# Converging Open Science and Respecting Indigenous Knowledge to Enrich Capacity of Zooarchaeological Comparative Collections: An Example from the University of Victoria

By

Kathryn Eileen McKenzie

B. Sc., University of Victoria, 2021

A Thesis Submitted in Partial Fulfillment of the

Requirements for a Degree of

MASTER OF ARTS

in the Department of Anthropology



We acknowledge and respect the Lək̓ʷəŋən (Songhees and Esquimalt) Peoples on whose territory the university stands and the Lək̓ʷəŋən and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

# Supervisory Committee

Converging Open Science and Respecting Indigenous Knowledge to Enrich Capacity of Zooarchaeological Comparative Collections: An Example from the University of Victoria

By

Kathryn Eileen McKenzie

B. Sc., University of Victoria, 2021

**Supervisory Committee**

Dr. Iain McKechnie, Supervisor

Department of Anthropology

Dr. Stephanie Calce, Departmental Member

Department of Anthropology

# Abstract

Zooarchaeological comparative collections, like natural history collections, hold latent information, and are fundamental to archaeological research on human-animal relationships. Additionally, these collections can extend their capacity with linked data informing biodiversity research, conservation efforts, and related contemporary and Indigenous management practices. Accessible digital information about specimens in these smaller, and usually regional, collections remain rare but can advance integrative synthetic research through links to taxonomic classifications, languages, as well as geospatial, biometric data, and 3D models and imagery. My research presents a framework for open comparative collection curation, enhanced zooarchaeological practices, and transdisciplinary collaboration by transforming the physical archive describing the comparative osteology specimens at the University of Victoria Zooarchaeology Lab into open “extended specimens” for 2,922 individual animals representing 671 distinct species. This diverse regional collection influentially informs zooarchaeological identifications for assemblages from sites across the North Pacific Coast and western North America. My research synthesizes information about the comparative collection including the development and application of data management, annotation, and publishing methods following FAIR (Findable, Accessible, Interoperable, Reuseable) principles to facilitate broader collection discovery and use. To achieve this, I adopt open data standards to uncover, broaden, and add depth to each skeletal specimen and enable integrative biodiversity repository publishing. This process creates citable “extended specimens” and ensures comparability by standardizing vocabulary and terminology, and annotating with life history stages, collection locations, and specimen specific details. Additionally, I develop a geocoding tool that connects Indigenous language areas and specimen collection locations. This work supports Indigenous data governance protocols following CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles and engages with Indigenous data platforms to confront how colonial practices are reflected in the creation and uses of anthropological and archaeological knowledge. This augmented collection helps bridge relationships with Indigenous communities whose legacies of engagement with archaeology has shaped, and continues to shape and enrich, landscapes and seascapes in the past, present, and future. This contribution to open science seeks to respect

Indigenous data sovereignty by considering FAIR and CARE principles to create a digital resource that connects audiences and enhances zooarchaeological research capacity.

Keywords

Archaeology; Biodiversity; CARE; Collection Management; Darwin Core; Data Governance; Data Management; Ecology; FAIR; Indigenous Knowledge; Linked Open Data; Natural History; Northwest Coast; Open Science; Osteology; Zooarchaeology

# Table of Contents

*Supervisory Committee* ..... *ii*

*Abstract* ..... *iii*

*Table of Contents* ..... *v*

*List of Tables* ..... *viii*

*List of Figures* ..... *ix*

*Preface* ..... *xi*

*Acknowledgements* ..... *xii*

*Dedication* ..... *xiv*

*Abbreviations and Acronyms* ..... *xv*

*Chapter 1: Introduction* ..... *1*

1.1 Introduction ..... 1

Zooarchaeology's unique position ..... 1

Zooarchaeological comparative collections ..... 3

Open science, FAIR, and CARE principles ..... 5

Thesis Structure ..... 8

1.2 Research Questions ..... 9

1.3 Theoretical Framework ..... 14

Archaeological approaches ..... 14

Personal Statement ..... 17

***Chapter 2: Open Science and FAIR Collection Case Study: The University of Victoria Zooarchaeology Lab Comparative Collection*** ***.......19***

2.1 Chapter Summary.......19

2.2 The University of Victoria Zooarchaeology Lab Comparative Collection.......20

Creating a FAIR open collection.......21

2.3 Methods.......23

Workflow processes.......23

2.4 Results.......33

FAIR open collections.......33

Extended specimens and linked data.......44

2.5 Discussion and Recommendations.......52

A path for FAIR open collections.......52

Planning for future capacity.......60

***Chapter 3: Facilitating an Ecology of Knowledges and Implementing CARE Principles in Zooarchaeology Collections*** ***.......64***

3.1 Chapter Summary.......64

3.2 Introduction.......65

Toward an ecology of knowledges: Creating epistemological relationships.......66

Creating a relationship between knowledges with open collections.......68

3.3 Methods.......72

CARE Principles in a FAIR Open Collection.......72

3.4 Results.......81

Specimens with coordinates.......81

Indigenous language areas map.......82

3.5 Discussion.......90

Animal taxonomies and Indigenous nomenclature.......90

Open collections with Indigenous language areas and animal names ............ 93

***Chapter 4: Conclusions*** ............ ***96***

4.1 Review of Research Questions ............ 96

4.2 Further Steps for Open Collections ............ 99

4.3 Implications ............ 100

Zooarchaeological methods ............ 100

Indigenous resource management, biodiversity, and conservation ............ 102

***Appendix: Supplementary Information for Chapter 2*** ............ ***104***

5.1 Data Management Plan ............ 104

5.2 R Code ............ 105

5.3 Darwin Core Terms ............ 106

***A Compendium and Glossary of Terms for Open Collections*** ............ ***113***

6.1 FAIR Principles ............ 113

6.2 Open access methods and tools ............ 118

Data management plans (DMP) ............ 118

Metadata standard – Darwin Core ............ 118

Linked data ............ 119

Workflow platforms and storage ............ 125

Data repositories and registries ............ 127

***References*** ............ ***129***

# List of Tables

**Table 1** Trait attributes recorded on excel spreadsheets and catalogue cards ............................27
**Table 2** Assemblage search results of 31 online zooarchaeological comparative collection. Orange section indicates United States/North America/Regional collections. Green section indicates NWC/Northern North America. Blue section indicates European/British/Continental/Middle East. Yellow section indicates Southeast Asia/Pacific. *Numbers are approximate...........................................................................................35
**Table 3** How UVicZL Project Data meets FAIR Principles......................................................39
**Table 4** Specimens with taxonomic serial numbers (TSN) by provider....................................45
**Table 5** Animal specimens (n=561) following the tDAR faunal sex ontology terminology. The sex field <dwc:sex> contains the term and link..........................................................................50
**Table 6** Animal specimens (n=360) following the tDAR faunal age ontology terminology. The lifestage field <dwc:lifestage> contains the term and link........................................................50
**Table 7** Individual specimens by element type. Records from 1,412 specimens were transcribed with element related data. The preparations field <dwc:preparations> for each was updated with the tDAR element ontology terminology and link and/or the OLS UBERON terminology and PURL link...........................................................................................................................51
**Table 8** Darwin Core location information fields used to derive geocoordinates. The field <dwc:occurenceRemarks> contains verbatim location information from catalogue cards used to manually populate location fields. Each location field provides additional detail for the geocoder to refine the coordinate results.........................................................................................76
**Table 9** CSV results for Google Maps, Mapbox V6, and OpenCage geocoder versions ............82
**Table 10** Number of UVicZL comparative specimens present across 58 Indigenous language areas returned from the Native Land Digital API representing 84% (2,124/2,523) of manually derived coordinates. ...........................................................................................................85
**Table 11** X̱aad Kil / X̱aaydaa Kil (Haida) animal names from FirstVoices and Darwin Core terminology (occurrenceID, scientificName, and vernacularName) of the ten specimens shown in the X̱aad Kil / X̱aaydaa Kil (Haida) language area from Figure 20 ...........................................89
**Table 12** Darwin Core terms, metadata description, and examples ........................................106
**Table 13** FAIR Principles and Objectives. Modified from Australian Research Data Commons (ARDC) FAIR Data Self Assessment Tool(https://ardc.edu.au/resource/fair-data-self-assessment-tool/) .........................................................................................................................115

# List of Figures

**Figure 1** UVicZL Workflow diagram showing data inputs, seven processes, document creation, data outputs, database, and display results. Processes 1-6 are described in Methods, process 7 is described in chapter 3. ....25
**Figure 2** Fields described and data entry from specimen catalogue card. Darker blue rectangles represent original fields from excel spreadsheets and catalogue cards. Brighter blue rectangles represent standardized Darwin Core terms for field names. ....29
**Figure 3** OSF catalogue card UVicZL-00.09 with GUID https://osf.io/ktvxq ....39
**Figure 4** UVicZL dataset download events by GBIF users since 2021 (13,522) as of July 2024. ....41
**Figure 5** UVicZL occurrence data uses and citations in GBIF since 2021 (53) as of July 2024. .42
**Figure 6** UVicZL occurrence data has 53 literature citations as of July 2024, including 5 simple citations and 48 data use citations. The dataset has been cited in 4 peer reviewed open access journal articles and 1 GBIF dataset ID listing. GBIF-mediated data has been substantially used in 27 (25 peer reviewed, 16 open access) journal articles, 10 (9 open access) preprints, 4 theses, 4 peer reviewed book sections, 2 IUCN Red List reports, and 1 computer program. ....42
**Figure 7** Geospatial data points generated by GBIF of UVicZL specimen collection location coordinates. ....43
**Figure 8** Number of specimens by country. Of the 2,519 specimens with geospatial data points, Canada has the most numerous with 1,891 specimens followed by USA with 550. ....44
**Figure 9** Number of species in the UVicZL collection (n=671) grouped by broad common taxonomic class names. ....46
**Figure 10** The top ten orders of the 2,922 animal specimens categorized by broad common taxonomic class names. These include 1,258 fish, 872 birds, 369 mammals, 387 invertebrates, and 36 reptiles and amphibians. 1,012 of the 1,258 (80%) fish belong to one class, Actinopterygii, and five orders (Scorpaeniformes, Perciformes, Salmoniformes, Pleronectiformes, and Gadiformes). 562 of the 872 (64%) birds belong to one class, Aves, and three orders (Passeriformes, Anseriformes and Charadriiformes). 250 of the 369 (68%) mammals belong to one class, Mammalia, and two orders (Carnivora and Rodentia). ....47
**Figure 11** Top ten animal specimens by taxonomic order and family. There are 385 specimens from the order Scorpaeniformes made up of 138 Cottidae, 129 Hexagammidae, 60 Sebastidae, and 58 from other families. There are 258 specimens from the order Passeriformes and 203 specimens from the order Perciformes. There are 187 specimens from the order and family Salmoniformes Salmonidae and 172 specimens from the order and family Anseriformes Anatidae. Of the 170 Carnivora specimens 74 are from the Canidae family, 46 from the Mustelidae family and 50 from other families. There are 141 specimens from the order Pleuronectiormes and 132 specimens from the order Charadriiformes. Of the 96 Gadiformes specimens 76 are from the Gadidae family, and 20 from other families. There are 80 specimens from the Rodentia order. ....48
**Figure 12** The top ten species by order and family. Salmoniformes Salmonidae contain 143 specimens represented by five salmon species *O.mykiss* (33), *O. tshawytschia* (31), *O. clarkii* (31), *O. keta* (25), and *O. kisutch* (23).There are 96 specimens of *P. monoterygius* from the order and family Scorpaeniformes Hexagrammidae . There are 63 specimens of *C. lupus* from the order and family Carnivora Canidae. There are 26 specimens of *G. chalcogrammus* from the

order and family Gadiformes Gadidae and 23 specimens of *A. platyrhynchos* from the order and family Accipitriformes Accipitridae........................................................................................49
**Figure 13** UVicZL-92.64 as an "extended specimen" with links to taxonomies, ontologies, geospatial data, images, related research, and Indigenous language areas. ................................59
**Figure 14** The biodiversity informatics landscape with connections to GBIF in 2017. Source: modified from Bingham et al. 2017. ........................................................................................63
**Figure 15** Map of 2,523 specimens with geocoordinates with Indigenous language areas outlined in black from Native Land Digital. Location data from which geocoordinates could be derived is available for 86% (n=2,523) of the specimen records in the collection. Each specimen is indicated by a clickable blue teardrop pin on the map which is linked to the specimen record in the dataset. Specimens located near each other are grouped by the coloured circles with number figures. ...................................................................................................................83
**Figure 16** Specimens collected from Puget Sound. Searching the waterBody field for "Puget Sound" results in 70 specimens visualized on the map..........................................................83
**Figure 17** Hakai Native Lands Digital API query for specimen locations based on manually derived coordinates from Mapcarta. Most specimens are from British Columbia and Washington State, as well as around the Aleutian Islands in Alaska. ........................................................86
**Figure 18** An example of specimen numbers within a specific Indigenous language area. .........87
**Figure 19** Information window shows detail for specimen collected from this location. In this example the field 'IndigenousName' is shown as an example where the animal name, "hatsawe'"[13], from the Indigenous language area Kwak̓wala could be inserted..........................87
**Figure 20** X̱aad Kil / X̱aaydaa Kil (Haida) language area showing ten specimens from the native language facet map including an isolated specimen on the north end (blue teardrop). ................88
**Figure 21** Open to Collaborate Notice from Local Contexts....................................................92
**Figure 22** Biocultural Notice from Local Contexts.................................................................92
**Figure 23** Data Management Plan.......................................................................................104
**Figure 24** GBIF ITIS TSN Rscript written to pull ITIS numbers from the GBIF API .............105
**Figure 25** tDAR Faunal Age Ontology depicting hierarchical levels (Source - doi:10.6067/XCV8PZ574Q)..................................................................................................124

# Preface

Chapter 2 and 3 are written to be submitted to peer-reviewed journals as co-authored manuscripts. Substantial knowledge contributions and editing have been made by Iain McKechnie.

# Acknowledgements

I would like to acknowledge the numerous territories from which the UVicZL specimens were collected where the Indigenous peoples of those territories stewarded their land for thousands of years, creating abundant food systems, upholding community governance, and developing vast trade networks. It is with great colonial privilege that I complete this thesis today from my home in ancestral and unceded Quw'utsun territory. It is my hope that we may begin to reconcile our histories and, with respect to those whose connection to this land was and remains to be intrinsically linked to language, reclaim names that were formerly replaced. This thesis seeks to imagine a world where settlers and Indigenous communities develop relationships to hear, write, and speak local languages and use Indigenous names as daily vocabulary.

I would also like to acknowledge Rebecca 'Becky' Wigen, for collecting and curating a majority of the UVicZL zooarchaeology comparative collection. Without Becky's incredible contribution to the creation of the collection, archaeology of the NWC of North America would not be the same. Her commitment to, and expertise in faunal analysis is truly inspiring.

Thank you to my supervisor, Dr. Iain McKechnie for all your insight, support, and guidance. Your knowledge and expertise have been invaluable. Who knew digitizing catalogue cards would lead down this path! Thank you to my committee member, Dr. Stephanie Calce, for all your advice and mentorship. Thank you to my external examiner, Dr. Emily Lena Jones, for your time and expertise. Also, thank you to all the HECA Lab members, past and present, including Dylan Hillis, Taylor Vollman, Robert Gustas, Katie Dierks, Alyssa Ball, and Kristina Barclay. Special thanks to Freya Harrison, who spent countless hours transcribing catalogue cards. Thanks to Dr. Marla McKinnon for your patience and time spent sharing expertise in 3D scanning.

Sincere thanks to Eric Peterson and Christina Munk of the Tula Foundation for financial support and to Hakai's IT and geospatial teams for my internship guidance and assistance. I'd especially like to thank Steve Van Der Valk for developing the geocoders and Brett Johnson, Tim Van Der Stap, and Will McInnis for all the R training, code writing, and geolocation assistance. Also, thanks to Standardizing Marine Biological Data (SMBD) members, Dr. Stephen Formel and Abigail Benson of the USGS, for coding, GBIF, and DwC assistance. Thank you to Carole Sinou, manager of the Canadensys network, for assistance publishing the collection to the repository.

Thank you to the Department of Anthropology faculty, staff, and cohorts for an excellent learning experience and support throughout my both undergraduate and graduate degrees. Special thanks to Jindra Belanger for all you do to support graduate students to ensure funding and paperwork are in order. Thanks to Drs. Brian Thom, Helen Kurki, Yin Lam, Ann Stahl, and Ran Donaldson for imparting your wisdom, experience, and knowledge. Thank you to Dr. Marie Hopwood who inspired my journey into Anthropology and prompting me to take the path not otherwise taken.

And finally, my deepest gratitude for the love and support of my close friends and family. This includes my five wonderful children, Scot, Nico, Cera, Elena and Namieka, and Lisa Rush, who followed me through this journey. Thank you especially to my partner, Jessop Torrie, for your love, encouragement, and patience.

# Dedication

This thesis is dedicated to my father, Dr. Alvin H. McKenzie, a lifelong learner who has been my inspiration for acquiring knowledge.

# Abbreviations and Acronyms

| | |
|---|---|
| ARK | Archival Resource Keys |
| API | Application Programming Interface |
| CARE | Collective benefit, Authority to control, Responsibility, and Ethics |
| CoL | Catalogue of Life |
| DMP | Data Management Plan |
| DOI | Digital Object Identifier |
| DwC | Darwin Core |
| EA | Effective Archaeology |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| FNIGC | First Nations Information Governance Centre |
| GBIF | Global Biodiversity Information Facility |
| GIDA | Global Indigenous Data Alliance |
| GRSciColl | Global Registry of Scientific Collections |
| GUID | Globally Unique Identifier |
| HCA | Heritage Conservation Act |
| HE | Historical Ecology |
| ICZN | International Code of Zoological Nomenclature |
| ITIS | International Taxonomic Information System |
| ITP | Integrated Publishing Toolkit |
| IUCN | International Union for Conservation of Nature |
| LOD | Linked Open Data |
| MoF | Measurement or Facts |
| NCBI | National Center for Biotechnology Information |
| NGO | Non-Governmental Organization |
| NWC | Northwest Coast |
| OBIS | Ocean Biodiversity Information System |
| OCAP® | Ownership, Control, Access, and Possession |
| OLS | Ontology Lookup Service |

| | |
|---|---|
| ORCID | Open Researcher and Contributor ID |
| OSF | Open Science Framework |
| PID | Persistent Identifier |
| PURL | Persistent Uniform Resource Locator |
| SQL | Structured Query Language |
| tDAR | The Digital Archaeological Record |
| TDWG | Taxonomic Databases Working Group |
| TSN | Taxonomic Serial Number |
| URI | Uniform Resource Indicator |
| URN | Uniform Resource Name |
| URL | Uniform Resource Locator |
| UNDRIP | United Nations Declaration of Rights of Indigenous Peoples |
| UUID | Universal Unique Identifier |
| UVic | University of Victoria |
| UVicZL | University of Victoria Zooarchaeology Lab |
| WoRMS | World Registry of Marine Species |
| ZCC | Zooarchaeological Comparative Collection |
| ZooMS | Zooarchaeology by Mass Spectrometry |

# Chapter 1: Introduction

## 1.1 Introduction

This thesis explores ways to increase access to and broaden the capacity of a zooarchaeological comparative collection (ZCC) using principles of open science and ethical methods in anthropological research and collections management. I examine what makes a comparative collection useful and consider ways to improve its accessibility to those who use it currently. I identify the significant hidden information a collection holds, what other information it has capacity for, and how this information can be disseminated while respecting ethical considerations. Such disciplinary self-reflection is key to ensuring anthropological information is shared in ways that recognize and look for solutions to historical harms. Natural history collection managers and, more broadly, the disciplines of natural and social science are increasingly embracing open science tools and principles. My research examines open science applicability to ZCCs. To identify opportunities and challenges in collection management, I evaluate best practices to determine where gaps occur and how these might apply to the University of Victoria's Zooarchaeology Lab (UVicZL) comparative collection. This work links, extends, and converges the past, present, and future to create a framework that is accessible and empowering for researchers, students, and especially First Peoples, the original knowledge holders.

### Zooarchaeology's unique position

Practitioners in anthropology, archaeology, palaeoanthropology, and subdisciplines like zooarchaeology and archaeobotany, who often integrate data from zoology, botany, ecology, and conservation, should have an appreciation of current interdisciplinary issues that are relevant to both social sciences and biological sciences. Social science disciplines are also key to bringing a broader perspective to the natural sciences and a greater understanding of deep-time human impacts on species interactions, distributions, and extinctions. As anthropologists we bring cultural awareness and disciplinary introspection on the history of our professions and the harms created and hold as our legacy. This greater consideration of acknowledging extractive and salvage orientated practices directly reflects forward thinking actions and aspirations which advocate for more representative interpretation and participation, increased knowledge accessibility, and diverse perspectives.

Zooarchaeology is an increasingly key aspect of archaeological interpretation as identification of animal species from archaeofaunal assemblages can shape analysis and interpretation of human history. Faunal assemblages are studied to interpret past environments, human resource use, cultural adaptations, ecological impact, and changes to animal ecology. Zooarchaeologists also study animal domestication and population management practices including community structure, diet, and habitat by identifying the species, age, or sex of animals based on preserved skeletal elements in a site, as well as human and taphonomic modifications. The study of animal remains in archaeological sites often reveals enduring human-animal relationships that have impacted evolutionary and ecological aspects of both animals and human society. Understanding the history and potential trajectory of these long-term relationships is pertinent to contemporary conservation, resource management, and cultural heritage preservation.

