scales, and eel otoliths from Burrishoole populations. In addition, the collection comprises samples from pollack (*Pollachius pollachius* L.), Arctic char (*Salvelinus alpinus* L.), and Bluefin tuna (*Thunnus thynnus* L.), and other species, collected during various research projects. During the amalgamation of the SRAI and the FRC (Fisheries Research Centre) into the MI, many biological samples were physically moved from locations around Ireland to the Burrishoole research station. Therefore, the resulting collection of biological samples is a mix of species, sample types and populations, all of which needed to be catalogued and stored correctly in one CM system.

3. System design – physical sample repository

The fish scales and otoliths in the MI collection were often stored in acid-free envelopes, in bundles of 50–100 samples. These bundles typically had a common metadata grouping factor, for example, they were the same species and collected in the same season. Scales and otoliths had also been mounted and stored on glass microscope slides, which were stored vertically in either slide cabinets or lying flat in slide trays within black archival cardboard boxes. The slides were also typically grouped according to a common metadata factor. The CM system was designed to accommodate these original sample metadata groupings. Once samples had fulfilled their purpose, they were put into arbitrary storage by the researcher who had collected them, Appendix A.

The envelopes were re-housed in clear plastic storage containers, fitted with desiccant pouches to prevent mould outbreaks. The sample repository used nested storage; 21 smaller containers fit into one large, mobile 'unit'. Each small container was allocated a bundle of envelopes. The glass slide samples were re-housed in the new archive-quality cardboard boxes. A globally unique repository identifier was assigned to each sample (envelope or glass slide) using archival quality labels and a thermal transfer label printer.

4. System design – database

4.1. The MI data catalogue & IFBA database

The MI's metadata repository, known as the Data Catalogue, is an online infrastructure which supports the storage of datasets generated by systems, programs, and individual researchers within the Institute (Leadbetter et al., 2019). The Data Catalogue is operated through Drupal, an open-source, web-based content management system. Drupal enables database specialisation by permitting developers to build custom 'modules', which are a set of instructions written for a specific function. The database developed, named IFBA (Irish Fish Biochronology Archive) is a self-contained module, that operates as an extended feature of the MI Data Catalogue module. This system was specifically designed for scales and otoliths, because they are, globally, the most commonly collected and used biomineral structures from fish used for aging purposes (Panfili et al., 2002). The system is flexible enough to be adapted for any biological sample type (e.g. shells, corals, vertebra, etc.). The documentation for the IFBA database is publicly available in the GitHub repository, https://github.com/IrishMarineInstitute/IFBA. The IFBA dataset (Ó Ó Maoiléidigh et al., 2020) is described in the Marine Institute's Data Catalogue (Leadbetter et al., 2020).

The IFBA database is access controlled, to ensure security and data quality, so users are allocated a login and password to upload sample metadata. First a digital 'sample set container' is created for a bundle of envelopes (Appendix B), which corresponds to a physical small storage container. After the digital and physical sample set container is generated there are three options for uploading samples within the sample set container. The first option is to manually upload the fish sample metadata (e.g. length, weight) for each sample into the sample set container table within the IFBA database. The second option, if time is limited, is to upload one sample from the container, named a 'Representative Sample' (RS), as a tag to indicate what is in the sample set container. For example, a container may have one RS to indicate it contains 100 salmon scale samples, from 1975, from a specific location. The third option is for an elevated privilege user to upload the sample metadata through a standard, pre-populated CSV template. The unique archive id links each physical sample to its entry within the IFBA database (Appendix B).

4.2. Data model & vocabularies

A conceptual diagram of the many types of data that can be generated from one fish is shown in Fig. 1. There are 4 parameters of critical information for digital storage: fish sampling location, date, species, and life stage. The CM database design is centred on each physical sample, Fig. 2. The metadata, such as fish length, weight, and day collected, is typically written on the sample envelopes. Some envelopes and glass slides contained no metadata, or were labelled with only a lab number and species. Furthermore, older labels were often either illegible or damaged. Therefore, samples that contained the necessary critical information were prioritized to be catalogued first in the database and repository. The internal vocabulary (and shorthand) used during sample labelling within the MI field station changed through time, and between researchers. Furthermore, the local vocabularies used were often not in accordance with best practice (Hedden, 2010). The usage of controlled vocabularies, which is a pre-defined list of terms, enables consistent and accurate information labelling (Mao et al., 2005). The field station's local vocabulary was mapped to internationally recognized controlled vocabularies, with strong content governance models. Controlled vocabulary terms from FishBase (Froese and Pauly, 2010), the World Register of Marine Species (WoRMS Editorial Board, 2020), and the International Council for Exploration of the Seas (ICES) were used to ensure interoperability and transparency (Table 1). The model allows the physical repository to be adapted to accommodate various laboratory storage systems.

4.3. Geo-referencing sampling location

For contemporary samples, geographic coordinates of the sampling location are often recorded. For older samples the sampling location can be ambiguous and may only refer to a lake, estuary, catchment or region where the fish was captured. Depending on the planned analyses for the archive samples, precise sampling location may not be required, and a broad location on classification can still be useful. Water Framework Directive (WFD) management areas provided the geographic reference boundaries for sampling locations (EPA, 2015) for the IFBA CM system. Each of these geographically referenced boundaries are classified as a 'Geographic Feature' within the MI Data Catalogue, as defined in Leadbetter et al., 2020. A Geographic Feature defines the feature type, and stores the dataset's (in this case the sample's) spatial information as points, lines, or polygons, in alignment with the European Commission's INSPIRE spatial data infrastructure. A custom python script was used to create WKT (Well-known text) files for the WFD catchment boundaries (Herring, 2011), and can be found in this repository; https://github.com/IrishMarineInstitute/IFBA. Within IFBA, scale and otolith samples with a known sampling point are allocated coordinates, and coordinate data are nested within the appropriate catchment boundary, for high-level search capabilities (e.g. searching