Zooarchaeology spans the natural and social sciences and is well placed to combine biology, ecology, and social science theory to broaden and illuminate past human-animal interactions and relationships. Zooarchaeological data from archaeofaunal remains embodies past animal populations viewed through a human activity lens and provides spatiotemporal biogeographical and settlement patterns, provisioning networks, and technological innovation data (Peres 2010, Orton 2016). This environmental and cultural evidence encompassing thousands of years, provides an extended perspective that can contribute to our understanding of ecosystem change over time, cultural preferences, and resource use in the past (McKechnie and Moss 2016). As I argue in this thesis, zooarchaeology can reflect inclusive archaeological approaches of Effective Archaeology (EA) and Historical Ecology (HE), open science principles, and Indigenous data sovereignty. These practices improve methods, expand audiences, and broaden knowledge discourse. My research, centered on the University of Victoria Zooarchaeology Lab (UVicZL) comparative collection, looks for ways to contribute to the discipline theoretically using open science to enable better discovery and access, and in practice with applied collection management methods.

**Zooarchaeological comparative collections**

Zooarchaeological comparative collections (ZCCs) contain osteological (skeletonized) specimens of known modern animal species and are used as morphological guides in the identification of fragmentary archaeological remains. Collections vary in composition and are often limited by size and species variation. How is each specimen is arranged and catalogued, how species are organized by taxonomic categories, and where this information is kept are key questions researchers may have. While there are a wide range of conventions, most collections track the provenience of each specimen and additionally, details about the animal recorded during collection and processing. Robust collections are representative of regional contemporary animals and include small and large taxa, with multiple single species specimens of varying age (Lyman 2010). These collections, often termed reference collections, are essential in zooarchaeology as the representative comparative specimens contained within them are crucial to the accurate taxonomic identification of fragmentary faunal remains as well as estimates of sex, size, and age (Betts et al. 2011, Driver 2011, Lyman 2010, McKechnie et al. 2020, Meighan et al. 1958, Niven et al. 2009). Archaeofaunal remains are key indicators that facilitate archaeological, biogeographical, ecological, and ethnobiological interpretation of the dynamic and interactive history of human-animal relationships. Accurate identification of these fragmentary archaeological bones and shells influences research questions and the direction of future studies beyond zooarchaeology.

Using ZCCs for identification to a genus and species level from archaeological contexts has consequences for the interpretation of archaeological and environmental phenomena. ZCCs allow researchers to answer the fundamental question: What bone is it? The identification process begins at an archaeological site, where animal remains are recovered, and then moves into a lab setting with a comparative collection. A primary objective is confident identification of anatomical element and taxon (Gifford-Gonzalez 2018) following the standard method of comparative morphology (including bone shape, size, and texture). This entails matching a zooarchaeological specimen to a modern 'known' analogue by element and taxa (Lyman 2019). At an analyst's discretion, when the archaeological specimen appears practically identical to a collection specimen, it is identified as a likeness of the modern analogue. Although modern analogues may not exactly mirror archaeological specimens, extensive regional collections do provide a broad overview of species that inhabit contemporary ranges that may have been present in the past. In many cases, accurate identification beyond genus requires excellent preservation of morphometrically diagnostic elements, a comprehensively representative comparative collection, and species-specific osteological criteria. Often when bone fragments may only be distinguishable by skeletal element, yet not identifiable beyond class or order, these remains are placed into size descriptive family-order groups such as “large ungulate” to assist with comparative analysis (Driver 2011). Precise identification is difficult and although it may be possible to identify to the species level using anatomical methods, it remains rare to identify to subspecies. Using regional ZCCs, confidence in the identification to a species-level has implications for the interpretation of human-animal relationships (Lyman 2017, Monfils et al. 2020).

**Open science, FAIR, and CARE principles**

Open science is a set of methods and practices that take advantage of an increasingly connected world to generate opportunities to connect research data and embrace accessible digital data stewardship. The traditional practice of keeping scientific knowledge guarded and inaccessible has shifted in recent years to move toward transparent and publicly available data by promoting transparency, accessibility, and reproducibility. This shift may be part of a larger change in social consciousness seeking to recognize and amend inequalities, and as a coordinated response to global ecological crises and climate change. Sharing knowledge efficiently in a network can quickly lead to insights and solutions not otherwise possible in isolation. These practices advance scientific research to make data and methods more open and available for scholarly and public use. In zooarchaeology, improving data stewardship, analytical transparency, collaborative expertise, and public involvement can be addressed by adopting open science practices (Marwick et al. 2017). Open science practices encourage researchers to make research findings, methods, and data accessible and preferably published in open access journals. Open data is a component of open science where datasets, typically in digital form, are freely accessible and supplemented with metadata that provide standardized data descriptors. Best practices in data stewardship encourage researchers to collect and share their datasets by uploading them to public repositories, data aggregators and workflow platforms.

In archaeology researchers have noted that improved access to digital information from varied and more refined data sources can stimulate research collaboration, more integrative research questions, and the reuse of scientific data in meaningful ways (Kintigh et al. 2018, Faniel et al. 2018). Datasets can be synthesized to create meta-analyses that incorporate larger spatiotemporal scales to better understand phenomena such as domestication and traditional hunting practices, and cultural preferences (Kintigh et al. 2018). Open science is improving standardization in zooarchaeological data, particularly via metadata documentation and well defined terminology, vocabularies, and ontology categories, to allow for better analytical transparency and the integration of datasets. Repository infrastructure, open access journals, and data management practices have emerged to support digital access to knowledge and broaden public engagement. Open science also facilitates education and outreach programs and improved scientific literacy in popular media coverage. Attention to the significance and implications of human-animal relationships of the past, present, and future is achieved by disseminating zooarchaeological knowledge. Open science practices hold the potential to generate entirely novel knowledge and enhance transdisciplinary collaboration, increased public engagement, and greater awareness of zooarchaeological comparative collections (ZCCs). Accessing this information about natural history is critical as impacts of climate change are leading to shifts and extinctions as well as the loss of cultural heritage (Hollesen 2022). The deep-time perspective provided by zooarchaeological evidence of cultural heritage requires an open science approach to bridge knowledge gaps and ensure preservation for future generations.

FAIR is an acronym that stands for Findable, Accessible, Interoperable, and Reusable and is described by specific objectives outlined in Compendium **Table 13**. The FAIR principles emerged in 2016 in relation to the management and stewardship of research data, particularly in the context of the life sciences (Wilkinson et al. 2016). Advocates argue that research data should be easy to find for humans and be machine readable. The FAIR principles are particularly important in research discovery, where data sharing and collaboration are essential for advancing knowledge. These principles aim to make research data more accessible and useful to the broader research community, fostering transparency and collaboration. Associated processes derived from library science, are often laden with technical terminology, but crucial for information sharing. For example, assigning unique identifiers (e.g., DOI), using standardized metadata, and indexing data repositories or catalogues ensures data is more easily found and readily accessible. This means that data should be stored in a way that is easy to retrieve, whether it's stored locally or in a data repository. Access to data should also be granted to all who need it while respecting privacy and security concerns. Data should be in a format that can be used and understood by different systems and applications. To be interoperable it should use common standards and vocabularies to enhance its usability across different contexts. Additionally, data should be documented and structured in a way that makes it easy for others to understand and reuse. This includes providing clear metadata, provenance information, and any necessary context for defining and interpreting the data.

Following the development of FAIR principles social scientists and Indigenous communities highlighted the need to promote and include ethical considerations for the open data, particularly where and for whom research is conducted. Hence the development of CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles for Indigenous data governance by the Research Data Alliance (RDA)[1] International Indigenous Data Sovereignty Interest Group and published by the Global Indigenous Data Alliance (GIDA)[2] in September 2019, were conceived to be congruent with FAIR principles (Carroll et al. 2021). . Essential to managing the data life cycle, CARE principles build on the Indigenous control of data tenets under United Nations Declaration of Rights of Indigenous Peoples (UNDRIP)[3], CARE principles call for Indigenous collective benefit, authority to control, responsibility, and ethics in relation to their data (Gupta et al. 2023). Like the FAIR, CARE principles are based on criteria and require digital infrastructure and tools to be effectively implemented. CARE principles are also aligned with Ownership, Control, Access, and Possession (OCAP®). Both CARE and OCAP® are frameworks for ensuring ethical dissemination of data that respects the rights of Indigenous communities. In contrast, FAIR principles primarily seek to facilitate data discoverability, access, interoperability, and reuse. Open data and non-open data can both be FAIR however, open science principles that facilitate the sharing of Indigenous data fall short of recognizing "relationships, power differentials, and the historical conditions associated with the collection of data (Carroll et al. 2021 p. 3)." Contemporary western science, including archaeology, have frequently been extractive and colonial, but aspects are changing with more awareness of structural inequities (Gupta et al. 2023, Jennings et al. 2023).

**Thesis Structure**

This chapter describes the role of zooarchaeology and ZCCs within the discipline of archaeology and generally how practices of open science apply. Chapters 2 and 3 are interrelated and describe

[1] https://www.rd-alliance.org/
[2] https://www.gida-global.org/
[3] https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_E_web.pdf

my research creating an open zooarchaeological comparative collection following FAIR and CARE principles respectively. Chapter 2 describes the technical aspects of creating a FAIR open collection with Chapter 3 builds on this by describing connected CARE practices within the context of the project. Chapter 4 provides a synthesis of methods and offers recommendations for the users and managers of zooarchaeological comparative collections globally. Finally, an addendum offers a compendium of tools, providers, and methods that can be applied to create an open collection.

## 1.2 Research Questions

My research is motivated by three questions that are framed around zooarchaeological comparative collections (ZCCs) and specifically how zooarchaeology (and archaeology and anthropology, more broadly) is linked and intertwined with life and earth sciences, natural history, and Indigenous and ethnobiological knowledge. As archaeology is increasingly recognizing the importance of information management and use of large-scale datasets (Anderson 2024), zooarchaeological research has become increasingly influential to interdisciplinary and collaborative approaches. However, not as much attention has been focused on the comparative collections that generate zooarchaeological data.

My first research question asks, first: **Can open science and FAIR principles contribute to collection accessibility and broaden discoverability by innovating documentation methods and connecting with diverse audiences?** To address this question my research identifies primary specimen data for the UVicZL comparative collection then consolidates and formats this data to create a basic and accessible digital archive. I determine the audience who is most interested in the collection currently and the type of information most beneficial to them. I then determine the broader audience and more diverse information to target.

Making this information more available by adopting open science practices improves zooarchaeological identification and use of lab facilities and generates an opportunity to create comparative 'type' specimens.

Understanding of how a zooarchaeological lab facility is arranged and what resources are available to students and researchers is a rarely examined aspect of zooarchaeological practice. Collections that have been in existence for many decades are usually developed by researchers and, in some cases, lab managers who may work with assistants to prepare, label, and accession specimens into a collection. Specimen information often includes collection location, date, donor, and biometric data such as weight and length. Although this primary data may exist in written form, there is no standard operating procedure, and often this key information is not fully digitized. To be useful, a zooarchaeological comparative collection must also easily allow users to compare a range of specimens and find primary data on each specimen. For instance, a researcher may need to know the kind and location of flatfish in a collection, where they were collected, and additional primary data such as weights and lengths. Other questions include whether the collection is organized by articulated specimens, taxonomic categories, elements, or a combination of these.

Recording and digitizing this primary data and making it available to researchers is an essential piece of creating an accessible and comprehensive collection (Peterson et al. 2005). Although specimens are often not recognized in similar ways in which biologists identify taxonomic 'type' specimens (Federhen 2015), the opportunity to recognize and describe comparative 'type' specimens is excluded when primary data do not exist or are not accessible. Omitting such detail reduces confidence in identification. Considering the importance of regional comparative collections for generating identifications across a region over time, documentation about animal specimens in a collection can have important implications for archaeological interpretation.

After I initially focus on discoverability and accessibility, I then explore how information can be shared and reused, which frames my second question: **How can open science tools like linked open data (LOD), data standards, osteological and life-stage terminology, and methods to display geospatial, occurrence, and biometrics broaden collection specimen data and collection use across disciplines?** To address this question, I explore emergent digital collection management practices based on open biodiversity initiatives. I follow the "extended specimen" concept to link latent collection information and data. This question considers the potential for zooarchaeological comparative collections to learn from open science initiatives developed by museum-based natural history collections (NHCs). NHCs hold abundant biodiversity data and preserved biological specimens collected over the past 300 years which serve as regional, national, and global biodiversity records (Colella et al. 2020, Johnson et al. 2023). Despite their importance to conservation research, many NHCs collections currently face challenges including funding and support. This has led to a loss of expertise, facility deterioration and, in some cases, the consolidation and/or disposal of specimens. Globally, there are thousands of NHCs that contain an estimated 3 billion specimens (Cook et al. 2014, Hedrick et al. 2020, Johnson et al. 2023) which are often an untapped source of primary scientific data. Technological advances have allowed NHCs to collaborate in international open science initiatives and share specimen attributes, imaging, audiovisual, and biochemical data (Hedrick et al. 2020, Miller et al. 2020, Johnson et al. 2023). This has enabled smaller *regional* NHCs to participate, strengthening the utility of this data even for modest collections of less than 10,000 specimens (Monfils et al. 2020). Although regional NHCs often strongly represent local diversity and provide insight into unique taxa and geographic distributions, zooarchaeological comparative collections, currently absent in this domain, can help fill a gap in knowledge.

My research seeks to bridge gaps in disciplinary boundaries between archaeology, museums and archives, ecology, biology and related subdisciplines by adopting these information management techniques to broaden access and discovery, and aid in sharing and reuse of data across disciplines and beyond academic borders. Addressing these questions, however, would not be complete without ethical consideration and self-reflection, leading to my third research question: **How can collections adopt CARE principles and connect with Indigenous histories and communities to foster more equitable sharing and reuse of data to positively impact communities and the histories zooarchaeologists study in the present and future?** While open science provides a framework for organizing and networking categorical data, it often ignores and embodies some of the historical inequities of the Western science paradigm. As many have noted, it is necessary for science to acknowledges and consider Indigenous perspectives, inherent rights, and alternative ways of being in relation with biodiversity (Salomon et al. 2023). To foster a more inclusive and collaborative approach to collections management, my research seeks to develop digital infrastructure to acknowledge relationships, attend to colonial histories, and work towards decolonization in zooarchaeology.

For many decades, a small community of ethnobiologists and ethnoecologists have been working with Indigenous peoples to better understand fauna and flora in their territories and how they name and classify organisms. More recently, there have been calls to restore use of Indigenous animal names, and recognize the richness of “folk taxonomies” (Gillman and Wright 2020, Gomes 2020, Franco 2021). Such Indigenous taxonomies differ from the hierarchical classifications following the Linnean system of categorization. These ways of knowing and naming living organisms are based on relationships, differences in observation, and utilitarian practices. Folk nomenclature systems may generate descriptive names for species derived from cultural perceptions in an Indigenous language (Gomes 2020), are rooted in local knowledge and practically reflect decolonial identity (Gillman and Wright 2020, Franco 2021). Considering the majority of the world's Indigenous languages are considered "critically endangered" (Bromham et al. 2022), revitalizing Indigenous animal taxonomy and nomenclature has the potential to help address and bring attention to threatened biodiversity and traditional knowledge. For example, in Hawai'i, the colonial impact on Indigenous people and the ecological landscape led to the extinction of upwards of 68% of known avian species (Gomes 2020) and interrupted the multigenerational transfer of native bird knowledge.

Although there is general agreement that scientific naming is useful and standardization is necessary for practical, educational and research continuity purposes, Indigenous names and common names are important for regional or local identification and are inclusive of different languages. In the public sphere, the changing of names has often been contentious and encompassed broader issues including school, museum, and street name changes. Rather than changing scientific names, open science platforms and tools can enable linkages to Indigenous names that can contribute toward revitalizing knowledge systems and integrate Indigenous ways of knowing in relation to Western science practices. Revitalizing usage of Indigenous names is part of the decolonizing process and a step toward reconciliation with Indigenous peoples who have faced colonial injustices including language loss. In anthropology this is a relevant and reflexive step which helps to recognize our discipline's historical role in colonial practices and policies. Importantly, zooarchaeological collections can provide a platform for engaging with Indigenous communities in this process (Sholts et al. 2016). Local ethnobiological knowledge shared by Indigenous knowledge keepers works to broaden and deepen our understanding of biodiversity and ecosystems (ibid).

For this question, I employ data sovereignty principles in Chapter 3 to identify the regional aspects of collection specimens in relation to their Indigenous histories. I use existing open science and open Indigenous data platforms to enable future collaboration with Indigenous communities whose engagement with the 671 species represented in the collection can provide a foundation for animal names.

## 1.3 Theoretical Framework

### Archaeological approaches

Over the past few decades there has been increasing appreciation and archaeological recognition of the persistence or *longue durée* of Indigenous habitation along the Pacific Northwest Coast (NWC) (Ames 1991, McLaren et al. 2015, McMillan and McKechnie 2015). The coast, with its rich marine biodiversity and dynamic shoreline, has been managed by Indigenous peoples “since time immemorial” (Mathews and Turner 2017 p. 169). To navigate and thrive within this liminal intertidal landscape, where terrestrial and marine environments meet, required closely held long-term knowledge, language, and skill traditions passed on through many generations (ibid) and is reflected in the way animals were named, described, and considered as human relatives across numerous Indigenous languages and communities (Blaser 2016).

Indigenous knowledge, ways of knowing, and “worldview” has been characterized as part of an Indigenous epistemology that often stands in contrast to how academic knowledge and settings are configured. For instance, knowledge is categorized within Western science to become disciplines and methodologies wherein academics and academic departments may claim to authoritatively represent or investigate past and present realities. Similarly, the differing worldviews of peoples across the globe have often been depicted as being in opposition to the predominant knowledge of Western science which purportedly "best *reflects* reality" (Henare et al. 2007 p. 11). In the disciplines of anthropology and archaeology, particularly with respect to Indigenous peoples, the consequence of this Western epistemology is to create a duality that separates 'real' and 'constructed' ontologies (ibid). However, reality requires situated practices, materials, tools, specialized skills, and competencies to exist (Santos 2015). This shift referred to in anthropology as the "ontological turn" attempts to address this epistemological failing (Henare et al. 2007). To reconcile reductionist interpretations of cultural beliefs, the 'turn' envisions alternative ways of being in the world by reconfiguring reality as a "multinaturalism" or many worlds (ibid). This idea of many worlds, however, creates a circumstance in which sets of knowledge may be "incommensurable" (ibid).

Anthropologists have looked for solutions that reconcile incommensurability of worldviews and realities but have sought this through deeply referenced technical writing. Placing the ironies of academic communication aside, these efforts to reconfigure reality risks placing the control of ways of knowing back into the hands of the dominant power relations and effectively reinforces the protection of Western authority by its determination of the nature of reality (Graeber 2015). While the turn attempts to avoid the subject-object enigma, it cannot fully address its disproportionate power imbalance (ibid). As noted by Stahl (2020), anthropologists are left with the question of incommensurability and the challenge of recognizing diversity in non-divisive ways. These issues are pertinent and reflected within zooarchaeology and this thesis as it concerns the names of hundreds of animal species and the multitude of relationships these animals hold for interpreting human and environmental histories. New approaches and interpretive frameworks, such as “effective archaeology” (Stahl 2020), have been developed to address the tensions created by considering history and situated reflexivity. Effective archaeology (EA) attends to anthropology’s historical foundations and envisions a new set of collaborative relationships and dialogue to work though this conceptual incommensurability.

Archaeological research that incorporates a landscape approach provides a longer-term view of time, space, place, and history. Historical ecology (HE) is one such an approach which is motivated by addressing how Indigenous peoples and their history have a role in confronting contemporary ecological and climate crises (Armstrong et al. 2017). HE uses historical and archaeological data to demonstrate ancient human-environmental long-term trajectories, often reflects Indigenous land-use pattern legacies, and can identify solutions to natural or anthropogenic change agents (Hardesty 2007). HE challenges the historical construction of nature as lacking humans and seeks to make new connections between humans and the environment they shaped and created (Balée 1998). It is premised on human and historical events as responsible for changing relationships between humans and their environments. Climatic change, environmental degradation, and development is placing increasing pressure on archaeological heritage sites. Along the coast of British Columbia there exists an urgency for protecting and recovering information from known and unrecorded sites to further document deep time Indigenous history and former management practices. Archaeological recovery and analysis of animal bone specimens from these sites reveal ecological changes that extend historical baselines relevant for conservation and resource management (Fordham et al. 2020). The study of zooarchaeological remains from sites across British Columbia and the Northwest Coast (NWC) supports a deep time perspective on human-environmental interactions over the past 14,000 years of human occupation (McKechnie and Moss 2016, Mackie et al. 2018, McLaren et al. 2020, Gauvreau et al. 2023). An immense amount of biological information contained in archaeological sites highlights the need to prevent their destruction and preserve them as one of “the only available records of past biodiversity” (Fossile et al. 2020 p. 39). HE asks what the archaeological record can reveal about Indigenous agency by tracing intergenerational traditional practices. Indigenous people are seen as active managers of the environment who have structured the landscape though millennia (Armstrong and Veteto 2015).