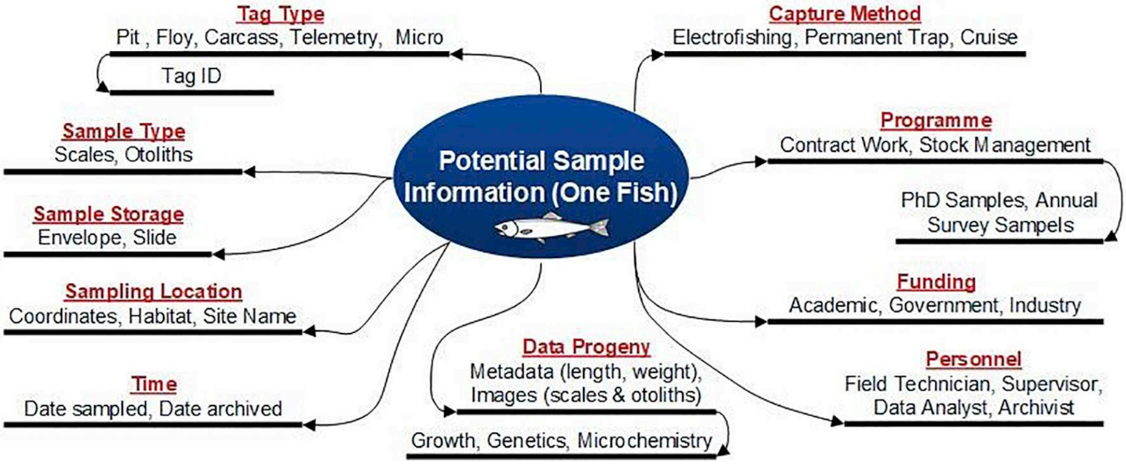


Fig. 1. A mind map of the potential information associated with a fish scale or otolith sample.

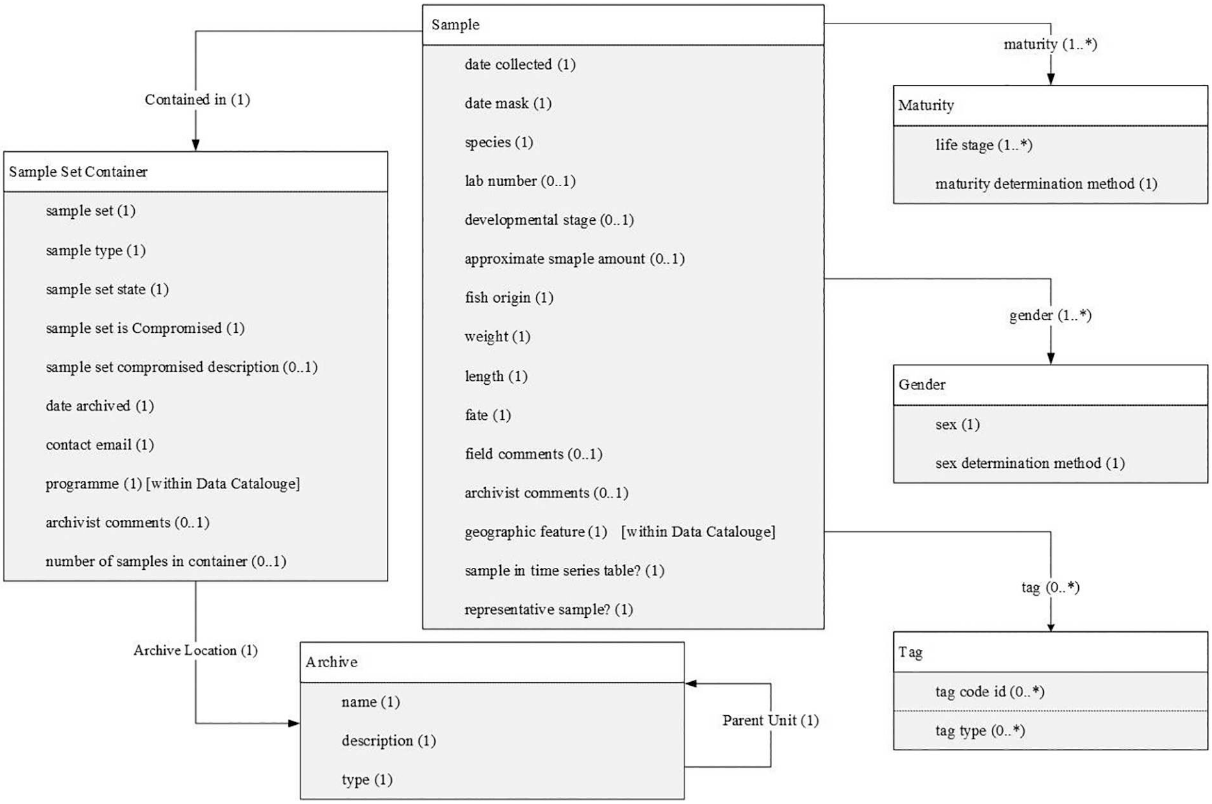


Fig. 2. Data model for the IFBA fish scale and otolith archive.

Table 1
Some of the external vocabulary servers used for various parameters in the data model.

Vocabulary server	CM Parameter	Description	Link to parameter	Terms used
International Council for the Exploration of the Seas (ICES)	Fish Origin	The genetic origin of the sampled specimen	https://vocab.ices.dk/?ref=153	Hatchery Reared Local Wild Stock
	Developmental Stage	The life stage of sampled specimen	https://vocab.ices.dk/?ref=52	Adult Juvenile
	Gender	The gender of the sampled specimen	https://vocab.ices.dk/?ref=1478	Male Female
World Register of Marine Species (WoRMS)	Species	The species of the sampled specimen	https://www.marinespecies.org/	*Salmo salar* *Salmo trutta* *Anguilla Anguilla* *Perca fluviatilis* *Rutilus rutilus* *Salvelinus alpinus*
FishBase	Maturity	The observed maturity of the sampled specimen	https://www.fishbase.se/search.php	Multi-sea-winter Kelt Slob Grilse Finnock Parr Smolt Elver Springfish

samples by WFD catchment). If the sampling point is unknown, but the WFD catchment is known, samples are allocated to the catchment boundary.

4.4. Image library

Scale and otolith images are used for age and growth estimations. In IFBA, images are stored in a collaborative online space. The image file name for any image is the same as the unique archive sample id in the IFBA database. An example of the data progeny resulting from a fish scale sample is illustrated in Appendix B. During imaging, the sample id is permanently marked onto the image to prevent file name corruption issues. Furthermore, there is Boolean phrase within the IFBA database to indicate if the sample catalogued is held within the time-series image library.

4.5. Process flow

Formal documentation of process flows is often used in industry to document logistical pathways and reveal areas where efficiency could be improved upon. In line with the Marine Institute's Data Management Quality Management Framework (Leadbetter et al., 2019), to document

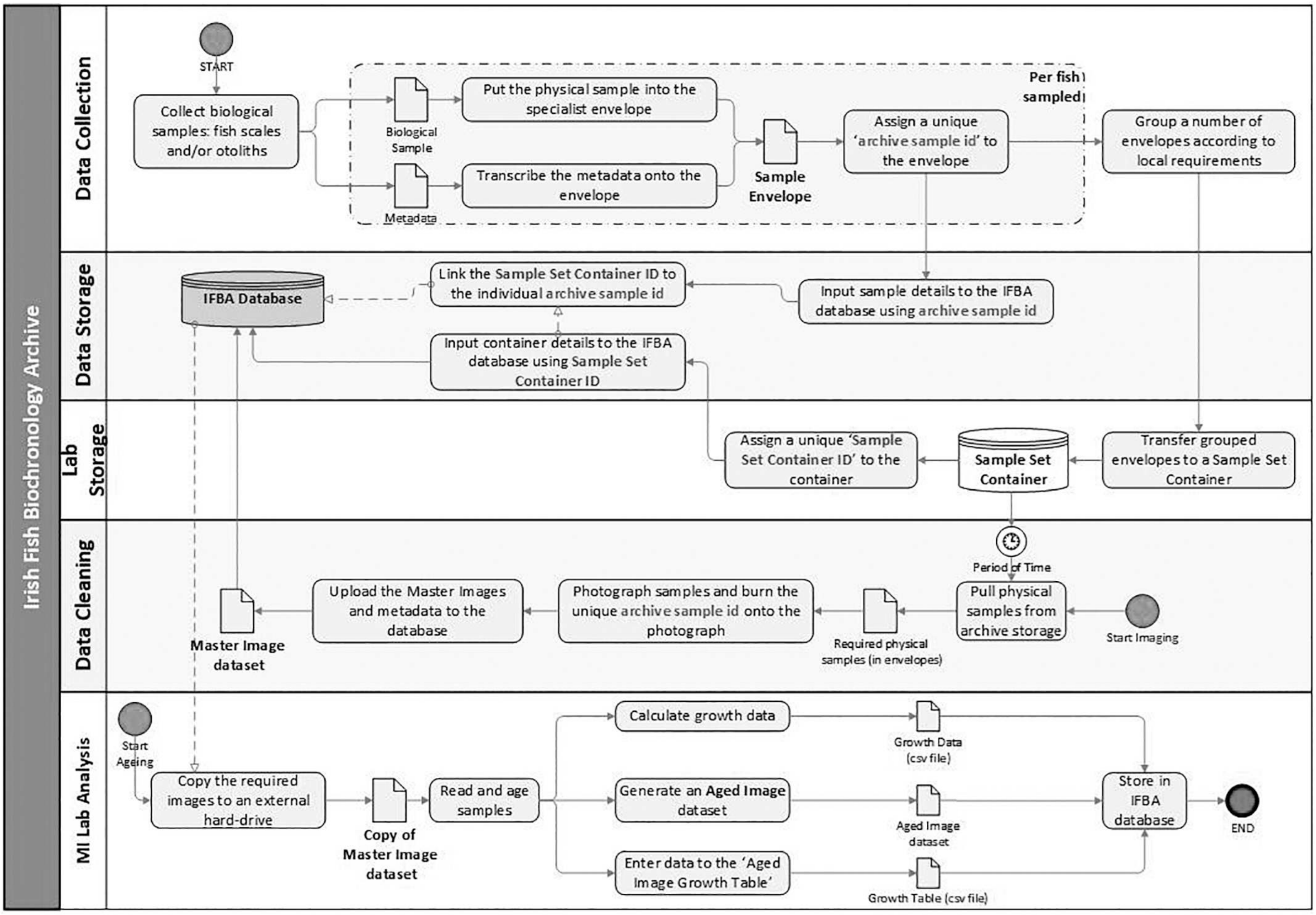


Fig. 3. A process flow for the scale and otolith CM system.

the procedures involved in storing, uploading metadata, and imaging samples, a high-level process flow was developed (Fig. 3). The process flow outlines where the sample metadata is generated, analysed, and eventually stored. Furthermore, it documents how the physical samples and digital data are linked. This cross-disciplinary process flow is novel because it outlines how the operational part (routine sampling of fish for government mandated stock assessment) of an organization (the MI), and the research capacity of an organization (imaging and reading growth on scales and otoliths), can be interlinked. Additionally, the management of the IFBA dataset follows the procedures and roles outlined within the MI's Data Management Quality Framework (Leadbetter et al., 2019).