Effective archaeology and HE approaches redefine the conceptualization of biodiversity and landscape interrelationships between humans and the environment as an ongoing dialectic, applicable to resource use today (Balée 1998). The spatiotemporal lens of EA and HE provide insight on climatic and anthropogenic ecosystem change and variability, persistent or novel landscape features, and species habitation variation (Beller et al. 2020). When applied to ecosystem conservation, restoration, and management this perspective can result in the alteration and prioritization of strategies (ibid). Effective archaeology and HE strive to recognize and acknowledge the deep-time relationships, environmental interactions, and landscape stewardship practices of Indigenous people are as integral social-ecological systems (Mathews and Turner 2017, Beller et al. 2020). By incorporating past and present human agency to assess ecological changes, more accurate historical baselines can be integrated in research. Rather than aiming for “pristine wilderness” restoration, it develops “forward looking” goals for resource management (Beller et al. 2020).

**Personal Statement**

As an anthropological archaeologist I am engaged in continuous learning and unlearning as I consider the knowledge I produce. I acknowledge my positionality as an archaeologist and producer of knowledge whose responsibility is to follow practices that “do no harm” (Stahl 2020

p. 40). I am a white settler of maternal Ukrainian, and paternal Scots Irish ancestry. I lived on Cree lands in Treaty Six Territory on a farm along the Sturgeon River near ᓇᒣᐤ Namêw (Namao, Alberta) for the first half my life. I now reside on unceded Quw'utsun' (Cowichan) territory on Ts'elxwíń (Cherry Point, British Columbia). Where I grew up, the only word of Cree I learned was the placename ᓇᒣᐤ Namêw, which is Cree for sturgeon. I attended the same school from kindergarten to grade twelve and had many friends who were of First Nations ancestry. We were taught the colonial history of Canada, but not once do I recall learning any local history. It is only recently, since returning to university, that I have embarked on an education in Indigenous history. I situate myself as an anthropologist and archaeologist with my place-based history and life lessons learned over the past fifty years and as being open to new, and possibly radical ways of thinking and perceiving through continuous learning and unlearning. Accomplishing this unlearning "requires overcoming forms of ignorance *learned* through and *internal* to, [my Western] epistemology", reconciling my responsibility to decolonize archaeology, and being aware of the impact of my work on Indigenous communities (Stahl 2020 p. 40). This critical engagement recognizes the epistemological hierarchies that have permeated knowledge production and dichotomous ideology. Resolving ontological issues in archaeology must overcome the relativist position created by placing Western thought as autonomous and exempt from critique. Taking a knowledge dialogue approach works to respect and complement other ways of knowing to potentially reveal novel understandings.

# Chapter 2: Open Science and FAIR Collection Case Study: The University of Victoria Zooarchaeology Lab Comparative Collection

## 2.1 Chapter Summary

This thesis research focuses on the University of Victoria Zooarchaeology Lab (UVicZL), a physical lab space and comparative collection used to assist in the identification of fragmentary archaeological bones at Indigenous heritage sites. In this chapter I discuss digital management of the collection, which contains over 2,900 fish, mammals, birds, shellfish, reptiles, and amphibians collected and curated over the past 44 years. My research uses digital techniques to expand the reach and visibility of the Lab. I describe how open science tools and adoption of biodiversity data standards enable collection specimens to be converted into “extended specimens”, defined in detail later in this chapter, and annotated with linked data on taxonomy, ontology, images, and geospatial location. I describe how I implement these methods, including transcription and digitization of catalogue cards to enable data preservation and add further depth to each skeletal specimen. I consider how my research connects and enhances research capacity and integration with biodiversity repositories and workflow platforms to make data available across networked open science initiatives. Since publishing this dataset to a biodiversity repository, it has become a highly cited open collection used in scientific publications with over 13,000 dataset downloads. This exemplifies the value of adopting FAIR principles of discovery, access, interoperability, and reuse, but requires further steps required to augment the collection and make it more accessible to archaeologists, Indigenous communities and scholars, and educators and students across the Northwest Coast of North America.

## 2.2 The University of Victoria Zooarchaeology Lab Comparative Collection

Zooarchaeological comparative collections (ZCCs), like most scientific collections, now require both physical and digital management to better facilitate research and educational activities. In archaeology, this includes practices that enable zooarchaeological research on human-animal relationships as well as access to biogeographical, ecological, and ethnobiological data. This chapter describes the digitization process to better recognize and share the intrinsic body of knowledge held in a regional ZCC. The UVicZL is a unique facility, and my research uses the comparative collection to explore ways to extend knowledge and broaden its capacity within and beyond the zooarchaeology domain. The lab is a key resource used by many students, faculty, visiting researchers, and consulting archaeologists who recognize the value of the tangible aspects of comparative collections and physical laboratory space. The development of this diverse comparative osteological collection of known modern animal species was supported by the Department of Anthropology and curated by Rebecca (Becky) Wigen, who was a graduate student and later a Senior Lab Instructor between 1980-2017. In addition to building and maintaining the collection, Becky and her colleagues, Susan Crockford and Gay Frederick, have been instrumental in contributing to regional zooarchaeological research and understanding the Indigenous use of animals across the Northwest Coast (NWC) of North America and the Northeast Pacific over the past 50 years.

While there is considerable awareness of the UVicZL collection among archaeologists, including academics, students, and professional consultants, knowledge of the purpose and utility of the collection outside of this domain varies immensely. The lab manager, some UVic faculty and students, and local and regional archaeologists and zooarchaeologists are most familiar with the collection. The lab manager administers specimen acquisition via donation and/or preparation, and loans. Work study students help expand the collection base and accession specimens. Faculty who teach undergraduate zooarchaeology courses or supervise honours and graduate students whose research requires specimen identification also access the lab. Occasionally cross-department collaboration occurs, for example with biology, or requests for access from visiting researchers from other universities. The lab facility includes a specimen preparation room with a dermestid beetle colony, fume hood, microscopes, reference literature, and bench space. Professional consultants regularly book bench space in the lab facility to identify skeletal or invertebrate remains from archaeological sites. One such company, Pacific Identifications Inc., has used the lab collection and carrels regularly over several decades and through biological research contacts, facilitated the acquisition of numerous additional fish, bird, and mammal species. As a type of Natural History Collection (NHC), the UVicZL collection is representative of regional animal species where a majority of remains from NWC archaeological sites have been identified (McKechnie et al. 2020). My research digitizes and broadens the primary biodiversity and ethnobiological information embedded in this extensive collection.

**Creating a FAIR open collection**

Like many small NHCs, comparative collections in zooarchaeology remain under recognized and generally unknown outside the disciplines of anthropology and archaeology. Although open science methods are widespread and increasingly discussed in archaeology, they are not yet common in zooarchaeology. For instance, the ability to locate comparative collections and individual specimens is often lacking online. While many university-based comparative collections exist, encountering information such as the location of the facility, access, and details of collection content is not easily obtained and depend on precise search terms. Moreover, the primary information about specimens in such collections is often inaccessible and/or undervalued and underutilized. In the case of the UVicZL collection (prior to this research), to access this information a researcher would have to inquire about individual specimens and then visit the separate preparation lab and scrutinize the catalogue card for a particular specimen. Fortunately, the card catalogue system was carefully designed by Becky Wigen for the UVicZL collection and maintained for 37 years, providing a granular account of thousands of specimens. However, other than a few columns of spreadsheet data, this precious written archive had not been digitized, scanned, or itemized. My research to bring this information forward is designed to reveal primary data and improve collection awareness and reach, display the richness and depth of collection data, and facilitate connections to users and data sources. Additionally, this research strengthens the transparency of zooarchaeology identifications that have been conducted in the lab over the past 40 years. This is especially relevant given that zooarchaeological evidence is becoming ever more relevant to contemporary environmental management and conservation challenges (Rick 2023). Hence, my approach takes advantage of digital opportunities to connect the knowledge held in collections like UVicZL more broadly.

In this chapter I first ask if open science and FAIR (Findable, Accessible, Interoperable, Reuseable) Principles can contribute to collection accessibility and broaden discoverability. To address this question, I explore biodiversity and natural history archive methods to make data findable and accessible. This question seeks to innovate documentation methods and connect with diverse audiences. I then ask how data sharing and reuse can be enhanced to broaden collection specimen information and use across disciplines. I adopt open science tools that follow data standards, including vocabularies and terminology, linked open data (LOD), and geospatial methods. The UVicZL comparative collection, formally curated between 1980 and 2017 by Rebecca Wigen, is fortunate to have a consistent and well documented specimen record system including handwritten catalogue cards, which I digitized. These previously unscanned cards and records were the only physical copy and form basis for my research to design a digitally accessible open collection. This required research on best practices in open science, FAIR principles, data management plans (DMPs), workflow and storage platforms and online repositories. Also imperative was to understand how specimen data could be aligned with and connected to the considerable range of science conventions, initiatives in taxonomy, and biodiversity research. Accordingly, I conducted background research on which specific tools to bring the UVicZL collection closer to these goals and thus began a journey to connect people and collections and consider the animals in the collection from diverse viewpoints. This research aims to strengthen collaborative efforts for current and future users, addresses the broader terrain of archaeological research, and inspires greater use of the resulting open collection.

## 2.3 Methods

### Workflow processes

To create an open collection, I followed established open access methods and tools which are further described and referenced in an Addendum to this thesis. Specifically, I designed a workflow with six key processes to address my research questions (**Figure 1**). To apply FAIR principles, I created a data management plan, undertook scanning of primary information cards and secured data storage with the support of the Department of Anthropology. I then began amalgamating primary data and selected a metadata standard, which involved deciding on a workflow platform and storage options. After these first four steps, I began annotating specimen records to extend specimen data. As described further below and in chapter 3, these methods were assisted with the use of tools from GitHub and RStudio, publishing tools in data repositories, and developing of a geocoding system to locate specimen collection locations. **Figure 1** visualizes these processes and provides a framework for my methods. To understand the prevalence of ZCC collection information available online, I conducted Google web searches for collections using search terms such as “zooarchaeology/ical”, “comparative”, “collection”, “reference”, “skeletal”, “bone”, “faunal”, and “modern” and consulted experienced colleagues.

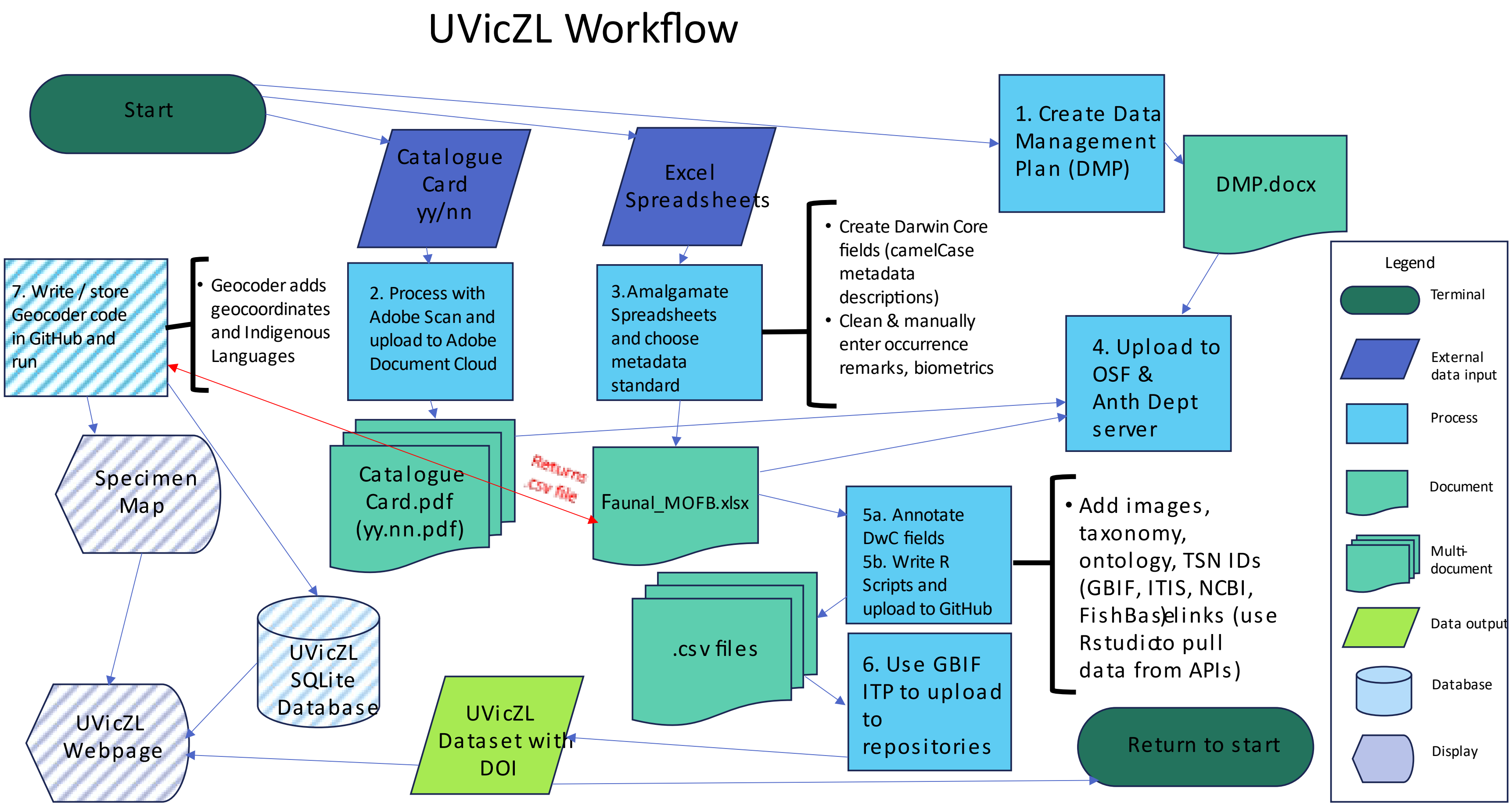



**Figure 1** UVicZL Workflow diagram showing data inputs, seven processes, document creation, data outputs, database, and display results. Processes 1-6 are described in Methods, process 7 is described in chapter 3.

*Process 1: Data management plan (DMP)*

To develop a DMP, I considered best practices for managing research data and zooarchaeology data (Kansa 2015, Kansa et al. 2020) over the data life cycle. The UVicZL project DMP was modified following the Digital Research Alliance of Canada's[4] institution specific Portage's DMP Assistant[5] with participation and guidance following a University of Victoria Library Digital Scholarship Commons workshop.

*Process 2: Scanning and data storage*

The original collection records consisted of physical catalogue cards, a specimen recording log, and excel files separately listing birds, fish, mammals, and other animals. The card catalogue was organized taxonomically by class (using a common name e.g., Birds, Mammals, Fish) and then by family name (e.g., Accipitridae or 'birds of prey'). In 1980, Becky Wigen established the specimen catalogue numbering system using of the last two digits of the year of collection or donation followed by the numerical sequence of animals received in that year. For example, the 21st animal entering the collection in 1997 was given the catalogue number 97/21. The excel files contained some of the catalogue card information, but not all, and sometimes included novel information not on the cards (i.e., measurements, collection location). Neither of these data sources had any accessibility outside of the physical lab.

The initial process of digitizing these records involved scanning each catalogue card as a .pdf document using Adobe Scan on an Android device. These scans were initially stored on a

[4] https://alliancecan.ca/en
[5] https://assistant.portagenetwork.ca/

local laptop and a Cloud server before being uploaded to a UVic Systems supported secure server in folders organized by taxonomic family.

*Process 3: Amalgamate primary data and choose metadata standard*

I gathered the physical and digital data available for the collection with the intention of maximizing accessibility, interoperability, and reuse. Handwritten catalogue cards and a logbook were key physical records that needed to be preserved digitally. Excel worksheets for each collection type (mammal, bird, fish, shellfish) contained data recorded from the catalogue cards and often additional information not otherwise recorded. These primary source worksheets containing attributes of each of the mammal, fish, shellfish, and bird specimens were compiled carefully to ensure that original fields (columns) were in line. Trait attribute data contained in each field were cross referenced with data from catalogue cards. These included scientific and common names, sex and age, measurements, and other relevant information associated with each animal (**Table 1**).

**Table 1** Trait attributes recorded on excel spreadsheets and catalogue cards

| **Original Trait Attribute** |
| --- |
| Year of collection |
| Catalogue number |
| Scientific name |
| Common name |
| Comments |
| Date of death |
| Sex |
| Weight & length |
| Date donated |
| Donor |
| Where collected |
| Original catalogue card images |

I used an excel workbook file, (Faunal_MOFB.xlsx) containing several worksheets, as a 'master workbook' and a basis for data entry and annotation described in Process 5. The primary worksheet 'faunal_collection' was created using standardized field headings (i.e., column headings) developed from the Darwin Core (DwC) metadata vocabulary. Formatting followed the Darwin Core Quick Reference Guide[6], and used an excel template generator[7] to add DwC terms as lowerCamelCase style field headings (columns) to the dataset. The compiled worksheets within Faunal_MOFB.xlsx, were standardized with the selected DwC terms to enable interoperability. This step took considerable time as not all fields (columns) describing original trait attributes seemed to have equivalent DwC terms, nor was the utility of the extensive range of more specific fields immediately apparent. However, using the scanned catalogue cards as a reference, each specimen record could be further described using the DwC terms (Appendix **Table 12**) and required significant time to conduct manual conversion and data entry of catalogue numbers, names, dates, locations, donors/collectors, preparations, occurrence remarks, and biometric data from the corresponding card image. An example of the data transferred by transcription into corresponding fields is shown in **Figure 2**. Hundreds of hours were spent conducting data entry and cleaning manually. This involved line by line transcription, writing excel formulas, and copy and paste functions. Some of this work was performed by a workstudy student and a NSERC USRA student. In addition, the original catalogue card text was recorded verbatim in the field <dwc:occurrenceRemarks>. If the original excel worksheet contained additional information about the specimen in a "comment" field (renamed "occurrenceRemarksOriginal"), this was also included in <dwc:occurrenceRemarks>.

---

[6] https://dwc.tdwg.org/terms/

[7] https://gbif-norway.github.io/dwc-excel-template-generator-js/

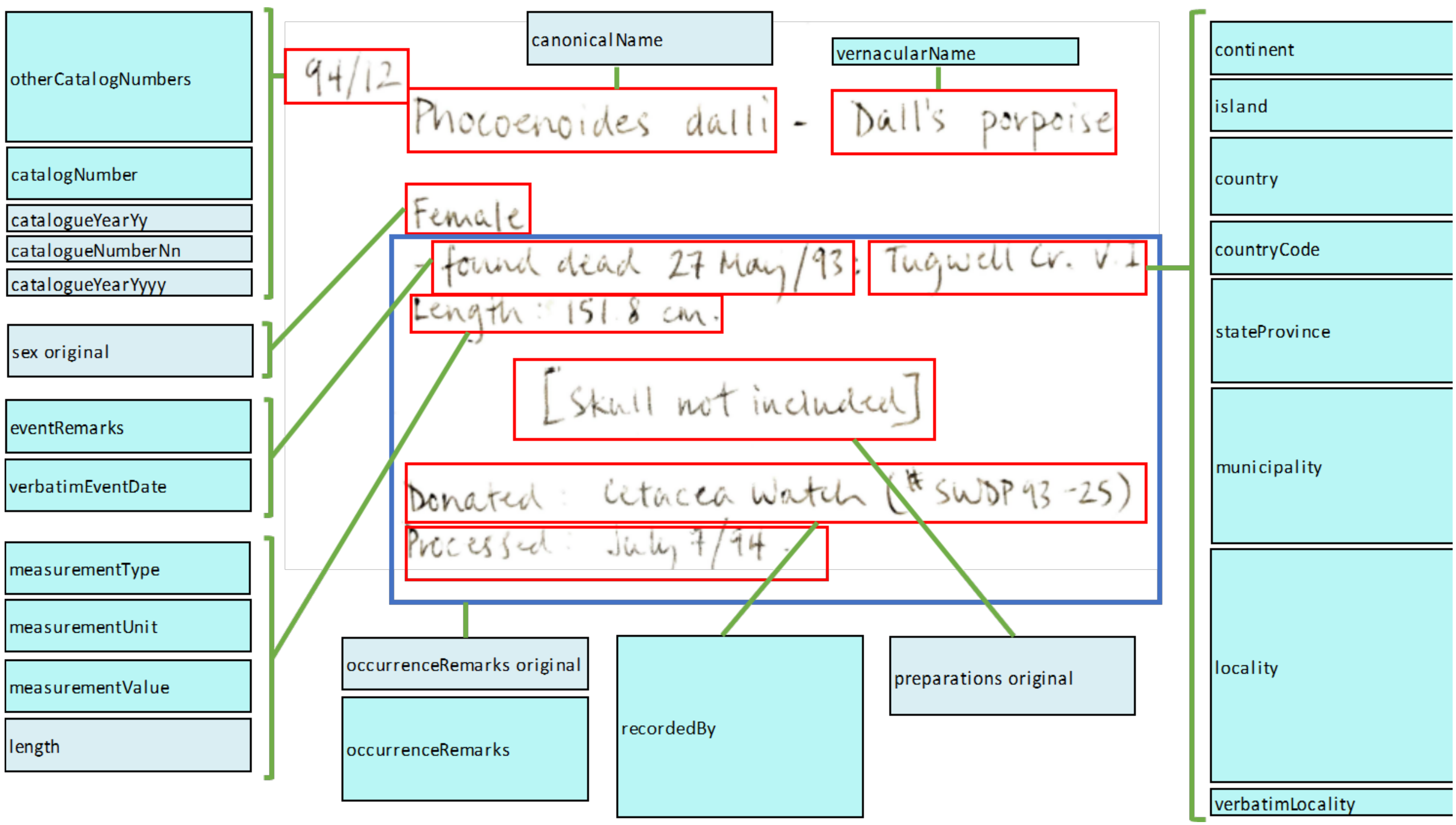



**Figure 2** Fields described and data entry from specimen catalogue card. Darker blue rectangles represent original fields from excel spreadsheets and catalogue cards. Brighter blue rectangles represent standardized Darwin Core terms for field names.