4.6. FAIR data principles

The criteria of how the IFBA database addresses the requirements of the FAIR (Findable, Accessible Interoperable, and Reusable) data management principles (Wilkinson et al., 2016) can be found in Table 2. The FAIR principals were created by a group of stakeholders from academia, industry and funding agencies to enhance data reusability and findability by machines. In utilizing the FAIR principles, researchers can allow for interoperability between their datasets.

4.7. Observations and measurements mapping

A representation of the IFBA database in the Observations and Measurements (O&M) model has been created to comply with both the ISO 19156 standard and the INSPIRE regulations, Fig. 4. Within O&M model, an Observation is action whose result is an estimate of the value of some property of the feature of interest obtained using a specified procedure (Cox, 2016). For the O&M representation of IFBA, the sampling portion of O&M has been used, and as such Observations have been made against a specimen. The specimen, in this case a fish, was removed from the environment and analysed ex-situ. Multiple Observations are made against an individual specimen, each leading to a subsequent result. Following Leadbetter and Vodden (2015), the property being observed may be complex and have multiple attributes. Within IFBA, for example, the length of the fish specimen has units of measure of metres. The specimen is linked to a geographic feature of interest from which it was collected. The geographic feature to which the specimen is linked is defined within the MI Data Catalogue. This O&M representation demonstrates the extensibility of the IFBA database to allow for other biological samples. This approach can also be extended to representing the content of the IFBA database to other data serialisation models, such as Darwin Core (Wieczorek et al., 2012; see listing in Appendix C). Furthermore, the general concept can be applied to other sample types.

5. Discussion

Properly maintaining and curating archives of fish calcified structures and their metadata is essential to realize their full scientific value. By developing and implementing an open-source and on-site CM system for scale and otolith archives, researchers can maintain the physical and digital integrity of biological samples at a notably lower cost than if using a proprietary software or a large-scale open data repository. Furthermore, an open-source CM archive system is more likely to endure changes in budgets and employee turnover. Another benefit of

Table 2

How the IFBA CM system meets the FAIR principles of Data Management (Wilkinson et al., 2016) (Table adapted from Leadbetter et al., 2020).

FAIR term	FAIR principle	How the IFBA system meets the FAIR principle
Findable	F1. (Meta)data are assigned a globally unique and persistent identifier	Samples are assigned a persistent unique identifier physically, and then catalogued into the digital system using the identifier.
	F2. Data are described with rich metadata	The Data Model in Fig. 2 and the O&M Model in Fig. 4 outlines the rich metadata model for sample data description.
	F3. Metadata clearly and explicitly include the identifier of the data they describe	The unique identifier is clearly outlined on the physical sample, and links it to the digital system (Fig. 2). Furthermore, resulting images of the sample are also permanently linked to the unique identifier.
	F4. (Meta)data are registered or indexed in a searchable resource	Metadata are indexed and searchable locally, and further work is being doing to make it more widely searchable
Accessible	A1. (Meta)data are retrievable by their identifier using a standardized communications protocol	The metadata are stored in a Web-based system which uses a standard HTTP to retrieve metadata from a Web address based on the sample's unique identifier.
	A2. Metadata are accessible, even when the data are no longer available	Even if samples have been destroyed for analyses, their metadata is still recorded and preserved. The dataset to which the destroyed samples belonged can be accessed from within the MI Data Catalogue system.
Interoperable	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	The data have been mapped to the Observations and Measurements model, as described in Section 4.7, and therefore comply with ISO, Open Geospatial Consortium, and INSPIRE Spatial Data Infrastructure standards.
	I2. (Meta)data use vocabularies that follow FAIR principles	The controlled vocabularies the IFBA system utilized can be found in Table 1.
	I3. (Meta)data include qualified references to other (meta)data	Data stored within IFBA references dataset metadata and geospatial feature metadata stored within the MI Data Catalogue (Leadbetter et al., 2020).
Reusable	R1. Meta(data) are richly described with a plurality of accurate and relevant attributes	Figs. 3 and 4 show the breadth of attributes recorded in the metadata. The mapping to O&M demonstrating that more than a minimum set of attributes is described.