*Process 4: Choose workflow platform and storage*

The DMP, the master workbook Faunal_MOFB.xlsx, and any existing primary lab documents, including original excel worksheets, a laboratory map and listing of physical specimen locations, and a listing of lab literature resources were then uploaded to the Open Science Framework (OSF) UVic Zooarchaeology Lab (UVicZL) Data Management Project[8]. Once all the catalogue cards were scanned, they were also uploaded and the OSF project was made publicly available with its own citable DOI (McKenzie and McKechnie 2024). In addition to OSF, all project documents are updated with copies stored on the secure Anthropology Department "uviczooarchlab" server which is regularly backed up.

*Process 5a: Annotate specimen records*

Taxonomic names, classifications, and identifiers, links to element, age and sex ontologies, and card images added to each specimen record were based on the original excel worksheet and catalogue card records. Initially, each record was transcribed line by line. Taxonomic names were guided by the Global Biodiversity Information Facility (GBIF) backbone taxonomy[9]. Approximately one third of the records were completely transcribed and annotated manually. All specimen records required manual transcription into <dwc:occurrenceRemarks>, from the verbatim comments on each catalogue card

<u>Taxonomies</u>

To ensure standardized species names, specimen records were annotated with links and taxonomic serial numbers (TSNs) using four international open science initiatives: International Taxonomic Information System (ITIS), GBIF, National Center for Biotechnology Information

[8] https://osf.io/45bw9/
[9] https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c

(NCBI), and FishBase. Taxonomic serial numbers (TSNs) from these taxonomic authorities and databases are linked to each specimen by Uniform Resource Indicators (URIs).

Ontologies

Ontologies are terms used to describe categories of things. In this thesis, I use accepted ontologies described by specialized repositories as explained below to link to specimens in the collection to a broadly understood set of categorical descriptions. I include the names for skeletal elements in the field <dwc:preparations>. I adopt the UBERON skeletal anatomy to standardized terms for skeletal elements which have been vetted and stored in the broader Ontology Lookup Service (OLS) repository. For animal ages, I follow terms defined by the Digital Archaeological Record (tDAR) repository who compiled a faunal age ontology specific to common zooarchaeological data. Specimens with age estimates are linked to these terms in the field <dwc:lifestage>. Similarly, for animals with sex information, I used the tDAR faunal sex ontology to indicate sex information in the field <dwc:sex>.

Digital images and media

The majority of specimen records received a stable GUID from OSF that links the image of the original catalogue card to the dataset. I provided links to additional open-source digital images have been added including a few collection specimens that have been 3D scanned by the UVic Library. Those links are available from the Idaho Virtual Museum[10], the UVic Library's SketchFab page[11], and specifically the UVic Zooarchaeology Lab collection[12].

[10] https://virtual.imnh.iri.isu.edu/Osteo
[11] https://sketchfab.com/uviclib
[12] https://skfb.ly/oWD6P

*Process 5b: GitHub and RStudio*

Hours of manual entry time was saved using GitHub and RStudio to develop code-based queries to populate the dataset with hierarchical taxonomy terms, taxonomic serial numbers, and catalogue card images for each specimen record. In collaboration with the Geospatial and IT team at the Hakai Institute, as part of a MITACs accelerate internship, we devised script queries in the programming language R to siphon taxon and image data from APIs including GBIF, Wikidata, and OSF. These APIs enabled taxonomic data to be harvested from multiple open science platforms (GBIF, ITIS, NCBI, and FishBase), and card image data from OSF. The scripts written in RStudio were pushed to a private Hakai Institute GitHub repository. This repository will be migrated to a UVic repository and made public once coding from Process 7 is complete (see Chapter 3). See Appendix **Figure 24** for an example of the code written to add the ITIS TSNs to each specimen record.

*Process 6: Data publishing - Canadensys and GBIF*

The UVicZL excel workbook file (Faunal_MOFB.xlsx) worksheets, 'faunal_collection', 'faunal_collection_MoF', and 'faunal_collection_simple_multim' were converted in excel to CSV files and uploaded as DwC archive records to the Canadensys ITP. These records include a DwC Occurrence file, and two Extension files: a Measurement or Facts (MoF) file, and a Simple Multimedia file. The occurrence file contains 105 fields, 73 of which are mapped to DwC terms (Appendix **Table 12**). The extended MoF file contains biometric data that can include specimen weight, length, and/or diameter as derived from the occurrence file. For example, depending on the species, fish lengths can have one of several terms to quantify body size: standard, total, and fork length measurements. The Simple Multimedia file contains GUIDs for the catalogue cards and links to specimens with additional imaging such as 3D scans.

*Process 7: Create geocoder for geocoordinates and Indigenous language area map*

In Chapter 3, I provide more detail on the creation of a geolocator workbench that uses code to generate locations and display specimen information in an interactive and queryable map. The current version of the map includes Indigenous language areas obtained from the open data initiative Native Land Digital which also uses open source geocoordinates to show polygon locations where Indigenous language are spoken. As described in Chapter 3, this is particularly relevant to specimen collection locations.

## 2.4 Results

### FAIR open collections

*Finding zooarchaeological comparative collections*

To understand the extent of accessible collections, a web search for ZCCs using search terms such as “zooarchaeology/ical”, “comparative”, “collection”, “reference”, “skeletal”, “bone”, “faunal”, and “modern” produced information about 31 institutionally held zooarchaeological comparative collections (**Table 2**). Notably, fewer than half of these institutions provided links to access or download collection information or provide a total count of comparative specimens. Knochenarbeit[13], an international directory of archaeozoological resources was developed by biologist, Hans Christian Küchelmann, and posted to the International Council of Archaeozoology (ICAZ) in 2020. This directory resource lists 168 Archaeozoological Institutions[14], that conduct zooarchaeological research, 29 of which are in North America (6 in Canada, 23 in US). The directory also lists comparative Skeletal Collections and Identification

[13] https://www.knochenarbeit.de/?lang=en
[14] https://www.knochenarbeit.de/archaeozoological-institutions/?lang=en

Keys[15]. Of the 68 listings less than 30% are in North America with only two Canadian sources[16] listed.

[15] https://www.knochenarbeit.de/skeletal-collections-identification-keys/?lang=en
[16] http://www.royalbcmuseum.bc.ca/Natural_History/Bones/homepage.htm and http://vzap.iri.isu.edu/

**Table 2** Assemblage search results of 31 online zooarchaeological comparative collection. Orange section indicates United States/North America/Regional collections. Green section indicates NWC/Northern North America. Blue section indicates European/British/Continental/Middle East. Yellow section indicates Southeast Asia/Pacific. *Numbers are approximate.

| Institution / Lab | Name of Collection(s) | Primary region of coverage | Webpage | Access or download format | Fields described | Number of specimens |
|---|---|---|---|---|---|---|
| **U of Arizona - Arizona State Museum** | Western Archaeological & Conservation Center Collection; Stanley J. Olsen Zooarchaeological Collection | SW/SE United States | No | N/A | N/A | 4,000* |
| **U of California, Davis – Department of Anthropology** | Zooarchaeology Lab Comparative Collection | North America & Arctic | A | Listings by animal type | 7-8 | 1,800* |
| **U of Florida – Florida Museum Environmental Archaeology Program** | Zooarchaeology Comparative Collection | SE North America & circum-Caribbean | B | Searchable database | 14 | 11,784 |
| **Colorado State University – Anthropology and Geography** | Zooarchaeology Lab Comparative Collection | North America, Africa | No | N/A | N/A | N/A |
| **U of Nevada, Las Vegas** | Modern Comparative Collection | SW United States SW Asia & eastern Mediterranean | No | N/A | N/A | N/A |
| **UCLA Cotsen Institute of Archaeology** | Zooarchaeology Lab Modern Comparative Collection | Regional, North America, Central America, Peru | C | N/A | N/A | 5,100* |
| **New York University – Center for the Study of Human Origins** | Comparative Zooarchaeology Collection | North America & Europe | D | Spreadsheet | 4 | 300* |
| **Indiana University William R. Adams Zooarchaeology Laboratory** | Modern Comparative Collection | North America | E | N/A | N/A | 10,000* |
| **Mississippi State University** | Comparative Faunal Collection | SE United States | No | N/A | N/A | 450* |
| **Illinois State Museum** | Zooarchaeology Laboratory Comparative Collection | Midwestern United States | No | N/A | N/A | N/A |
| **U of Georgia – Department of Anthropology** | Zooarchaeology Lab Comparative Collection | SE United States, Caribbean | F | N/A | N/A | 5,000* |
| **U of Maine – Climate Change Institute** | Zooarchaeology Lab Comparative Collection | NE and SW United States | G | N/A | N/A | N/A |
| **Wake Forest U Archaeology Lab** | Comparative Faunal Collection | North Carolina | No | N/A | N/A | 200* |
| **Harvard Peabody Museum** | Zooarchaeology Laboratory Reference Collection | New England | H | N/A | N/A | 1221 |

[A] https://zooarchaeology.ucdavis.edu/
[B] https://www.floridamuseum.ufl.edu/scripts/dbs/zooarch_comp_pub.asp
[C] https://www.ioa.ucla.edu/content/zooarchaeology-lab
[D] https://s18798.pcdn.co/csho/wp-content/uploads/sites/2067/2017/05/CrabtreeFaunalSpecimens_Updated5December2016.xlsx
[E] https://zooarch.sitehost.iu.edu/home.php
[F] https://anthropology.uga.edu/research/zooarchaeology-lab
[G] https://www2.umaine.edu/climatechange/Research/facilities/zoo/index.html
[H] https://peabody.harvard.edu/zooarchaeology-lab

| Institution / Lab | Name of Collection(s) | Primary region of coverage | Webpage | Access or download format | Fields described | Number of specimens |
|---|---|---|---|---|---|---|
| **U of Toronto Mississauga – Department of Anthropology** | Zooarchaeology Lab Deborah J. Berg Faunal Collection | Ontario | I | .pdf listing | 7 | 1,500* |
| **U of Alberta – U of Alberta Museums** | Zooarchaeology Reference Collection | Alberta & the Arctic | J | Searchable database | Up to 13 | 903 |
| **Portland State University (PSU)** | Zooarchaeology Reference Collection | NWC, Interior Plateau | K | Spreadsheets, HTML Charts | 15 | 379 |
| **The Virtual Zooarchaeology of the Arctic Project (VZAP)** | Virtual, interactive, osteological reference collection | Arctic & North Pacific | L* | Searchable database | N/A | N/A |
| **Simon Fraser University** | Zooarchaeology reference materials | N/A | No | N/A | N/A | N/A |
| **Central Washington University Zooarchaeology Laboratory** | Comparative Osteology Collection | Western North America, North Pacific | M | .pdf listing | 5 | 600* |
| **U of Oregon Zooarchaeology Lab** | Comparative Faunal Collection | North Pacific | N | N/A | N/A | N/A |
| **U of Sheffield Zooarchaeology Laboratory** | Reference Collection | European | O | Spreadsheet | 43 | 2,096 |
| **Historic England** | Zooarchaeology Reference Collection | Britain & Continental Europe | P | .pdf listing | 11 | 3,308 |
| **UCL Institute of Archaeology** | Osteological Comparative Collection | Regional? | No | N/A | N/A | <5,000 |
| **The Institute for Aegean Prehistory Study Center for East Crete (INSTAP SCEC)** | Modern Cretan Faunal Skeletal Collection | N/A | No | N/A | N/A | N/A |
| **American School of Classical Studies at Athens** | Modern Faunal Comparative Collection | Greece & Eastern Mediterranean | No | N/A | N/A | N/A |
| **STARC, The Cyprus Institute** | Faunal Reference Collection | Regional | Q | Spreadsheet | 45 | 202 |
| **U of Haifa, at the Zinman Institute of Archaeology** | Modern Comparative Reference Collection | Regional? | No | N/A | N/A | N/A |
| **UCL - Zooarchaeological Research Laboratory** | Osteological Comparative Collection | Middle East & Eurasia | No | N/A | N/A | N/A |
| **U of Leicester** | Reference Collection | Britain & Europe | R | Spreadsheet | 24 | 641 |
| **Australian National University** | Zooarchaeology Collection | Southeast Asian & Polynesia | S | 57 3D models | N/A | N/A |

[I] https://www.utm.utoronto.ca/anthropology/research/ontario-archaeology-utm/zooarchaeology/utm-zooarchaeology-lab-and-deborah-j-berg-faunal

[J] https://search.museums.ualberta.ca/search?initial_submit=true&initial_collections[0]=20&type_view=items&items_view=list&item_groups_view=list

[K] https://web.pdx.edu/~virginia/zooarchcatalog.html

[L] https://boneswall.iri.isu.edu/ *Collaboration between Idaho State University, Canadian Museum of Civilization, Idaho Museum of Natural History, Port Townsend Marine Science Center, Sitka Sound Science Center & Burke Museum of Natural History and Culture (University of Washington)

[M] https://www.cwu.edu/academics/anthropology/facilities/zooarchaeology-laboratory.php

[N] https://pages.uoregon.edu/mmoss/Zooarchaeology-at-Oregon/

[O] https://docs.google.com/spreadsheets/d/1ehX5GLg9C2t0ZoFkS-DdhMYDaBMXtra7BkuDqr-nU3E/edit?usp=sharing

[P] https://historicengland.org.uk/images-books/publications/zooarchaeology-reference-collection/zooarchaeology-ref-collection-jul20/

[Q] https://sachrofics.cyi.ac.cy/reference-collection

[R] https://le.ac.uk/-/media/uol/docs/academic-departments/archaeology/bone-laboratory/animal-bone-holdings.xlsx

[S] https://sketchfab.com/CDHR_ANU/collections/skullbook-digital-bone-library

*Data management and FAIR data*

The Data Management Plan (Appendix **Figure 23**) lists types of data that have been collected and suggests ways to create, link, acquire and/or record this information. This includes file formats that enable data re-use, interoperability, sharing and long-term preservation and access. Project management tools that dictate data structure, filename, and version-control ensure proper and consistent conventions and procedures. The plan estimates data storage space requirements, length of storage, and outlines permissions to ensure data is securely managed and changes are approved by members of the project. Also specified are data storage and back up methods and names data research team members and other project collaborators who can modify and contribute to the project. For preservation, the plan specifies data repositories considered for long-term storage and outlines preservation friendly, proprietary, and non-proprietary file formats to ensure file integrity, and inclusion of supporting documentation. To manage legal, ethical, and intellectual property issues end-user licenses were specified as CC4.0. Also included in the is the intent to improve access to researcher, educators, students, and the community, and strategies to address secondary uses of data. Major data management tasks are listed in the plan for future managers of collection data for which they will be responsible. Finally, resources required to implement the DMP are listed.

According to the FAIR criteria as summarized in **Table 3**, the dataset meets expectations as the components are publicly accessible and available online without requiring specialized protocols or tools. The CC4.0 license means the data are accessible by anyone and can be shared and reused by researchers and educators. Using the Darwin Core vocabulary, data can be viewed, sorted, and filtered widely and easily following established terminologies. This includes the versions of data published in several locations, including GBIF, GRSciColl, and OSF, each which are downloadable in structured, open-standard, machine-readable format (CSV), non-machine-readable format (PDF), and/or proprietary format (XLSX). In OSF the project was registered and received a DOI[36]. Each document, including catalogue card images also receive a unique identifier from OSF (**Figure 3**). The dataset, when published in Canadensys[37], is simultaneously uploaded and registered with GBIF, and is assigned Persistent Identifiers (PIDs) including a GBIF UUID[38] and a collection specific DOI[39]. The publication of the dataset also registers collection, the digitized specimens, and The University of Victoria as a data publisher with the Global Registry of Scientific Collections (GRSciColl) and includes an administrative contact and point of contact[40]. The UVicZL collection is available as an accessible and searchable dashboard via GRSciColl[41] which makes information about the collection readily discoverable for researchers, both local and international, to find information about specimens in the collection. While not every link in the UVicZL dataset is persistent[42], many are. Each specimen record, in addition to trait and collection data, has linked data on taxonomy, ontology, digital images, and geospatial location.

---

[36] https://doi.org/10.17605/OSF.IO/45BW9
[37] https://data.canadensys.net/ipt/resource?r=faunal_collection
[38] https://www.gbif.org/dataset/5b11525e-c116-48a0-956f-4147ecd97237
[39] https://doi.org/10.5886/jej09d
[40] https://www.gbif.org/publisher/e6743ca5-c776-4fdd-9afa-cfbf8e79d86c
[41] https://scientific-collections.gbif.org/collection/1cbebd16-1e98-417a-b1fe-531208312892/dashboard
[42] see Compendium 6.2 Open access methods and tools: Linked data for more on persistent identifiers

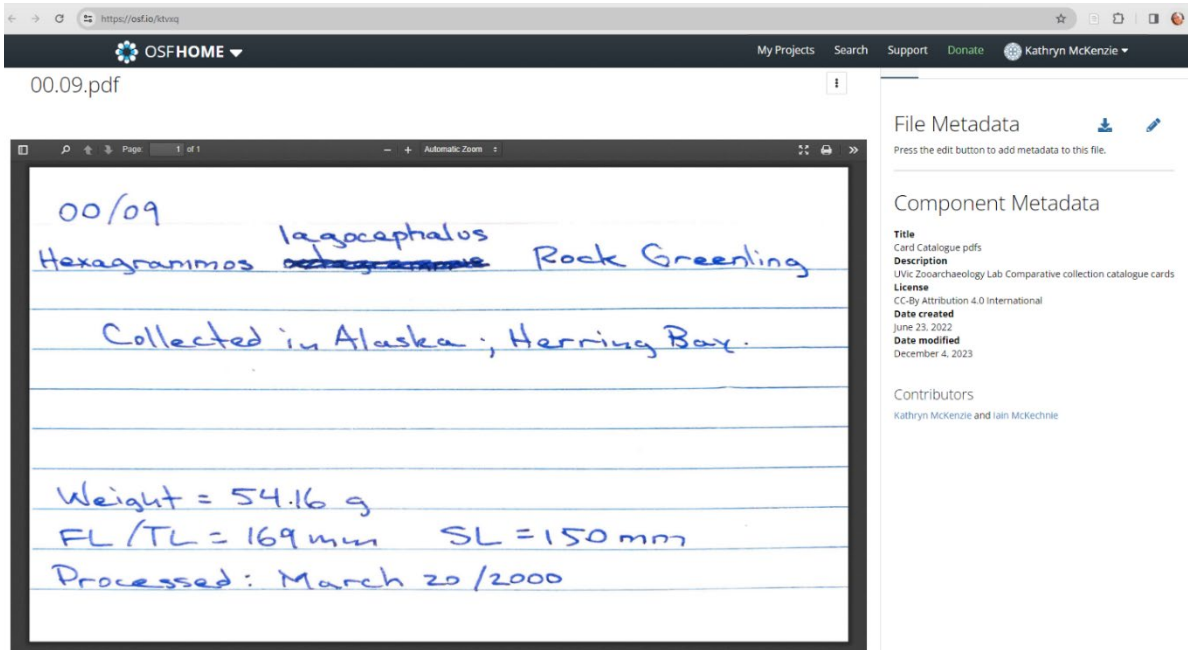



**Figure 3** OSF catalogue card UVicZL-00.09 with GUID https://osf.io/ktvxq

**Table 3** How UVicZL Project Data meets FAIR Principles

| ***FAIR Principle Question*** | ***UVicZL Project Data*** |
|---|---|
| ***Findable*** | |
| *Does the dataset have any identifiers assigned?* | Dataset is assigned a globally unique citable and persistent identifier (DOI) by GBIF. OSF Project has a DOI and GUID and each document in OSF has a GUID |
| *Is the dataset identifier included in all metadata records/files describing the data?* | Yes |
| *How is the data described with metadata?* | Metadata is comprehensive using a recognised formal machine-readable metadata schema |
| *What type of repository or registry is the metadata record in?* | Metadata is stored locally, in OSF and in GBIF and Canadensys repositories |
| ***Accessible*** | |
| *How accessible is the data?* | Publicly accessible |
| *Is the data available online without requiring specialised protocols or tools once access has been approved?* | Standard web service API, downloadable as CSV |
| *Will the metadata record be available even if the data is no longer available?* | Yes |
| ***Interoperable*** | |
| *What (file) format(s) is the data available in?* | Dataset is in a structured, open standard, machine-readable format (.csv); some documents and catalogue cards are structured, open standard, non-machine-readable format (.pdf, .txt); some documents are in proprietary formats (Word, Excel) |
| *What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?* | Standardised open and universal schema (Darwin Core) using resolvable global identifiers linking to explanations |
| *How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?* | The metadata record includes URI links to related metadata, data, and definitions |
| ***Reusable*** | |
| *Which of the following best describes the license/usage rights attached to the data?* | Standard machine-readable license (Creative Commons 4.0 License) |
| *How much provenance information has been captured to facilitate data reuse?* | Fully recorded in a text format |

* Modified from Australian Research Data Commons (ARDC) FAIR Data Self Assessment Tool([https://ardc.edu.au/resource/fair-data-self-assessment-tool/](https://ardc.edu.au/resource/fair-data-self-assessment-tool/))

*Repository publishing*

Global Biodiversity Information Facility (GBIF)

Since it was first uploaded in July 2021, the UVicZL collection hosted on GBIF (McKenzie and McKechnie 2024), global users have downloaded over 5.6 million specimen occurrence records spanning 13,500 download events by (**Figure 4**). The unique open science tools enabled by GBIF allows researchers to search, filter, and download their results by species and geographic occurrence data across numerous datasets published in the repository. The GBIF-mediated data can vary and range across taxonomic hierarchies, record type (preserved specimen for example), date ranges, continents, countries, and IUCN status categories. When a dataset or a subset of its occurrence records from GBIF are downloaded by a user as part of a search, a DOI is generated by GBIF to represent that query and specifies which datasets and the number of occurrence records from each dataset were downloaded. These DOIs can be searched to see which datasets and the number of occurrence records are included in each query. The generated DOI or the dataset DOI itself can be cited in a publication in a number of ways[43]. For the UVicZL collection, this is either "GBIF cited" or "GBIF used", where "GBIF cited" is what GBIF refers to as a "simple citation" (i.e. researchers used this dataset or a subset of its occurrence records in their research) and "GBIF used" is where the data was used more "substantially". As of July 2024, the UVicZL collection dataset has been used and cited 53 times (**Figure 5**). The UVicZL dataset has been "GBIF cited" in four peer reviewed, open access journal articles and one GBIF dataset ID listing. More substantial use of UVicZL occurrence data, i.e., "GBIF used", includes twenty-seven journal articles, twenty-five of which are peer reviewed, and sixteen are open

[43] https://www.gbif.org/es/news/1pHRMMpGbCueqAmmUGAekW/top-tip-literature-data-use-and-how-its-all-linked

access across a wide range of scientific publications, primarily relating to research in biodiversity, biogeography, climate change, and taxonomy. The remaining GBIF-mediated UVicZL occurrence data has been used in ten preprints (nine open access), four theses, four peer reviewed book sections, two IUCN Red List reports, and one computer program (**Figure 6**). One IUCN Red List reports on the Mountain Pygmy-owl (*Glaucidium gnoma*) of which there are three specimens in the collection. Journal venues span the biological, conservation, agri/aquaculture, forestry, and ecological sciences.