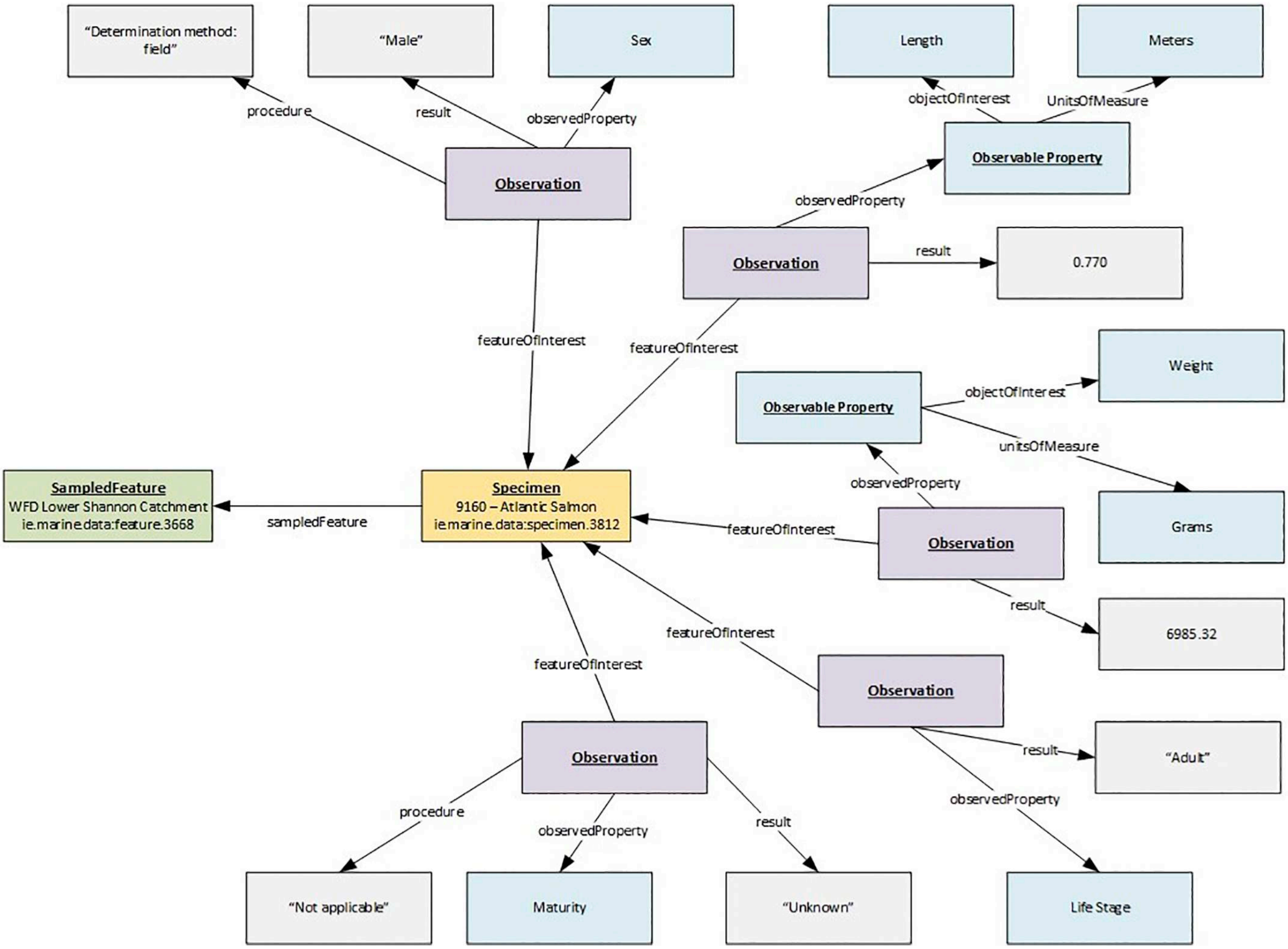


Fig. 4. An Observations and Measurements model for the IFBA database.

developing software as open-source, is that once created, modules can be readily shared, reused and extended by others.

To the best of our knowledge, IFBA is the first CM system to follow INSPIRE regulations and utilizes the FAIR open data principals. This feature is particularly important for biomineral archives which may be collected for one purpose but which could potentially support a much wider range of analyses (e.g. age/growth, genetics, stable isotopes, or trace elements). IFBA provides the infrastructure and system to properly store and catalogue material collected through publicly-funded programmes. The system links the physical sample with all associated data within one repository and digital catalogue, thus relinquishing individual ownership of or responsibility for samples and data and preserves them for posterity. Furthermore, the availability of geographically referenced samples from past years can facilitate research into unforeseen events. For example, the impacts on fish growth of a weather anomaly, such as a period of abnormally cold water temperatures, could be investigated by sourcing scale or otolith samples from the relevant time period. Since the samples have already been collected, they could be easily located within IFBA, with minimal resource requirements. Studies like these can only exist once previous biomineral samples have been physically archived and digitally catalogued, and made available to all researchers. The CM system presented here is fully scalable and can be used to store samples data from a range of sources. Data management plans (e.g. as required for Horizon 2020 projects) for short term projects as well as long term archival programmes could feature a national archive, such as the IFBA system created in this study, for sample deposition post-project, increasing the accessibility of the material and allowing for their re-use and reducing the proliferation of data archival centres.

5.1. Limitations & lessons learned

The material within IFBA was derived from numerous sources, and its collation involved many different agencies and researchers. Extensive research and effort was made to decipher shorthand used by researchers for various metrics (for example, fish length used to be recorded using the imperial system, and today it is recorded using the metric system). Accurate archiving of the material relied on communication with staff who had worked alongside the archive for long periods. In person communication was critical to decode the shorthand within the archive collection. The ability to decipher shorthand disappears with time, as researchers change jobs or retire, leaving invaluable collections meaningless (Rivers and Ardren, 1998; Fegraus et al., 2005). Appropriately archiving these collections before key personnel retire is paramount. This case study took over two years to develop and implement, one full-time staff member, numerous student work-placements, and an equipment budget of approximately 10,000 euros. Furthermore, the authors would like to note this project required extensive and persistent internal collaboration between fisheries scientists and data managers within the MI. For example, it was necessary for the fisheries scientists to explain the complexity of the data produced when sampling diadromous fish populations through time. In addition, it was equally important for the data managers to educate the fisheries scientists on informatics science, specifically on how to structure, model, and store data, in accordance with best practice. The data managers led and continue to have the role of maintaining the digital infrastructure for the archive, while the fisheries scientists played a key role in IFBA's design, and continue to collect samples and populate the database. The effort to catalogue all samples is still ongoing, and the IFBA database is continually being developed and improved upon.

5.2. Conclusions & future work

Machine Learning is revolutionizing many scientific disciplines, including marine science (Malde et al., 2019). Indeed, a large amount of images of fish scales or otoliths (in association with their interpreted age data), could be used to develop an image recognition algorithm (Mapp et al., 2017; Moen et al., 2018). The application of machine learning techniques to the analysis of calcified structures could automate fish aging. The IFBA system presents a digital and physical infrastructure to facilitate such application. Furthermore, with the advancements in genetic techniques, biomineral archives could also be seen as analogous to DNA banks, which can be critical to document changes in the genetic structure of fish populations. The IFBA infrastructure could also be adapted for any research institute or field office, and has the potential to catalogue material and images from other biological sample types, such as corals or shells.