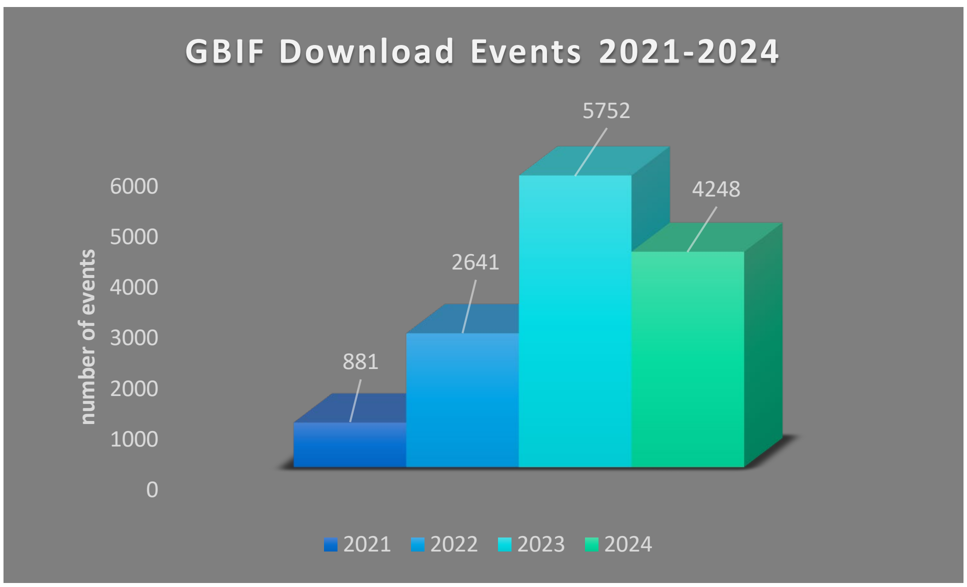



**Figure 4** UVicZL dataset download events by GBIF users since 2021 (13,522) as of July 2024.

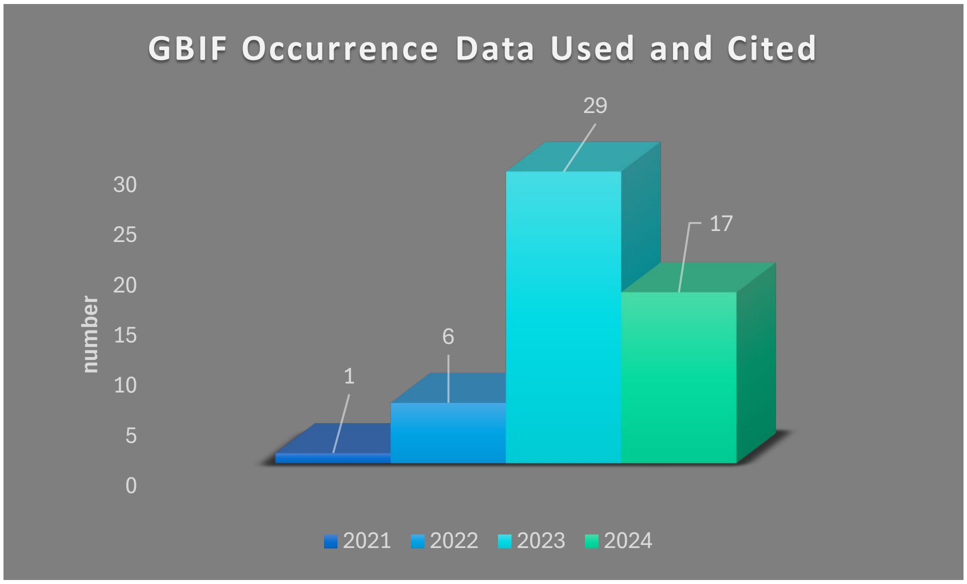



**Figure 5** UVicZL occurrence data uses and citations in GBIF since 2021 (53) as of July 2024.

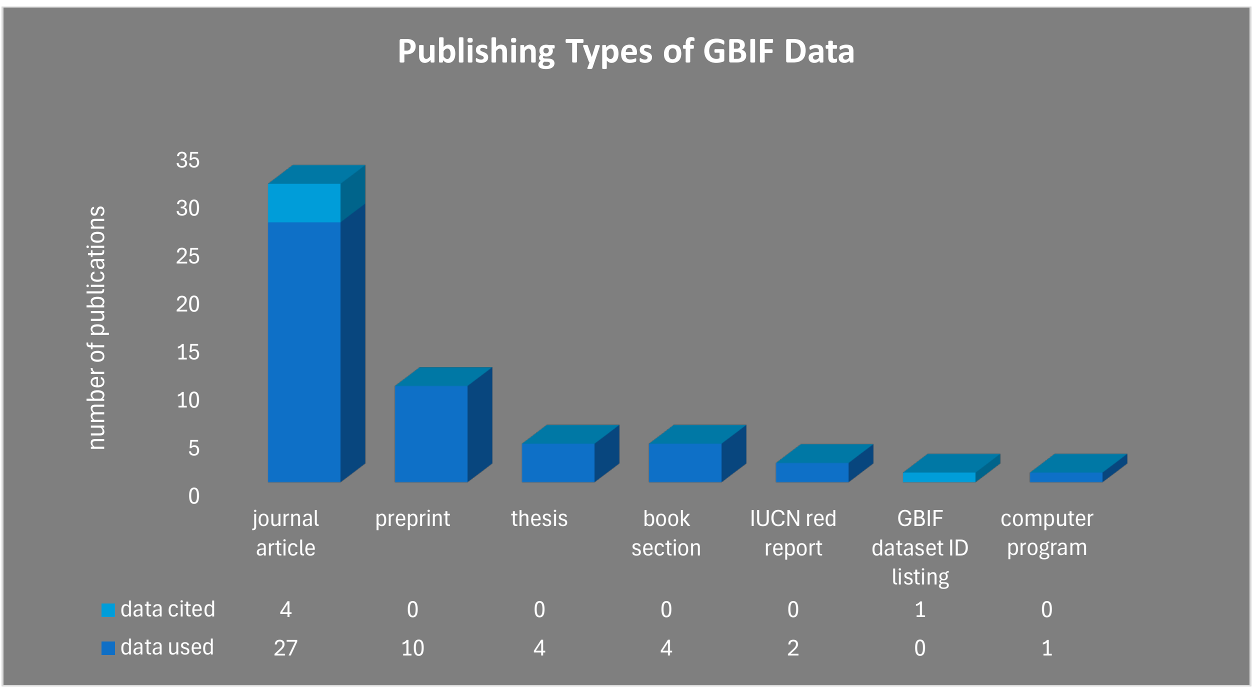



**Figure 6** UVicZL occurrence data has 53 literature citations as of July 2024, including 5 simple citations and 48 data use citations. The dataset has been cited in 4 peer reviewed open access journal articles and 1 GBIF dataset ID listing. GBIF-mediated data has been substantially used in 27 (25 peer reviewed, 16 open access) journal articles, 10 (9 open access) preprints, 4 theses, 4 peer reviewed book sections, 2 IUCN Red List reports, and 1 computer program.[44]

[44] https://www.gbif.org/resource/search?contentType=literature&relevance=GBIF_USED&relevance=GBIF_CITED&gbifDatasetKey=5b11525e-c116-48a0-956f-4147ecd97237

*Geospatial data*

Geospatial data for each specimen in UVicZL, where available, is published in GBIF and is visible in the map-based feature which displays locations based on geographic co-ordinates (**Figure 7**). Of the 2,922 specimens in the dataset, 2,519 (86%) have geographic co-ordinates. Dataset specimens can be queried and filtered using a variety of terms (i.e., DwC location terms such as <dwc:continent>, <dwc:waterBody>, <dwc:islandGroup>, <dwc:island>, <dwc:country>, <dwc:stateProvince>, <dwc:municipality>, <dwc:locality>) and by user drawn polygons. Of the 2,519 specimens, 1,891 specimens were collected in Canada, 550 in the United States, 43 in International Waters, and 17 in Russia (Russian waters) (**Figure 8**).

**Figure 7** Geospatial data points generated by GBIF of UVicZL specimen collection location coordinates.

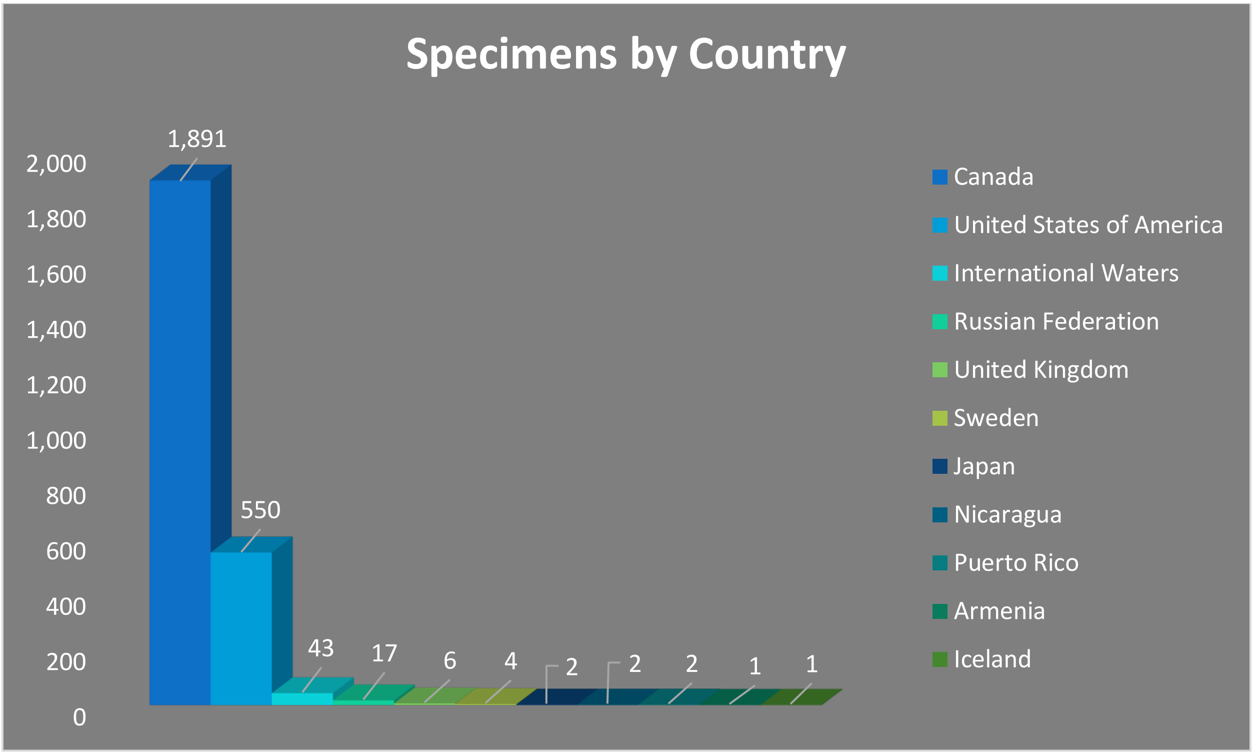



**Figure 8** Number of specimens by country. Of the 2,519 specimens with geospatial data points, Canada has the most numerous with 1,891 specimens followed by USA with 550.

**Extended specimens and linked data**

*Taxonomic authorities*

Specimen records were annotated with links and taxonomic serial numbers from ITIS (n=2,834), GBIF (n=2,916), NCBI (n=2,719), and FishBase (n=1,152). Taxonomy was determined using the GBIF backbone[45]. Depending on taxonomy recognition differences between authorities, some specimens do not have numbers from every provider, and only fish received a FishBase TSN, reflecting this focused initiative.

[45] https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c

**Table 4** Specimens with taxonomic serial numbers (TSN) by provider

| Number of Specimens with Taxonomic Serial Numbers (TSN) | |
|---|---|
| **GBIF** | 2916 |
| **ITIS** | 2834 |
| **NCBI** | 2719 |
| **FishBase** | 1152 |

*Animal type by taxa*

Prior to this research and digitization effort, specimen specific information about the overall composition of the collection was not thoroughly itemized or quantified, particularly given the numerous changes in biological systematics and species taxonomic classifications over the past 40 years. With this now updated, specimens in the collection can be searched and filtered across a much wider variety of fields. More importantly, this revised dataset provides revised counts of the absolute numbers of species and their various taxonomic groupings. The collection includes 2,922 individual animals, including 1,258 fish, 872 birds, 369 mammals, 387 invertebrates and 35 reptiles and amphibians. Of the 2,922 specimens in the dataset, the vast majority (2,887, 99%) are identified to species with the remaining 34 identified to genera. There are 671 distinct species including 281 fish species, 214 bird species, 72 mammal species, and 104 other species including invertebrates, reptiles, and amphibians (**Figure 9**). Figure 10, Figure 11, and Figure 12 show the top ten taxa by order, family, and species, respectively. Fish are well represented by order with 1,012 specimens including Scorpaeniformes (385), Perciformes (203), Salmoniformes (187), Pleuronectiformes (141), and Gadiformes (96) (**Figure 10**). The fish order Scorpaeniformes include 716 specimens from Cottidae (138), Hexagammidae (129), Sebastidae (60), and 58 specimens from other families (**Figure 10**). Birds are also well represented by order with 562 specimens including Passeriformes (258), Anseriformes (172), and Charadriiformes (132) (**Figure 10**). Of the Anseriformes, represented by family, are 172 Anatidae (**Figure 11**). Mammal specimens include 250 specimens from the orders of Carnivora (170) and Rodentia (80) (**Figure 10**). The Carnivora order is further refined by family to include 74 Canidae and 46 Mustelidae specimens (**Figure 11**). At the species level, fish specimens are again well represented with 96 *Pleurogrammus monopterygius*, 143 *Oncorhynchus sp.* (O. *mykiss, O. tshawytscha, O. clarkii, O. keta, O. kisutch*), 26 *Gadus chalcogrammus*, and 24 *Clupea pallasii* (**Figure 12**). Additionally, 63 *Canis lupus* and 23 *Anas platyrhynchos* specimens represent mammals and birds to round out the top ten species (**Figure 12**).

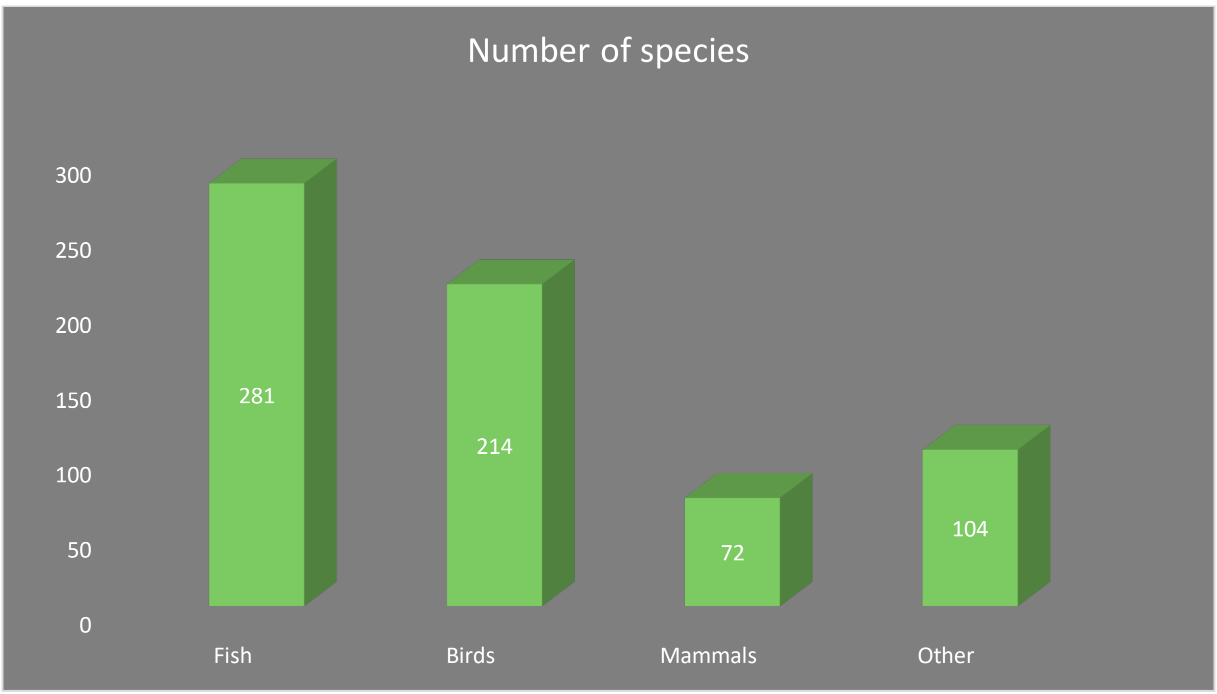



**Figure 9** Number of species in the UVicZL collection (n=671) grouped by broad common taxonomic class names.

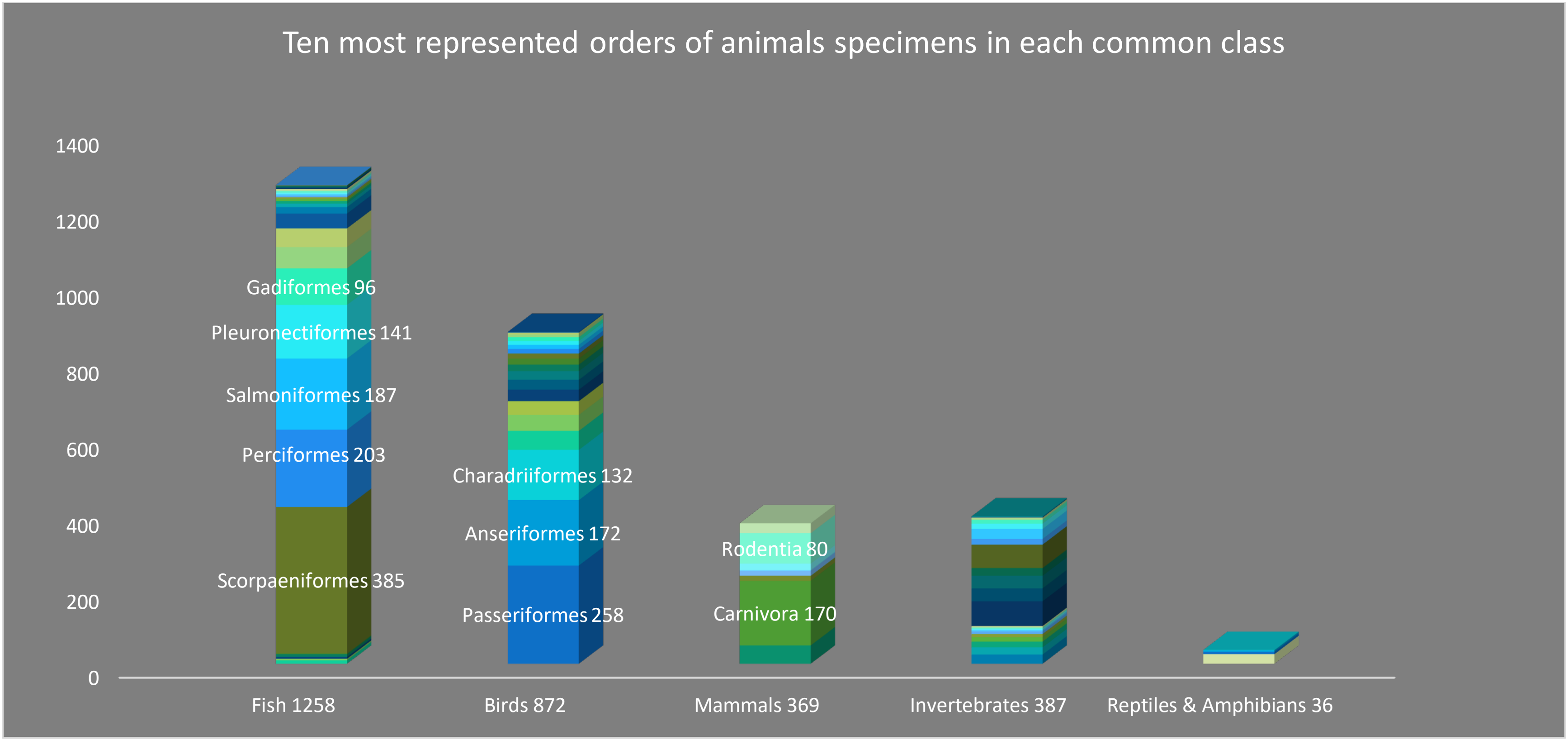



**Figure 10** The top ten orders of the 2,922 animal specimens categorized by broad common taxonomic class names. These include 1,258 fish, 872 birds, 369 mammals, 387 invertebrates, and 36 reptiles and amphibians. 1,012 of the 1,258 (80%) fish belong to one class, Actinopterygii, and five orders (Scorpaeniformes, Perciformes, Salmoniformes, Pleronectiformes, and Gadiformes). 562 of the 872 (64%) birds belong to one class, Aves, and three orders (Passeriformes, Anseriformes and Charadriiformes). 250 of the 369 (68%) mammals belong to one class, Mammalia, and two orders (Carnivora and Rodentia).