Funding

This work (Grant-Aid Agreement No. PBA/FS/16/03) was carried out with the support of the Marine Institute and is funded under the Marine Research Programme by the Irish Government.

Additional support was provided by the Irish Government and the European Maritime & Fisheries Fund as part of the EMFF Operational Programme for 2014–2020.

This work was part carried out under the COMPASS project. This project is supported by the European Union's INTERREG Iv Programme, managed by the Special EU Programmes Body (SEUPB). The views and opinions expressed in this document do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is the result of over 90 years of sampling effort, and we express our gratitude to all of the people, known and unknown, who contributed to the collection. The authors would like to thank all of the technicians and staff at the Marine Institute, Newport, Co. Mayo, Ireland. The authors especially wish to thank: Oisin Burke, Tara Keena, Seán Kelly, Louise Vaughan, Sarah Ryan, Rob Fuller, David and Kathleen Sweeney, Pat Nixon, Pat Hughes, Ger Rogan, Michael Murphy, Mary Dillane, Russell Poole, Ryan Magee, Signe Martin, Aisling Doogan, Stephanie Ronan, Olga Lyashevska, Katie Thomas, Sarah Flynn, Phillip McGinnity, Ciar O'Toole, David Currie, and Deirdre Cotter for the significant amount of expertise and support contributed to this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2020.101115.

References

Adam, M., Theeten, F., Herpers, J.M., Vandenberghe, T., Semal, P., Van den Spiegel, D., Duchesne, P.A., 2019. DaRWIN: an open source natural history collections data management system. Biodivers. Informa. Sci. Standards 3, e39054. https://doi.org/10.3897/biss.3.39054.

Adey, E.A., Black, K.D., Sawyer, T., Shimmield, T.M., Trueman, C.N., 2009. Scale microchemistry as a tool to investigate the origin of wild and farmed *Salmo salar*. Mar. Ecol. Prog. Ser. 390, 225–235. https://doi.org/10.3354/meps08161.

Arriaga, J.G., Cody, M.L., Vallejo, E.E., Taylor, C.E., 2015. Bird-DB: a database for annotated bird song sequences. Ecol. Informa. 27, 21–25. https://doi.org/10.1016/j.ecoinf.2015.01.007.

Begg, G.A., Friedland, K.D., Pearce, J.B., 1999. Stock identification and its role in stock assessment and fisheries management: an overview. Fish. Res. 43, 1–8. https://doi.

org/10.1016/S0165-7836(99)00062-4.

Brennan, S.R., Schindler, D.E., Cline, T.J., Walsworth, T.E., Buck, G., Fernandez, D.P., 2019. Shifting habitat mosaics and fish production across river basins. Science 364 (6442), 783–786. https://doi.org/10.1126/science.aav4313.

Campana, S.E., Thorrold, S.R., 2001. Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? Can. J. Fish. Aquat. Sci. 58 (1), 30–38. https://doi.org/10.1139/cjfas-58-1-30.

Clarke, A., Margetts, H., 2014. Governments and citizens getting to know each other? Open, closed, and big data in public management reform. Policy Internet 6 (4), 393–417. https://doi.org/10.1002/1944-2866.POI377.

Conze, R., Lorenz, H., Ulbricht, D., Elger, K., Gorgas, T., 2017. Utilizing the international geo sample number concept in continental scientific drilling during ICDP expedition COSC-1. Data Sci. J. 16 (2). https://doi.org/10.5334/dsj-2017-002.

Copp, A.J., Kennedy, T.A., Muehlbauer, J.D., 2014. Barcodes area a useful tool for labelling and tracking ecological samples. Emerg. Technol. 95 (3), 293–300. https://doi.org/10.1890/0012-9623-95.3.293.

Cox, S.J.D., 2016. Ontology for observations and sampling features, with alignments to existing models. Semantic Web 8 (3), 453–470. https://doi.org/10.3233/sw-160214.

Craglia, M., Annoni, A., 2007. INSPIRE: An Innovative Approach to the Development of Spatial Data Infrastructures in Europe. Research and Theory in Advancing Spatial Data Infrastructure Concepts. pp. 93–105.

Dassié, É., Delong, K., Kilbourne, H., Williams, B., Abram, N.J., Brenner, L., Brahmi, C., Cobb, K.M., Corrège, T., Dissard, D., Emile-Geay, J., 2017. Saving our Marine Archives.

EPA, 2015. An Approach to Characterisation as part of Implementation of the Water Framework Directive. Environmental Protection Agency. epa.ie/pubs/reports/water/other/wfdcharacterisationapproachmay2015.html.

Farley, E.V., Murphy, J.M., Adkison, M.D., Eisner, L.B., Helle, J.H., Moss, J.H., Nielsen, J., 2007. Early marine growth in relation to marine-stage survival rates for Alaska sockeye salmon (*Oncorhynchus nerka*). Fish. Bull. 105, 121–130.

Fegraus, E.H., Andelman, S., Jones, M.B., Schildhauer, M., 2005. Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. Bull. Ecol. Soc. Am. 86 (3), 158–168.

Flem, B., Moen, V., Finne, T.E., Viljugrein, H., Kristoffersen, A.B., 2017. Trace element composition of smolt scales from Atlantic salmon (*Salmo salar* L.), geographic variation between hatcheries. Fish. Res. 190, 183–196. https://doi.org/10.1366/0003702053085188.

Fraile, I., Arrizabalaga, H., Santiago, J., Goni, N., Arregi, I., Madinabeitia, S., Wells, D.R.J., Rooker, J.R., 2016. Otolith chemistry as an indicator of movements of albacore (*Thunnus alalunga*) in the North Atlantic Ocean. Mar. Freshw. Res. 67 (7), 1002–1013. https://doi.org/10.1071/mf15097.