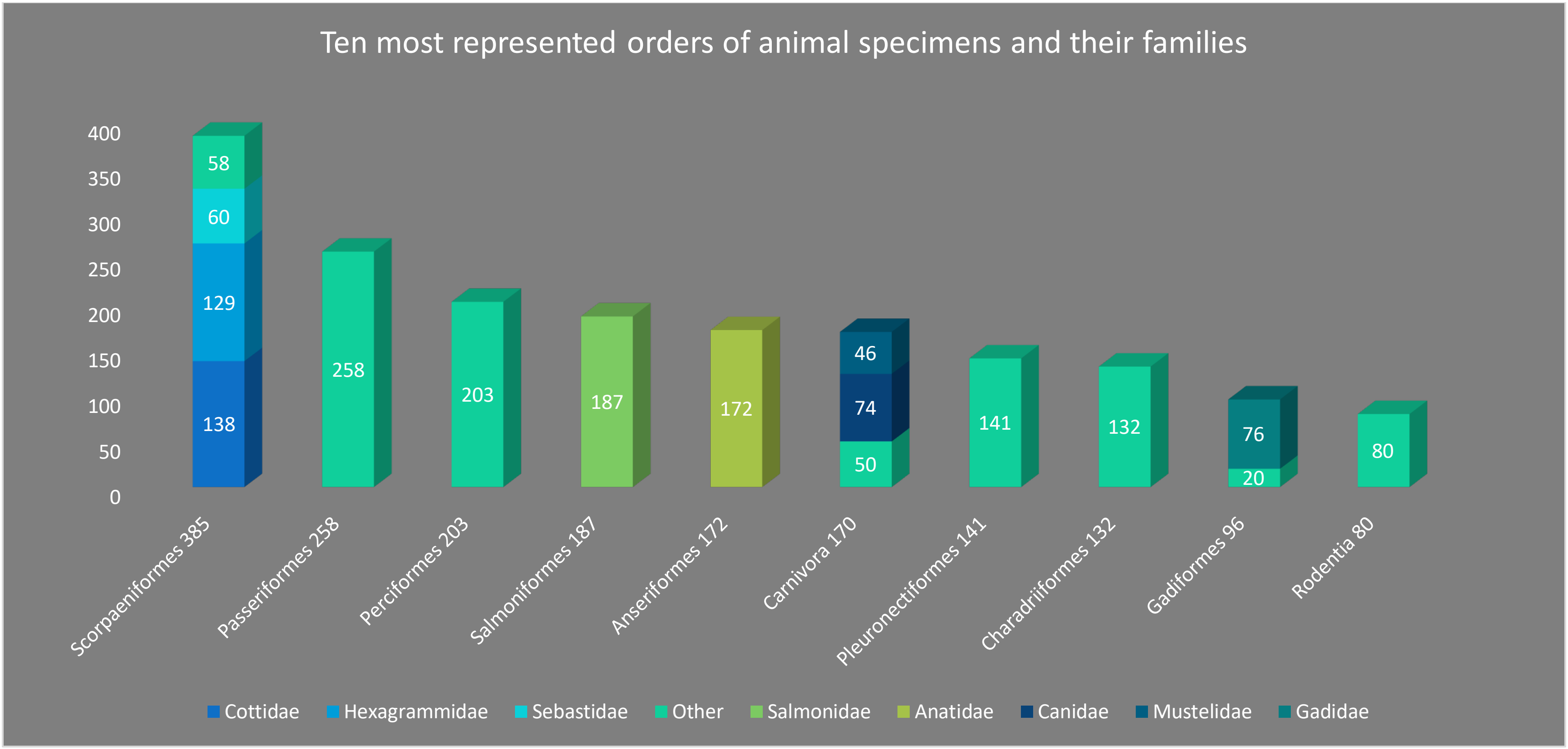



**Figure 11** Top ten animal specimens by taxonomic order and family. There are 385 specimens from the order Scorpaeniformes made up of 138 Cottidae, 129 Hexagammidae, 60 Sebastidae, and 58 from other families. There are 258 specimens from the order Passeriformes and 203 specimens from the order Perciformes. There are 187 specimens from the order and family Salmoniformes Salmonidae and 172 specimens from the order and family Anseriformes Anatidae. Of the 170 Carnivora specimens 74 are from the Canidae family, 46 from the Mustelidae family and 50 from other families. There are 141 specimens from the order Pleuronectiormes and 132 specimens from the order Charadriiformes. Of the 96 Gadiformes specimens 76 are from the Gadidae family, and 20 from other families. There are 80 specimens from the Rodentia order.

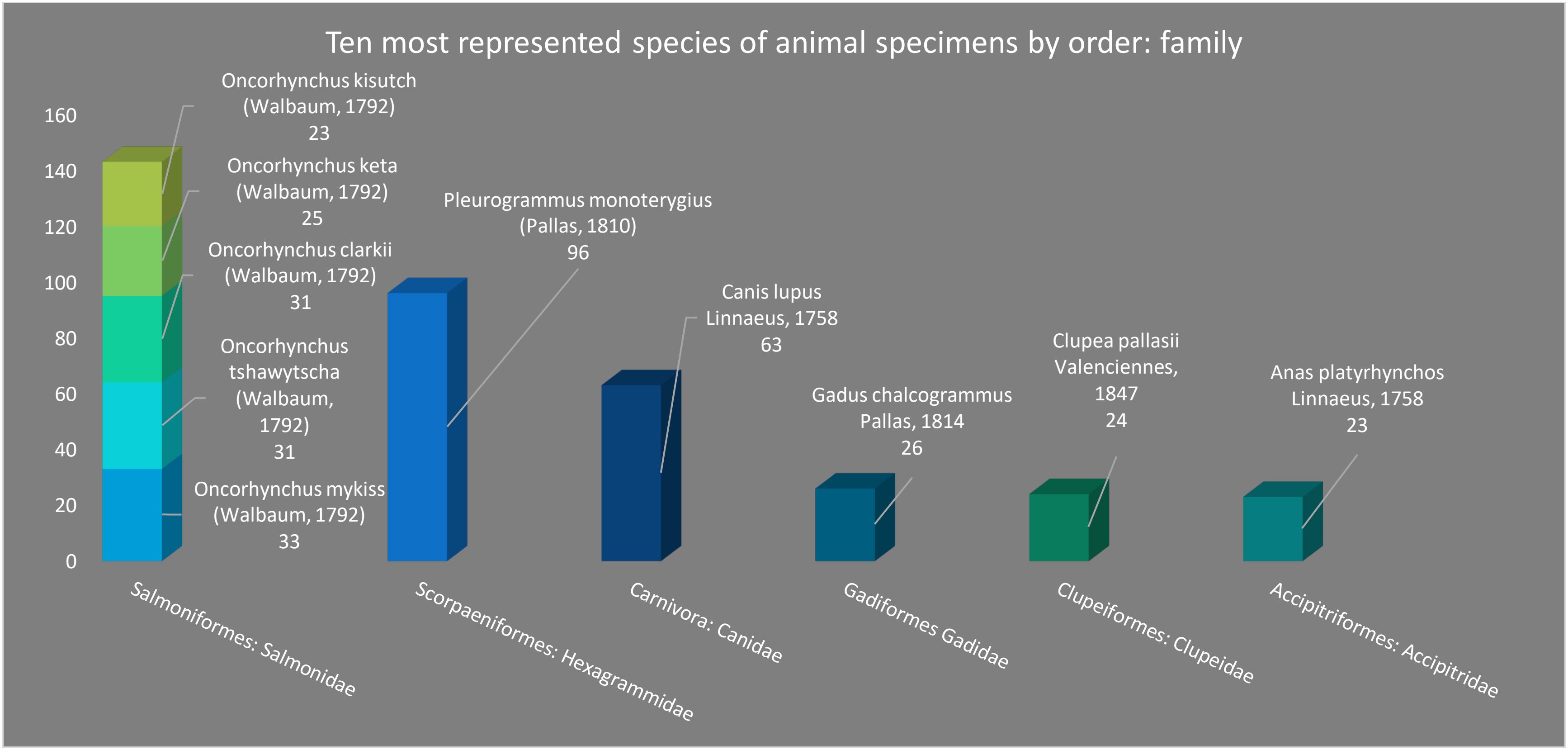



**Figure 12** The top ten species by order and family. Salmoniformes Salmonidae contain 143 specimens represented by five salmon species *O.mykiss* (33), *O. tshawytschia* (31), *O. clarkii* (31), *O. keta* (25), and *O. kisutch* (23).There are 96 specimens of *P. monoterygius* from the order and family Scorpaeniformes Hexagrammidae . There are 63 specimens of *C. lupus* from the order and family Carnivora Canidae. There are 26 specimens of *G. chalcogrammus* from the order and family Gadiformes Gadidae and 23 specimens of *A. platyrhynchos* from the order and family Accipitriformes Accipitridae.

*Using ontologies to describe traits*

The UVicZL dataset allows specimens to be queried and filtered by sex, age class or 'lifestage', and element type. Some of the original catalogue cards and associated excel worksheets contained sex and age determinations. For instance, 561 animal specimens have information recorded in the sex field, <dwc:sex>, including standardized terminology and a DOI link to the tDAR faunal sex ontology (**Table 5**). Similarly, 360 animal specimens have information recorded in the lifestage field, <dwc:lifestage>, which displays standardized age terminology and a DOI link to the tDAR faunal age ontology (**Table 6**).

**Table 5** Animal specimens (n=561) following the tDAR faunal sex ontology terminology. The sex field <dwc:sex> contains the term and link.

| Sex type | tDAR Faunal Sex Ontology https://core.tdar.org/ontology/3026/fauna-sex-ontology |
|---|---|
| **male** | 260 |
| **female** | 261 |
| **indeterminate** | 3 |
| **likely male** | 17 |
| **likely female** | 20 |
| | 561 |

**Table 6** Animal specimens (n=360) following the tDAR faunal age ontology terminology. The lifestage field <dwc:lifestage> contains the term and link.

| Age type | tDAR Faunal Age Ontology https://core.tdar.org/ontology/3030/fauna-age-ontology |
|---|---|
| **adult** | 174 |
| **indeterminate** | 3 |
| **subadult** | 62 |
| **juvenile** | 90 |
| **immature** | 26 |
| **neonate** | 3 |
| **fetal** | 2 |
| | 360 |

Individual specimens in the UVicZL comparative collection are mostly complete skeletons. 1,412 specimens had skeletal information recorded in either the original spreadsheets or the scanned catalogue cards, which was copied or transposed manually into the <dwc:preparations> field as element terminology. Also included was an ontology link(s) from tDAR and/or UBERON. Although some have missing elements or are only skulls, 315 fish specimens, in additional to skeletal remains, have scales (**Table 7**).

**Table 7** Individual specimens by element type. Records from 1,412 specimens were transcribed with element related data. The preparations field <dwc:preparations> for each was updated with the tDAR element ontology terminology and link and/or the OLS UBERON terminology and PURL link.

| Element type | OLS UBERON skeletal ontology | UBERON PURL | tDAR Faunal Element Ontology https://core.tdar.org/ontology/6029/fauna-element-ontology |
|---|---|---|---|
| vertebrate elements | | | 649 |
| skeleton | 649 | http://purl.obolibrary.org/obo/UBERON_0004288 | |
| skull | | | 64 |
| skull | 64 | http://purl.obolibrary.org/obo/UBERON_0003129 | |
| scales | | | 315 |
| scales | 315 | http://purl.obolibrary.org/obo/UBERON_0002542 | |
| mollusc | | | 384 |
| mollusc shell | 370 | http://purl.obolibrary.org/obo/UBERON_0008270 | |
| shell | 14 | http://purl.obolibrary.org/obo/UBERON_0006612 | |
| | 1,412 | | 1,412 |

*Digital images*

I scanned 2,795 images of the primary catalogue cards and uploaded these as pdfs to OSF. Each image is linked to the dataset by a unique OSF GUID, representing 96% of the collection specimens. Users are able to view these original sources of specimen collection data as seen in **Figure 3**.

## 2.5 Discussion and Recommendations

### A path for FAIR open collections

*Expanding collection capacity*

This study has demonstrated that open science tools and terminologies offer a path for comparative collections in zooarchaeology labs to be linked to other initiatives and become recognized for their value as identification resources and utility as natural history and biodiversity archives. While zooarchaeological comparative collections remain valuable within institutions, when the information is fully digitized and connected to other repositories they become more visible both inside and outside the subdiscipline of zooarchaeology. Achieving FAIR goals makes them more broadly useful resources. The urgency to preserve and make information more accessible from natural history collections is also an under-recognized value of zooarchaeological comparative collections, but requires greater attention to collection data management, expertise, training, and funding. The UVicZL has a history of influence that is especially relevant to archaeological records in the region and invested time in detailed documentation of the collection over 40 years. These collections are key to facilitating research and the interpretation of the dynamic natural and cultural history of human-animal relationships along the Pacific Coast and Northwestern North America. My approach using the extended specimen concept in annotated collections achieves each of the FAIR principles and, as evidenced by the thousands of downloads and 53 citations on GBIF, demonstrates the information contained in the collection is relevant to numerous contemporary environmental and ecological research efforts. The highly granular descriptions and standardized taxonomies, ontologies, locations, and biometrics (sex, age, weight, length, etc.) is a key aspect which enables these extended specimens to be integrated into larger natural history research data alongside internationally recognized museums and archives. This research also illustrates the value of placing zooarchaeological comparative collections among other ecological, historical, biological, and interdisciplinary archives. While often small or modest in size, the large number of regional and special purpose collections are particularly important for establishing biodiversity baseline data (Wiley et al. 2017) and generating data derived from zooarchaeological identifications. Although the UVicZL occurrence data that has been download from GBIF is collectively cited with many other natural history and biological datasets, it is one of three versions of open data (the other two being Canadensys and OSF) that enable citation of the collection dataset and make individual specimen citation possible. Literature using the GBIF citation recommendation and assigned DOI[46] shows that FAIR open collection citations are useful in archaeology (Hillis et al. 2022 p. 1396) and beyond. At this time, UVicZL appears to be the first standalone zooarchaeological comparative collection listed on GBIF. This research is a step toward engaging regional collections with biodiversity audiences.

By augmenting and making the data open in the collection, regular users as well as members of the wider research and educational community, including museums and archives, government agencies, biodiversity initiatives, and Indigenous organizations and communities now have access and can choose to engage this information. Using the methods outlined in this chapter also opens the collection to interdisciplinary researchers, organizations, and institutions outside of archaeology. These include professional consultants in biology and ecology, graduate students and faculty, and non-governmental organizations (NGOs) in conservation and resource

[46] McKenzie K, McKechnie I (2023). University of Victoria - Zooarchaeology Lab Collection. Version 1.10. University of Victoria. Occurrence dataset https://doi.org/10.5886/jej09d accessed via GBIF.org on yyyy-mm-dd.

management fields. Developing outreach education partnerships with wider reach, including school districts, Indigenous communities, and other organizations can further spark interest in archaeology, biology, and ecology and broaden the reach of the collection resources. Numerous regional museums who also hold collections, including the Royal British Columbia Museum (RBCM), the Museum of the North/Arctos, the University of British Columbia (UBC) Beaty Biodiversity collection, UBC Museum of Anthropology (MOA), and the Reciprocal Research Network, may find sharing collection strategies to be useful. Government agencies such as the BC Archaeology Branch[47], the Department of Fisheries and Oceans (DFO)[48], and provincial conservation officers can also provide links to the collection for researchers and consultants. Researchers can readily link specific specimens and details about animals in the collection. Which is important for potential future identification refinement and to judge the adequacy of taxonomic representation. Such efforts to promote increased transparency in methods is key to improve confidence and quality control in the zooarchaeological research process. This starts by being transparent about the identification process in zooarchaeology in relation to the comparative specimens. This is particularly critical in the issuance of Heritage Conservation Act (HCA) permits to conduct archaeological research in British Columbia, where a large number of projects have occurred. Currently many HCA permits do not stipulate standards for zooarchaeological bone identification. Numerous biodiversity initiatives, including MarineGEO - Smithsonian[49], Group on Earth Observations Biodiversity Observation Network (GEO BON)[50] and Canadian Biodiversity Observation Network (CANBON), Hakai Institute (Tula

[47] https://www2.gov.bc.ca/gov/content/industry/natural-resource-use/archaeology
[48] https://www.dfo-mpo.gc.ca/index-eng.html
[49] https://marinegeo.si.edu/
[50] https://geobon.org/

Foundation)[51], and the UN Ocean Decade[52], are in positions to share collection data within their networks. Indigenous communities and scholars are sought as collaborative partners and could benefit from accessing and utilizing the collection as well as the digitized data. Developing relationships with Nations harnesses greater potential to answer larger questions about animals that occur their territories and to advance their resource and land management practices, language revitalization, and food sovereignty objectives.

*Improving methods in zooarchaeology*

As many practitioners have noted, identification methods in zooarchaeology deserve greater attention and have larger implications for science. For instance, an influential 'blind test' of morphologically based identifications noted disjuncts between analysts and collections examining the same assemblages of fish bones (Gobalet 2001). Others have advocated for analysts to improve replicability by making methods and data available for critique and reuse (Wolverton 2013). The replicability crisis in science and archaeology depends on the practices of recording, describing, and sharing methods and data (Marwick et al. 2017). Practices that describe study limitations, explain inferences and assumptions, and clarify interpretation support improvements in standards. For example, documenting the sorting of remains, and the procedures for element and taxon identification, including comparative collection citation and anatomical attributes of specimens used for identification, reduces bias and lends credibility to replication.

Following best practice methods, analysts are encouraged to describe their identification methods and cite the reference collection, physical location of lab facilities, and the precise

[51] https://hakai.org/about/
[52] https://oceandecade.org/

osteological traits used for identification (Driver 2011, Gifford-Gonzalez 2018, Lyman 2017). The UVicZL collection has been used and cited in a majority of NWC identifications (McKechnie and Moss 2016, McKechnie et al. 2020), but despite active use of comparative collections by zooarchaeologists, individual comparative specimens are rarely reported in zooarchaeological research (Driver 2011, Gifford-Gonzalez 2018, Lyman 2010). Many digitized ZCC specimens include detailed information, and with accessible documentation some specimens have potential to be regarded as 'type' specimens.

The UVicZL collection is now formatted in a way where descriptive biometric data can be added to individual specimens. Each extended specimen used by analysts includes links to taxonomies, ontologies, geospatial data, images, related research, and, as discussed in the following chapter, Indigenous language areas (**Figure 13**). While seemingly elemental, these examples are applied broadly throughout the database and are crucial to the way data from this collection can be connected to diverse range of archives, collections, and observations globally. When these specimens are used as a data source, and the occurrence's assigned URI is documented as an individual specimen citation[53], this practice becomes a standard zooarchaeological method. Analysts who frequently use the UVicZL collection to identify faunal remains can now use the DOI to cite the collection as well as individual specimens when reporting on their methods. Disciplinary methods and data transparency improve when bone identification is paired with specific comparative collections and individual specimens. By citing the collection and specific specimens used to make their identifications, this additional step is a

[53] For example, if specimen UVicZL-16.1 (*Odocoileus hemionus*/deer) was used to identify an archaeological deer bone, the citation for this specimen would be:
McKenzie K, McKechnie I (2023). University of Victoria - Zooarchaeology Lab Collection. Version 1.10. University of Victoria. Occurrence dataset https://doi.org/10.5886/jej09d accessed via GBIF.org on yyyy-mm-dd. https://www.gbif.org/occurrence/3329069412.

vital opportunity to enable comparability and replication to assess accuracy. This practice also builds replicability with recurrent citation of specific comparative specimens as 'type' specimens (under which many identifications have and will be based).

Although currently there is no convention to cite specimens using PIDs, this is an increasingly relevant issue in zooarchaeological identifications and something the bioinformatics community is working to resolve (Mabry et al. 2022). Simply using catalogue numbers or DwC triplets (<dwc:institutionCode>: <dwc:collectionCode> : <dwc:catalogNumber>) for individual specimens as a citation practice does not prevent the existence of duplicates. Ideally, collection institutions should publish dataset PIDS for specimens to enhance tracking and attribution, and to better facilitate access to 'type' specimens. Better integration and linking of specimens across data resources, such as repositories, using APIs is a clear benefit of specimen PIDs (ibid). This practice will strengthen data and metadata for increased interoperability and reuse. Collections managers and data stewards are encouraged to include more detail about their collection specimens and document metadata. To broaden FAIR collection practices, I recommend collection managers and lab instructors implement and encourage the broader use of Darwin Core vocabularies and consider publishing datasets to make their comparative collections available on integrative repositories like GBIF.

By following open science initiatives to guide and enhance accessibility and discoverability, my research, as applied to the UVicZL collection, provides a framework for FAIR and open ZCCs globally. As an open and searchable dataset, the UVicZL comparative collection's regional taxonomic diversity is clearly demonstrated placing it as a valuable resource to zooarchaeologists, particularly those working in northwestern North America. Indexed sites like Knochenarbeit and the development of regional directories can increase visibility of a collection and allows researchers to peruse the inventory and understand the composition of specimens available, including taxa, elements, sex, etc. For example, a zooarchaeological researcher interested in comparing goats could determine the adequacy of a collection and how that may affect data results. The creation of an online ZCC registry with regional search functions would allow UVicZL and similar collections to be discovered by researchers focusing on specific taxa and regions. These regionally specific collections can provide data on species occurrence areas, useful knowledge for considering the range of animals that occur which is crucial for analyzing animal remains from archaeological contexts. Knowing the range of species that occur in an areas can also be used as a teaching tool for students, consultants and field technicians, and Indigenous archaeologists. The data from the UVicZL collection in this sense is reusable and can be used to support environmental background and faunal analysis reporting. Additionally, the documented biometrics shows collection diversity and provide analysts with multiple specimens of varying size, age, and sex for more refined comparisons such as domestication. Using tDAR and UBERON ontologies to describe the skeletal elements of specimens allows analysts to use the collection as a guide to map to their faunal remain terminology. Using a higher-level node such as “Vertebrate Element” or “Skeleton” as a basic description for each specimen leaves open the future possibility of linking identification of individual bones to a detailed ontological terminology after performing a more detailed collection inventory.

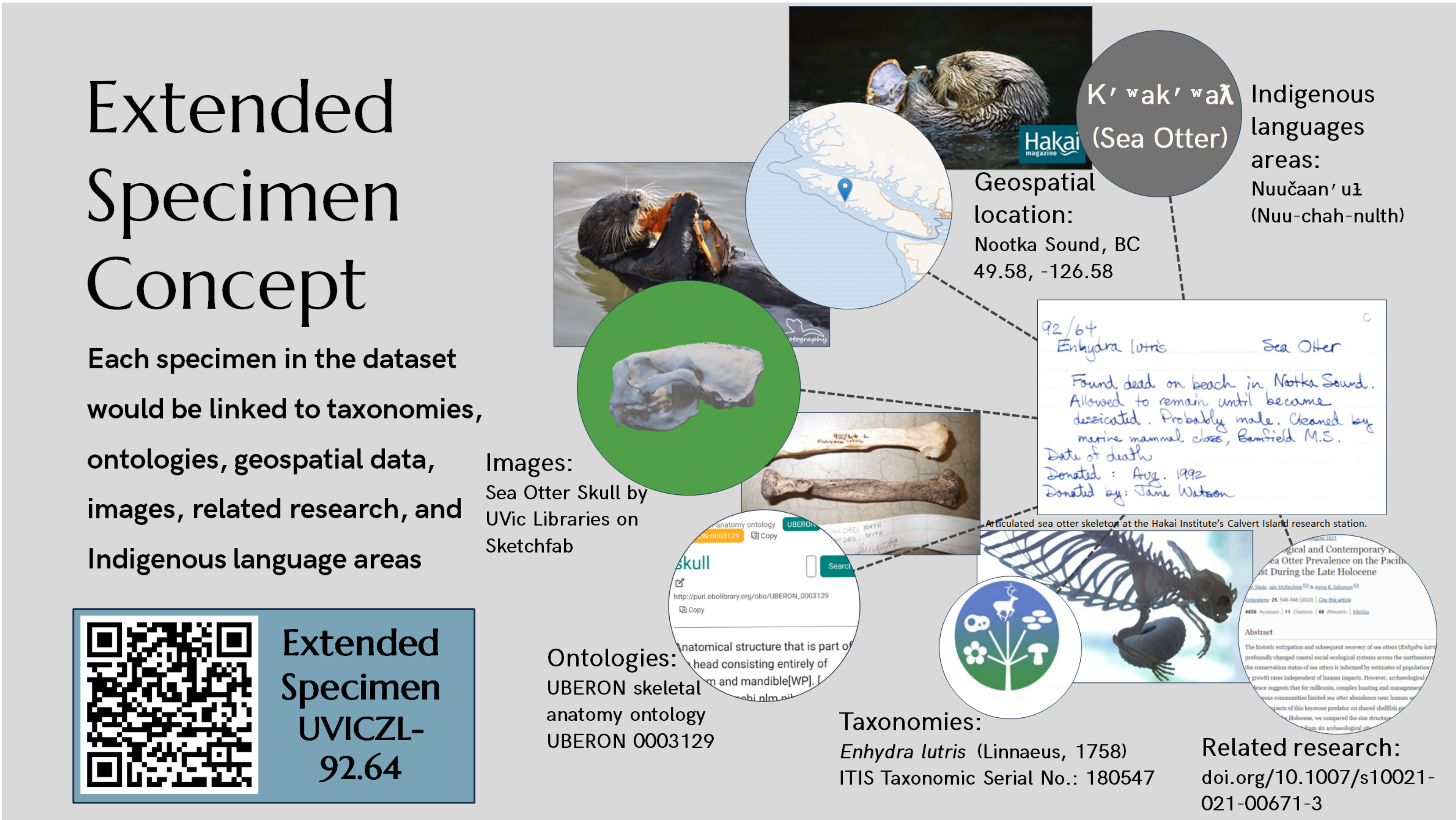



**Figure 13** UVicZL-92.64 as an "extended specimen" with links to taxonomies, ontologies, geospatial data, images, related research, and Indigenous language areas.

**Planning for future capacity**

*Ancillary identification methods*

The collection has the capacity to link additional identification sources to dataset specimens. These could include zooarchaeology by mass spectrometry (ZooMS), additional 3D imaging of collection specimen elements, and 2D photos. For example, open access ZooMS standardized peptide markers for taxonomic identification could be linked to reference specimen data (Brown et al. 2021 p. 234). ZooMS is the collagen peptide mass fingerprinting technique that provides a proteomic based identification of osteological animal remains (Richter et al. 2022). Several specimens from the collection have had ZooMS analysis and, for these, primary ZooMS data will be made available with the prospect of adding more in the future. Challenges arise with ZooMS DwC data entry as there is, to date, no suitable DwC term. Currently the only way to document collagen peptide mass fingerprinting is using the <dwc:dynamicProperties> term (LeFebvre 2022). Image files containing the sequences could foreseeably be uploaded as simple multimedia links providing researchers with identification confirmation and/or baseline data for otherwise unrecognizable fragmentary remains. This would provide more robust citation of specific specimens used for identification. Other open access projects with 3D images that can be downloaded with 3D model printing instructions could also be linked (Blackburn et al. 2024). Additional resources that would add value to researcher include links to morphology trait data, bone counts, and published journal articles that are related to the animal represented by the specimen.

*Repository and taxonomic authority expansion*

Taxonomy can be used to understand collection breadth and natural history. Using different open source tools such as ITIS, NCBI, FishBase, and GBIF is advantageous as they improve the biodiversity informatics landscape. These tools create more possibilities for communication and translation between dispersed datasets and collections which amplify the power of data compilation. In addition to more comprehensive coverage, each of these sources has unique strengths or specializations in certain taxonomic groups or geographic regions. For example, FishBase (Froese and Pauly 2024) was among the first open science efforts to focus on global distribution of fish species[54]. More recent efforts have built on this to create a broader database of marine species such as the World Register of Marine Species (WoRMS). These initiatives create a space for researchers with specific interests where access to TSNs and linked platforms are tailored to their focal taxa, time periods, and regions of interest. Similarly, databases, like ITIS and Catalogue of Life (CoL), are known for providing authoritative and well-curated taxonomic data. Cross-referencing information from multiple reliable sources enhances accuracy of taxonomic classifications and enables updates and revisions to often dynamic taxonomic information based on new research findings. While different databases may update their information at different rates, using multiple sources allows researchers to capture the most current taxonomic knowledge. Also, access to a diverse set of data sources in various formats allows researchers to choose sources that best suit their analytical tools and workflows. This flexibility in data integration and analysis can enhance cross referencing information from different sources to identify discrepancies and errors for a more comprehensive analysis. Some databases have a global focus, such as GBIF, while others may emphasize regional biodiversity.

[54] https://www.fishbase.org.au/manual/English/fishbaseThe_Making_of_FishBase.htm

Combining global and regional perspectives allows researchers to address questions at different spatial scales. The integration of data and collaboration across these platforms in biodiversity research offers a robust and comprehensive approach. It helps mitigate the limitations and biases inherent in individual databases, enhances the accuracy of taxonomic information, and supports a more holistic understanding of the diversity of life on Earth.

Collaborations within the biodiversity informatics landscape are highly connected and broadly represent the collection to appeal to current and future users. GBIF, Ocean Biogeographic Information System (OBIS), CoL WoRMS, FishBase, and ITIS were among the top ten connected sources in 2017 (Bingham et al. 2017) (**Figure 14**). These sources represent a few of the data curators, creators, and distributors within the diverse biodiversity informatics landscape. This quickly evolving network has since expanded with further collaboration and initiatives like the global species checklist. The global OBIS data platform uses the WoRMS taxonomic database for taxon identification. WoRMS aims to provide an authoritative and comprehensive list of names of marine organisms. The fish, shellfish, and marine mammals contained in the UVicZL collection are well suited to this platform and will reach even more biodiversity researchers focusing on marine species. The taxonomic information can be easily integrated using R scripts and the WoRMS API and published to OBIS similar to how it is published to GBIF. By adopting TSNs and links to these highly connected sources, UVicZL can greatly expand taxonomic information, and given that much of the collection is marine, further integration with OBIS using WoRMS, can improve the global biodiversity observation network being organized in the GEO BON.

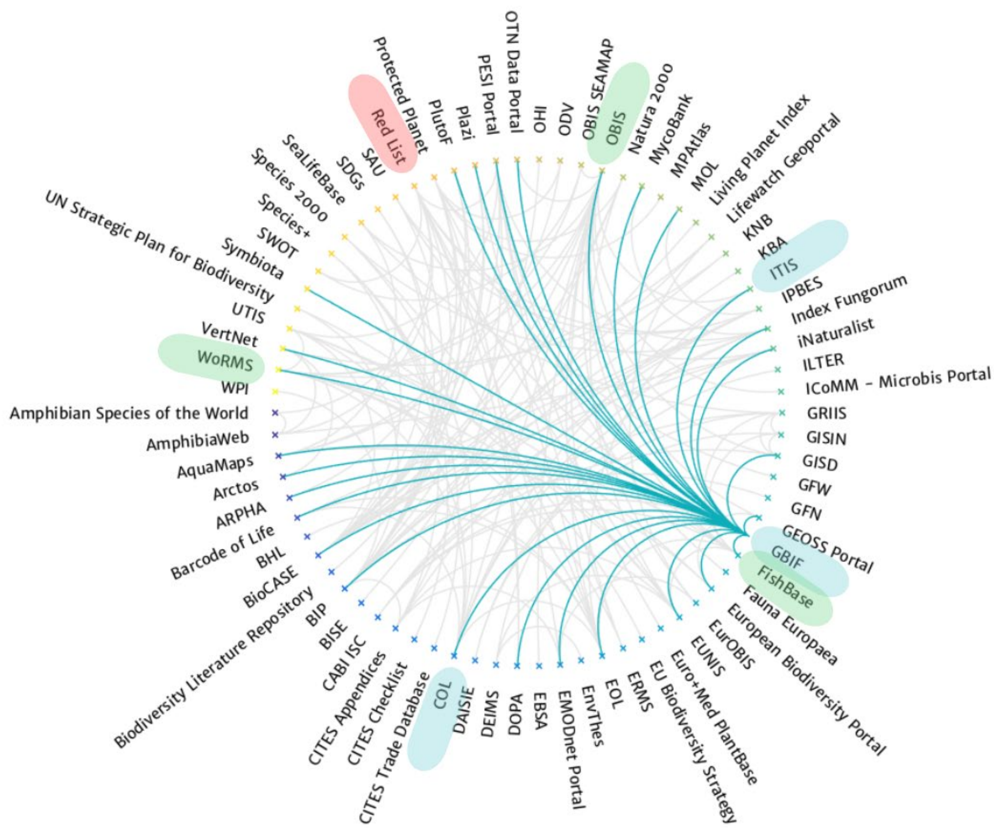



**Figure 14** The biodiversity informatics landscape with connections to GBIF in 2017. Source: modified from Bingham et al. 2017.

# Chapter 3: Facilitating an Ecology of Knowledges and Implementing CARE Principles in Zooarchaeology Collections

## 3.1 Chapter Summary

In this chapter, I examine my third research question: *How can open collections adopt CARE principles and connect with Indigenous histories and communities to foster more equitable sharing and reuse of data to positively impact communities and histories zooarchaeologists study in the present and future?* To address this, I explore how foundational knowledge, both within the discipline of anthropology and in Western science more broadly, has impacted the way information is collected, presented, and mobilized. Because the use of open science tools and the principles of FAIR (Findable, Accessible, Interoperable, Reusable) can come into conflict with Indigenous data sovereignty and CARE (Collective benefit, Authority to control, Responsibility, and Ethics) principles, it is critical to examine how control and access to ways of knowing can be accommodated. This begins by acknowledging power and positionality and leads to facilitating collaborative discourse beyond Western frameworks of knowledge production. Such work is needed to rebuild and strengthen relationships between academic and scientific institutions and Indigenous communities who hold inherent rights and are increasingly seeking to reclaim data about them held by these institutions. As an open collection, my efforts to make the UVicZL FAIR has created an opportunity to broaden collection scope to engage and connect with Indigenous communities and knowledge keepers in ways that also support CARE and Indigenous data sovereignty principles. Efforts toward reframing knowledge result in positive outcomes by stimulating engagement in education and research practices and strengthening contemporary knowledge of human and ecological resilience.

## 3.2 Introduction

As previously explored in Chapter 2, zooarchaeological comparative collections (ZCCs) are fundamental to archaeological inquiry and provide an empirical basis for animal bone identification. Archaeology, however, produces knowledge that invariably retains interpretive and methodological attachments to past epistemologies, particularly the foundations of Western science (Henare et al. 2007). While the UVic Zooarchaeology Lab (UVicZL) collection has unique design features specific to the identification of animal bone archaeology, it also falls squarely into this broader tradition, representing a subset of Natural History Collections (NHCs) constructed with a fully Western framework, and which remains central to broadly 'representative' collections in national museums of natural history globally (e.g., Washington DC, London, Paris, etc.). For NHCs, hierarchical Linnean taxonomic classification using the Latin binomial nomenclature convention (i.e., *Genus name, species name*) is an organizing principal and is widely recognized as the scientific standard internationally. There are rational reasons for this effort, including the scientific ambition to classify all forms of life and the more practical ability to name and reference species globally.  However, this categorical, hierarchical, and reductionist view of classifying types of lifeforms fails to reveal the many political, racist, and colonial motivations, circumstances, and actions involved in its formation (Trisos et al. 2021), nor how local peoples who may 'also' know these animals much more deeply understand the interactions and relationships between them. Akin to interpretation from archaeological data, without a broader perspective incorporating both knowledge and conceptual categories, scientific inquiry will only produce a partial view of complex phenomena and concurrences (Martindale and Nicholas 2014).

As has been widely noted, the legacy of anthropologists, and more generally, scientists and academics as “experts” deserves reflection when seeking to impose an ontological framework that perpetuates colonial legacies of categorization (Trouillot 2003, Stahl 2020). By shining a “critical lens” on conventionally deployed theory and methods, especially historically influential epistemologies, it becomes possible to set such hierarchical schema alongside emergent decolonizing practices (Stahl 2020). As researchers and academics are becoming increasingly cognizant of the history of extractive and inappropriately applied or acquired Indigenous knowledge, advocates argue for a shift in this trajectory towards supporting inherent Indigenous rights and sovereignty (Jennings et al. 2023).

**Toward an ecology of knowledges: Creating epistemological relationships**

Ann Stahl’s concept, of effective archaeology provides a point of departure to consider how multiple forms of knowledge might be reconciled. She discusses how to attain an “ecology of knowledges” from Boaventura de Sousa Santo’s 2016 book, *Epistemologies of the South*. Rather than refraining from assigning value to kinds of knowledge, the goal is to develop a knowledge discourse aimed at “overcoming forms of ignorance *learned through*, and *internal to*, an epistemology” (Stahl 2020 p. 40). From this perspective, ecology is, in definitive terms, a set of interdependent relationships or relations and interactions, and represents relationships between knowledges that can coexist in tension and countervailing dialogue. Set in the context of my research question, I recognize that a goal of an open zooarchaeology comparative collection is to provide a venue for broader understanding including the relationships between scientific and Indigenous knowledge of animals. In this sense, my work digitizing specimen information is an effort to consider how I can pursue “effective” archaeological inquiry and navigate within this space of tension by situating processes and providing a venue for connecting knowledge. In

Stahl's framing, "effective archaeology understands that often things were and could have been otherwise and that through our disciplinary practice we can illuminate the processes through which particular futures past came to be" (Stahl 2020 pp. 42–43). This is one of several paths toward decolonizing anthropological practices and diversifying knowledge.

One recent example is provided by Gupta and Stoolman (2022) who suggest a decolonial collaboration with Indigenous communities by asking whether research is beneficial to peoples with whom we work, and more broadly, what our discipline would have become "if it had been constituted as a decolonizing project?" (Gupta and Stoolman 2022 p. 5). For each scenario they envision the capacity for multidisciplinary training and dialogue, as well as richer collaboration, knowledge, and inclusion with Indigenous Nations. Although these interactions are bound to create "moments of incommensurability" and "epistemological disconcertment", such acts establish collaborative frameworks to develop a new knowledge discourse (Stahl 2020 p. 43). "Robust collaboration fosters critical engagement, forces reflection on whose and how facts are valued by bringing into view the epistemological hierarchies and ontological premises that configure fact-making" (ibid p. 43). Collaboration, as such, is not ancillary, it entails complete transparency and equality in power and challenges epistemological and causal contributions to archaeological interpretation and the configuration of power (Martindale and Nicholas 2014). Rather than expelling science, by engaging in collaboration, the discipline and its historical manifestations are more appropriately delimited. By setting boundaries around anthropological and Western science foundations, critical engagement in knowledge discourse with Indigenous communities can be thought of as "borderlands" or a "place of incommensurable contradictions" (Gupta and Ferguson 1997 p. 48). This liminal space between sets of knowledges are places where incommensurable ways of knowing can form a relationship – an ecology of knowledges (ibid). An ecology of knowledges views ontological premises as an avenue to envision novel ways of thinking and relating without determining the nature of reality. My research examines how the UVicZL collection can be used as tangible catalyst to engage in discourse relating to archaeological research on human-animal interactions.