Friedland, K.D., Hansen, L.P., Dunkley, D.A., MacLean, J.C., 2000. Linkage between ocean climate, post-smolt growth, and survival of Atlantic salmon (*Salmo salar* L.) in the North Sea area. ICES J. Mar. Sci. 57, 419–429. https://doi.org/10.1006/jmsc.1999.0639.

Froese, R., Pauly, D., 2010. FishBase.

Gries, C., Gahler, M.R., Hanson, P.C., Kratz, T.K., Stanley, E.H., 2016. Information management at the north temperate lakes long-term ecological research site - successful support of research in a large, diverse, and long running project. Ecol. Informa. 36, 201–208. https://doi.org/10.1016/j.ecoinf.2016.08.007.

Grønkjær, P., Pedersen, J.B., Ankjærø, T.T., Kjeldsen, H., Heinemeier, J., Steingrund, P., Nielsen, J.M., Christensen, J.T., 2013. Stable N and C isotopes in the organic matrix of fish otoliths: validation of a new approach for studying spatial and temporal changes in the trophic structure of aquatic ecosystems. Can. J. Fish. Aquat. Sci. 70 (2), 143–146. https://doi.org/10.1139/cjfas-2012-0386.

Hansen, M.M., Ruzzante, D.E., Nielsen, E.E., Bekkevold, D., Mensberg, K.-L.D., 2002. Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. Mol. Ecol. 11 (12), 2523–2535. https://doi.org/10.1046/j.1365-294x.2002.01634.x.

Hedden, H., 2010. Taxonomies and controlled vocabularies best practices for metadata. Journal of Digital Asset Management 6, 279–284. https://doi.org/10.1057/dam.2010.29.

Herring, J., 2011. OpenGIS® Implementation Standard for Geographic Information-Simple Feature Access-Part 1: Common Architecture [Corrigendum].

Hunter, E., Laptikhovsky, V.V., Hollyman, P.R., 2018. Innovative use of sclerochronology in marine resource management. Mar. Ecol. Prog. Ser. 598, 155–158. https://doi.org/10.3354/meps12664.

Hutchinson, W.F., Oosterhout, C.V., Rogers, S.I., Carvalho, G.R., 2003. Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proc. R. Soc. B Biol. Sci. 270 (1529), 2125–2132. https://doi.org/10.1098/rspb.2003.2493.

Leadbetter, A., Vodden, P., 2015. Semantic linking of complex properties, monitoring process and facilities in web-based representations of the environment. Int. J. Digit. Earth 9 (3), 300–324. https://doi.org/10.1080/17538947.2015.1033483.

Leadbetter, A., Carr, R., Flynn, S., Meaney, W., Moran, S., Bogan, Y., Brophy, L., Lyons, K., Stokes, D., Thomas, R., 2019. Implementation of a data management quality management framework at the marine institute, Ireland. Earth Sci. Informa. 1–13. https://doi.org/10.1007/s12145-019-00432-w.

Leadbetter, A., Meaney, W., Tray, E., Conway, A., Flynn, S., Keena, T., Kelly, C., Thomas, R., 2020. A modular approach to cataloguing marine science data. Earth Sci. Inf. https://doi.org/10.1007/s12145-020-00445-w.

Lehnert, K.A., Vinayagamoorthy, S., Djapic, B., Klump, J., 2006. The Digital Sample: Metadata, Unique Identification, and Links to Data and Publications, in American Geophysical Union Fall Meeting 2006 (San Francisco, CA) Abstract # IN53C-07.

Limburg, K.E., Casini, M., 2019. Otolith chemistry indicates recent worsened Baltic cod condition is linked to hypoxia exposure. Biol. Lett. 15 (12), 2–5. https://doi.org/10.1098/rsbl.2019.0352.

Limburg, K.E., Walther, B.D., Lu, Z., Jackman, G., Mohan, J., Walther, Y., Nissling, A., Weber, P.K., Schmitt, A.K., 2015. In search of the dead zone: use of otoliths for tracking fish exposure to hypoxia. J. Mar. Syst. 141, 167–178. https://doi.org/10.1016/j.jmarsys.2014.02.014.

López-Duarte, P.C., Fodrie, F.J., Jensen, O.P., Whitehead, A., Galvez, F., Dubansky, B., Able, K.W., 2016. Is exposure to Macondo oil reflected in the Otolith chemistry of marsh-resident fish? PLoS One 11 (9), 1–24. https://doi.org/10.1371/journal.pone.0162699.

MacKenzie, K.M., Trueman, C.N., Palmer, M.R., Moore, A., Ibbotson, A.T., Beaumont, W.R., Davidson, I.C., 2012. Stable isotopes reveal age-dependent trophic level and spatial segregation during adult marine feeding in populations of salmon. ICES J. Mar. Sci. 69 (9), 1637–1645. https://doi.org/10.1093/icesjms/fss074.

Malde, K., Handegard, N.O., Eikvil, L., Salberg, A.B., 2019. Machine intelligence and the data-driven future of marine science. ICES J. Mar. Sci. https://doi.org/10.1093/icesjms/fsz057.

Mao, X., Cai, T., Olyarchuk, J.G., Wei, L., 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21 (9), 3787–3793. https://doi.org/10.1093/bioinformatics/bti430.

Mapp, J., Hunter, E., Van Der Kooij, J., Songer, S., Fisher, M., 2017. Otolith shape and size: the importance of age when determining indices for fish-stock separation. Fish. Res. 190, 43–52. 4.2.. https://doi.org/10.1016/j.fishres.2017.01.017.

Marine Institute, 2019. The Stock Book 2019: Annual Review of Fish Stocks in 2019 with Management Advice for 2020. Marine Institute, Galway, Ireland.

Matta, M.E., Black, B.A., Wilderbuer, T.K., 2010. Climate-driven synchrony in otolith growth-increment chronologies for three Bering Sea flatfish species. Mar. Ecol. Prog. Ser. 413, 137–145. https://doi.org/10.3354/meps08689.

McNutt, M., Lehnert, K., Hanson, B., Nosek, B.A., Ellison, A.M., King, J.L., 2016. Liberating field science samples and data. Science 351 (6277), 1024–1026. https://10.1126/science.aad7048.

Moen, E., Handegard, N.O., Allken, V., Albert, O.T., Harbitz, A., Malde, K., 2018. Automatic interpretation of otoliths using deep learning. PLoS One 13 (12). https://doi.org/10.1371/journal.pone.0204713.

Morrongiello, J.R., Thresher, R.E., Smith, D.C., 2012. Aquatic biochronologies and climate change. Nat. Clim. Chang. 2 (12), 849–857. https://doi.org/10.1038/nclimate1616.

Morrongiello, J.R., Sweetman, P.C., Thresher, R.E., 2019. Fishing constrains phenotypic responses of marine fish to climate variability. J. Anim. Ecol. 88 (11), 1645–1656. https://doi.org/10.1111/1365-2656.12999.

Nonogaki, H., Nelson, J.A., Patterson, W.P., 2007. Dietary histories of herbivorous loricariid catfishes: evidence from δ 13 C values of otoliths. Environ. Biol. Fish 78 (1), 13–21. https://doi.org/10.1007/s10641-006-9074-8.

Ó Maoiléidigh, N., Tray, E., Dillane, M., de Eyto, E., Cooney, J., Hughes, P., Murphy, M., Nixon, P., Sweeney, D., Poole, R., Cotter, D., Rogan, G., 2020. Irish Fish Biochronology Archive (IFBA). Marine Institute, Ireland.

Panfili, J., Pontual, H.D., Troadec, H., Casselman, J.M., Moksness, E., Wright, P.J., 2002. Manual of Fish Sclerochronology.

Pannella, G., 1971. Fish otoliths: daily growth layers and periodical patterns. Science 173 (4002), 1124–1127. https://doi.org/10.1126/science.173.4002.1124.

Peluso, D., Bean, A., Rose, K., Mcintyre, D., Eck, R., Soeder, D., Crandall, D., Hakala, A., 2016. Geoscience Collection Management Systems: A Beginners Guide.

Peyronnet, A., Friedland, K.D., Maoileidigh, N.Ó., Manning, M., Poole, W.R., 2007. Links between patterns of marine growth and survival of Atlantic salmon *Salmo salar*, L. J. Fish Biol. 71, 684–700. https://doi.org/10.1111/j.1095-8649.2007.01538.x.

Popper, A.N., Ramcharitar, J., Campana, S.E., 2005. Why otoliths? Insights from inner ear physiology and fisheries biology. Mar. Freshw. Res. 56 (5), 497–504. https://doi.org/10.1071/MF04267.

Rivers, P.J., Ardren, W.R., 1998. The value of archives. Fisheries 23 (5), 6–9. https://doi.org/10.1577/1548-8446(1998)023<0006:TVOA>2.0.CO;2.

Short, J.A., Anderl, D.M., 2012. The Age and Growth Program Database. Resource Ecology and Fisheries Management Division. Alaska Fisheries Science Center NMFS, NOAA, 7600 Sand Point Way NE, Seattle WA 98115. Available from. http://www.afsc.noaa.gov/REFM/Age/default.htm.

Sirot, C., Grønkjær, P., Pedersen, J.B., Panfili, J., Zetina-Rejon, M., Tripp-Valdez, A., Ramos-Miranda, J., Flores-Hernandez, D., Sosa-Lopez, A., Darnaude, A.M., 2017. Using otolith organic matter to detect diet shifts in *Bardiella chrysoura*, during a period of environmental changes. Mar. Ecol. Prog. Ser. 575, 137–152. https://doi.org/10.3354/meps12166.

Smoliński, S., Mirny, Z., 2017. Otolith biochronology as an indicator of marine fish responses to hydroclimatic conditions and ecosystem regime shifts. Ecol. Indic. 79, 286–294. https://doi.org/10.1016/j.ecolind.2017.04.028.

Trueman, C.N., MacKenzie, K.M., Palmer, M.R., 2012. Stable isotopes reveal linkages between ocean climate, plankton community dynamics, and survival of two populations of Atlantic salmon (*Salmo salar*). ICES J. Mar. Sci. 69 (5), 784–794. https://doi.org/10.1093/icesjms/fss066.

Vollmar, A., Macklin, J.A., Ford, L., 2010. Natural history specimen digitization: challenges and concerns. Biodivers. Inform. 7 (2). https://doi.org/10.17161/bi.v7i2.3992.

Wells, R.D., Kohin, S., Teo, S.L., Snodgrass, O.E., Uosaki, K., 2013. Age and growth of

North Pacific albacore (*Thunnus alalunga*): implications for stock assessment. Fish. Res. 147, 55–62. https://doi.org/10.1016/j.fishres.2013.05.001.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin Core: an evolving community-developed biodiversity data standard. PLoS One 7 (1). https://doi.org/10.1371/journal.pone.0029715.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3. https://doi.org/10.1038/sdata.2016.18.

WoRMS Editorial Board, 2020. World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2020-01-20. https://doi.org/10.14284/170.

Zeigler, J.M., Whitledge, G.W., 2011. Otolith trace element and stable isotopic compositions differentiate fishes from the middle Mississippi River, its tributaries, and floodplain lakes. Hydrobiologia 661 (1). https://doi.org/10.1007/s10750-010-0538-7. 289–230.