**Creating a relationship between knowledges with open collections**

The development of comparative osteological collections containing modern faunal remains is a key archaeological method to understanding past peoples' relationships with animals and how they are represented and identified in an archaeological context. These biological science collections are typically organized using biological science with binomial nomenclature, phylogenetic relationships, and anatomical traits and categories. In an effort to minimize and refine ambiguous identification or categorization, the Linnean taxonomic naming system uses hierarchies and sub-classifications of the basal 'species' that fully embrace and reflect Western scientific approaches to biological definitions. Similarly, archaeologists and zooarchaeologists use and rely heavily on the species concept. However, given the fragmented nature of archaeological animal bones, zooarchaeologists can often only identify animals less than species specific. Many skeletal specimens can only be confidently grouped into less specific composite taxonomic categories such as flatfish or ungulate (Driver et al. 2011). Less consequential, but also having an impact on species concepts as well as collections management, is the need for zooarchaeological analysts to be aware of the frequent debates and formal revisions to animal taxonomies (Driver 2011, Lyman 2017). The ongoing refinement and revision to classification by taxonomists and biologists continues as genomic, geometric, and morphometric analyses, as well as phylogenetic and developmental research increases resolution and specificity. Changes to past names and renaming practices are expected and built into this scientific process. With current data science methods and networked technological capacities, incorporating multiple terms for a type of organism is feasible. Of relevance to this case study, associations between Linnean species names and Indigenous language terms for those animals exist and changes to taxonomic terms can be inclusive of many Indigenous names. With linked open data there is also the ability to cross-reference multiple names and expand associated definitions, as well as various international and practical orthographies (linguistically structured spellings and renderings of spoken words). The ability to create many categories and meanings holds the potential to extend the breadth of knowledge about diverse animal species. Such changes are also aligned with developing refined scientific methods of understanding such as mapping of locations where specific species occur as spoken in Indigenous languages.

A goal of verbal (and written) communication is to achieve mutual understanding. From a Western perspective, this requires an equivalency in meaning of words within languages and analytical categories. Archaeological data often seeks to create or adopt taxonomies and nomenclatures as these categories can be central to classifying and organizing information and aspects of archaeological interpretation. A core focus of my research into the specimens in the UVicZL collection includes determining how the animal species are classified, identified, and documented using binomial nomenclature and English common names. In the previous chapter, I explain how I address FAIR principles and describe the creation of machine-readable data from a largely paper-based record into a digitized open collection that uses taxonomic identifiers. In this chapter, I shift focus and discuss how I apply CARE principles using open science by considering Indigenous languages as an integral component of this dataset of animals primarily from the Northwest Coast of North America and the North Pacific. I develop digital tools to create visual geographic links to collection specimens and match specimen collection locations to associated Indigenous language areas. I also adopt digital pathways for Indigenous communities to collaborate to enable Indigenous peoples and organizations to pursue data sovereignty. This research makes efforts to bridge an aspect of the larger incommensurability of Western scientific and Indigenous knowledges by creating an open collection that is structured to be inclusive of animal names in both Western and ethnobiological taxonomies that are not necessarily equivalent. This effort to decolonize an aspect of zooarchaeology endeavours to restore relationships with Indigenous people and democratize control over the naming practices that are historically laden with hierarchical classifications established by scientific authorities. This work also seeks to broaden audiences, stimulate collaborative practices, and increase collection use in archaeological investigations and public education. I also consider the potential for policy implications as zooarchaeological bone assemblages identified using the UVicZL collection can influence resource management practices and assist in Indigenous food sovereignty. Working within anthropology, I follow an ecology of knowledges approach to reframe discourse, acknowledge disciplinary foundations, and encourage future research that works towards decolonization.

*CARE principles and Indigenous data governance*

Indigenous data has been defined as any "…data, information, knowledge, in any format, that impacts Indigenous peoples, nations, and communities at the collective and individual levels; data about their resources and environments, data about them as individuals, and data about them as collectives" (Carroll et al. 2021 p. 1). These can include formally or informally conceptualized data collected or created by Indigenous or non-Indigenous people or entities about Indigenous lands, water, knowledges, oral histories and stories, and cultural sites (Gupta et al. 2023). Metadata, or "data about data", describes and summarizes data attributes within a dataset. It is designed to assist users locate, apply, or manage data (Farnel 2018). Defining metadata as a type of data requires the development of ethics for its design, use, and sharing. Correcting or revising inadequate Indigenous data representation is an objective for (meta)data curation. A concept that has influenced the definition of Indigenous data is OCAP® (Ownership, Control, Access, and Possession), which advocates data curation as key to ethical access and sharing (Schnarch 2004). Since conception OCAP® has been revised and recently trademarked. It is currently administered by the First Nations Information Governance Centre (FNIGC) and endorsed by many Indigenous communities and research organizations in North America. OCAP® principles enable Indigenous communities to claim data ownership, control, and protection to be enacted by the Indigenous communities to which metadata applies. The role of context and culture, biases in (meta)data tools, and imbalances in power and structural issues should be considered to create culturally appropriate Indigenous (meta)data (Farnel 2018). Considerate, sincere, and respectful engagement with, and involvement of Indigenous communities is necessary to develop (meta)data enriched by relevant Indigenous language and alternate categorizations (Reid et al. 2024).

Beginning with a framework such as OCAP® allows for thoughtful reflection on (meta)data sharing and access in relation to FAIR (Findable, Accessible, Interoperable, Reusable) and open principles. Although the free and open exchange of metadata can be beneficial, the ethical and appropriate sharing of (meta)data requires the consideration of “sensitive” data (Farnel 2018). Risk reduction measures should be enacted to protect the well-being of individuals or communities and any associated traditional or local knowledge that may cause harm. Crucially, data curators should consider how the “structural inequalities in the practice of science" impact communities whose (meta)data is revealed through adopting open science principles (Gupta et al. 2023).

Developing an ethical framework to display (meta)data allows curators to ameliorate inequalities. CARE principles, drawing upon the United Nations Declaration of Rights of Indigenous Peoples (UNDRIP), provide such an ethical scaffold for archaeologists to follow as they collect, interpret, preserve, curate, circulate, and reuse (meta)data over the data life cycle. (Gupta et al. 2023). Although Indigenous perspectives are changing questionable cultural heritage practices, digital data so too deserve consideration regarding care, access, and ownership. Ethical stewardship of digital heritage considers how scientific practices are permeated by colonialism and disproportionate power. CARE principles allow Indigenous peoples to reclaim data and places them as "active innovators, policy makers, practitioners, and contributors" (ibid).

## 3.3 Methods

### CARE Principles in a FAIR Open Collection

As I described in Chapter 2, to create an open collection and apply FAIR principles to the UVicZL collection I implemented the open access methods and tools further described in the Addendum and Workflow processes 1-6 (**Figure 1**). Following the first six processes in the workflow, to address my third research question, process seven encompasses the development of a geocoder to determine specimen collection coordinates, a searchable specimen map with Indigenous language areas, as well as a database and web outputs.

*Specimen geolocation*

Knowing the geolocation of an individual specimen collection point serves two purposes. First, to satisfy FAIR principles, specimens published in the Global Biodiversity Information Facility (GBIF) repository have coordinate points recorded in their occurrence record and are visible on repository derived maps within a predetermined area of uncertainty. Second, as a preliminary step towards CARE principles, this location is associated with open-source Indigenous language areas. The Indigenous language(s) currently spoken at the collection point could be included in the specimen record and be used as a starting point for identifying Indigenous animal names within this language area.

To add specimen geolocation data, I intended to model my approach after the paper by van Erp et al. (2015) who devised a method to improve geolocation data accuracy for natural history collection (NHC) animal specimens using the Centre for Biodiversity Naturalis Collection in the Netherlands (Naturalis) [55]. Vast amounts of biodiversity data are held by NHCs, including location, which can be used to determine geographic distribution, environmental change effects, and impacts on regional species (van Erp et al. 2015). Using a georeferencing aid to determine precise coordinates of specimen origin is essential to these analyses. Because NHCs often have species collected over great time periods, many older specimens lack precise geographic information, let alone geocoordinates. The immense number of specimens held in collections make manual georeferencing each specimen a time intense and expensive pursuit. Some of the challenges of georeferencing include discerning between locations with the same name, vague localities, linear feature measurements or ambiguity, two or more location

[55] https://www.naturalis.nl/en

descriptors, topological nesting, complex interpretive descriptions, changing political borders, and shifting historical place names. Tools have been developed for batch georeferencing such as BioGeoMancer[56] and GeoLocate[57], however neither use species occurrence domain-specific knowledge. van Erp et al. 2015 draw on the specimen occurrence geographical distribution domain knowledge held in GBIF to enhance their technique. The paper uses a knowledge driven approach using 150 test records obtained from Naturalis which is analyzed in five rule-based modules written in a programming language (Perl). The records were processed using record filter, text parsing, gazetteer lookup, offset calculation, and disambiguation heuristics modules, each refining geolocation accuracy. The disambiguation heuristics include spatial minimality, expedition clusters, and GBIF species occurrence data. Each module indicator is assigned a confidence score (-12 to 12), with 12 being the highest, to represent the trustworthiness of the georeferenced results. Confidence scores are dependent upon the quantity of available information contained in a specimen record, with each indicator having varying values. The authors of the paper incorporated these modules into a geocoder workbench, GeoImp (Georeference Improver), that was developed to analyze single or batch CSV records. A geocoder workbench is a software tool for extracting geospatial coordinate data. Specimen data entry produces a map, coordinates, and a confidence score. Unfortunately, the server holding the GeoImp code failed without a secondary backup. As this code was unavailable to reproduce, I needed to create a new workbench to apply to the UVicZL collection to better understand its regional focus and identify relationships between specimen locations and Indigenous language areas.

[56] https://sites.google.com/site/biogeomancerworkbench/home?authuser=0
[57] https://www.geo-locate.org/

Given the large number of records to be georeferenced in the UVicZL collection, my approach intended to follow an automated georeferencing process to refine available specimen location data, improve accuracy, and decrease manual annotation time. UVicZL specimens typically have written descriptions of where they were collected or retrieved derived from their catalogue card. Most specimens, except for a few, lack specific latitude and longitude coordinates. Determining accurate coordinates is challenging as many UVicZL collection records are deficient in location data or use ambiguous or locally known location descriptions. For example, some specimen records are recorded from "Dr. Charles Morgan's farm", which was likely well known by the collector but does not return a specific location result when searched in Google Maps. Location information was recorded in the Darwin Core (DwC) location fields and derived from the occurrence remarks field (**Table 8**) which was transcribed verbatim from either the original catalogue card or an excel spreadsheet containing specimen records. To determine geocoordinate latitude and longitude data, as the automated portion (Process 7, **Figure 1**) of the georeferencing process was taking place, and to ensure results were accurate and specimens could be plotted in GBIF, each record with sufficient location information was manually determined using Mapcarta.com[58], an open-source map application. These coordinates were then entered into the specimen record using the fields, <dwc:decimalLatitude> and <dwc:decimalLongitude>. To evaluate the accuracy of the geocoder, the difference in distance (km) between manually derived coordinates and geocoder derived coordinates was also calculated in separate fields.

[58] https://Mapcarta.com/

**Table 8** Darwin Core location information fields used to derive geocoordinates. The field <dwc:occurenceRemarks> contains verbatim location information from catalogue cards used to manually populate location fields. Each location field provides additional detail for the geocoder to refine the coordinate results.

| **Field name** | **Description** | **Example** |
|---|---|---|
| occurrenceRemarks | Comments or notes about the Occurrence | from Lake Washington (freshwater), length: 17cm, weight: 152.5g, donated: L.S. Smith (Fall '88), processed: Sept. 19/89 |
| continent | The name of the continent in which the Location occurs | North America |
| waterbody | The name of the water body in which the Location occurs | Pacific Ocean |
| islandGroup | The name of the island group in which the Location occurs | Broken Islands |
| island | The name of the island on or near which the Location occurs | Vancouver Island |
| country | The name of the country or major administrative unit in which the Location occurs | Canada |
| countryCode | The standard code for the country in which the Location occurs | CA |
| stateProvince | The name of the next smaller administrative region than country (state, province, canton, department, region, etc.) in which the Location occurs | BC |
| municipality | The full, unabbreviated name of the next smaller administrative region than county (city, municipality, etc.) in which the Location occurs. Do not use this term for a nearby named place that does not contain the actual location | Saanich |
| locality | The specific description of the place. Less specific geographic information can be provided in other geographic terms (higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description | Long Beach |

*Geocoder development*

Implementing automated georeferencing is the seventh process of the UVicZL workflow (**Figure** 1) following the methods described in Chapter 2. After reviewing NHC literature on geocoordinate determination, I consulted with the Hakai Institute geospatial and IT teams as part of my internship with the Tula Foundation and the MITACS Accelerate program. To determine coordinates for specimen collection points, a custom geocoder workbench was designed and implemented using different geocoding toolkits. Steve Van der Valk, from Hakai's IT team, developed the Python code to run the geocoders and uploaded the versions to a Hakai GitHub repository. This repository, currently forked into my personal public repository, will be migrated to a public UVic repository once all coding work has been completed. The geocoder workbench processes the digital data to enable a variety of uses and queries. Using programming language, the geocoder uses multiple written descriptions of locations to estimate geographical coordinates for each individual specimen. The geocoding work for each version is done by open-source toolkits, these services include Google Maps[59], Mapbox[60] and OpenCage[61]. The documentation for toolkits is extensive and answers questions concerning methodologies, strategies, and trade offs inherent to geocoding tasks, which will not be addressed here. The workbench uses a Python API interface that interacts with each geocoding service. Python was chosen for its readability and simplicity. The Python script reads the dataset as an excel formatted spreadsheet (main worksheet 'faunal_collection' in Faunal_MOFB.xlsx) and coverts it to a Pandas memory DataFrame[62]. The DataFrame is converted to a CSV file for input to each geocoding service. For each specimen it extracts words describing geographic locations from field names in (**Table 8**). It then 'cleans' and sends applicable words to the geocoder API with configuration and parameters to aid the selection of likely correct locations. The geocoder attempts validation of the geographic coordinate responses and stores them alongside the spreadsheet data. These results are compiled into an CSV output file with new fields displaying the coordinates. The resulting geocoded coordinates for each specimen can be plotted onto an open-source map which visualizes them geographically at collection origin locations (Hakai Institute and Tula Foundation 2023). The workbench compares geocoder locations with the manual coordinates and calculates a Haversine distance in kilometers. These distance calculations are displayed in the CSV output file.

---

[59] https://developers.google.com/maps/documentation/javascript/geocoding
[60] https://docs.mapbox.com/api/search/geocoding/
[61] https://opencagedata.com/
[62] https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_excel.html

*Indigenous language areas and specimen map*

The time and many steps involved in deriving coordinate data for individual specimens in the comparative collection is considerable. This was made more challenging by the sheer number of specimens and the variety of terms used. Few zooarchaeological comparative collections have gone to this level of effort, but once achieved, this mapping enables connections to Indigenous peoples, environments, and languages. The data and maps are currently hosted on a publicly available website[63], and connect to Indigenous language areas in an effort to address OCAP®, FAIR, and CARE principles. Taking advantage of open-source geospatial data, I integrate spatial boundaries for Indigenous language areas using the open source API from Native Land Digital[64]. These areas are incorporated into the map as area polygons with black outlines. Native Land Digital is an Indigenous-led non-profit based in Canada who have also followed an open science model to create a user-friendly searchable map of Indigenous territories, languages, and treaties. Their stated mission is to facilitate discussions about colonial history, Indigenous knowledge, and settler-Indigenous interactions. They also provide educational tools like maps and Territory Acknowledgement Guides and sources, including links to Indigenous Nation websites and provide a platform for Indigenous communities to share their own stories. This helps create spaces for non-Indigenous individuals to learn about the land they live on, its history, and how they can contribute to a more inclusive and collaborative future. The Native Land dataset is publicly available and is under a CC0 license. Anyone can freely incorporate the dataset into their own maps and applications using the Native Land Digital API. However, as Native Land

[63] https://zooarch.server.hakai.app/opencage-geocoded-native-lands/opencage-geocoded-native-lands
[64] https://native-land.ca/

states, these maps do not represent or intend to represent official or legal boundaries of any Indigenous Nations.

To connect the UVicZL specimens to the Indigenous languages spoken where each specimen occurs, the geocoder queries the Native Land Digital API with specimen collection location site coordinates. The Native Land Digital API is queried whenever the geocoder is run with an updated dataset, uploaded to the GitHub repository, and results reflect recent updates by Native Land to language areas. The geocoder creates a new geocoded spreadsheet version of the input dataset formatted as a CSV file listing the languages associated with collection points (Hakai Institute and Tula Foundation 2023). Each specimen location visualized on the world map receives a pin icon and is linked to a collection specimen and its data. By clicking on a pin, data associated with each record is shown in an information window. The map is queryable using single or multiple fields derived from the input data such as native language, vernacular (common) name, catalogue number and linked data such as taxonomic categories, TSNs, and ontologies. The search function can be customized further should there be a need.

*Database and web outputs*

The spreadsheet with approximations of the native languages spoken in the specimen's area of collection (provided by Native Land Digital) is converted into a relational database to enable standardized and programmatic querying. This enables easy export of queries for subsections of data and easy reuse of the digital components by anyone else (Hakai Institute and Tula Foundation 2023). The project uses utilities from Datasette[65] developed in Python which

[65] https://github.com/simonw/datasette

converts the data into a SQLite database[66] and creates website content[67] with a web map component[68] hosting a user-friendly and programmatic interface to these data. The resulting website is currently hosted on AWS infrastructure supplied by the Hakai Institute using an open source Dockerfile which can be deployed on numerous cloud vendors for free which enables portable re-hosting of the data and website in a few simple steps (ibid). Hosting of the map will shift to UVic and be accessible through the UVic Zooarchaeology Lab webpage[69]. A web interface for accessing the collection dataset is another goal for the UVic webpage as the dataset is currently only publicly available through the map interface, GBIF, and OSF.

---

66 https://datasette.io/tools/sqlite-utils
67 https://docs.datasette.io/en/stable/publish.html#publishing-to-fly
68 https://github.com/simonw/datasette-cluster-map
69 https://www.uvic.ca/socialsciences/anthropology/research/facilities/zooarchaeology%20lab/index.php

## 3.4 Results

### Specimens with coordinates

The manual process of deriving geocoordinates using Mapcarta generated 2,523 coordinate points representing a majority (86%) of the 2,922 specimen records (**Table 9**). The geocoder produced CSV output spreadsheets with 2,486 rows of data for Google Maps and Mapbox V6, and 2,227 rows for OpenCage. The performance of the geocoders revealed challenges with accuracy issues in deriving accurate coordinate points for all location descriptions. The Google Maps and Mapbox V6 versions translated geocoordinate locations for 98.5% (2,486 of 2,523) of the specimen records in the dataset with sufficient location information, whereas OpenCage translated 88% (2,227 of 2,523). The accuracy was greatest for coordinates derived from the Google Maps version where 66% (n=1,642) of points were within 49km of the manually derived coordinate points, whereas a lesser amount of Mapbox V6 and OpenCage coordinates were within the same range (n=1,414, n=1,011).

Programming code for the geocoding components of each version varied with result errors seemingly more frequent with the OpenCage version. For example, the OpenCage geocoder did not return distance results for 296 records that had manually derived coordinates. One third (33%) of the Google Maps, 43% of the Mapbox V6, and 54% of the OpenCage coordinates were over 50km away from the manual coordinates using a Haversine distance calculation. The Google Maps geocoder placed 14% of the results 500+km away from the manual coordinates, whereas the Mapbox V6 placed 26% and OpenCage placed 41% 500+km away. Despite providing information in numerous location fields as per (**Table 8**), the accuracy of the derived geocoordinates was affected by how the geocoders read, parsed, and prioritized text in each field. Often, only a portion of some placenames (e.g., Salt instead of Salt Spring) or a less specific location (e.g., Vancouver Island rather than Sidney) would be chosen. The generated coordinates would place specimens in a central location by preferentially choosing remarks in one field over another. For example, island rather than municipality was chosen for many specimens. Specifically, OpenCage placed 434 specimens from Victoria, BC in Victoria, Australia. These results ultimately would affect the Indigenous language area by placing many specimens from one language area into another language area. After testing the geocoders against the manually derived coordinate points, the most reliable method to plot specimen points on the maps to derive the Indigenous language areas for each specimen is manually.

**Table 9** CSV results for Google Maps, Mapbox V6, and OpenCage geocoder versions

| | **Google Maps** | **Mapbox V6** | **OpenCage** |
|---|---|---|---|
| CSV results | | | |
| number of rows of geocoded coordinates | 2486 | 2486 | 2227 |
| number of Mapcarta coordinates | 2523 | 2523 | 2523 |
| | | | |
| Distance results | 2486 differences calculated | 2486 differences calculated | 2227 differences calculated |
| exact-<1km | **843** | **278** | **796** |
| 1-49km | **799** | **1136** | **215** |
| 50-99km | 58 | 90 | 50 |
| 100-499km | 426 | 329 | 242 |
| 500-999km | 207 | 364 | 198 |
| 1000+km | 153 | 289 | 726 |
| Total geocoder coordinates | **2486** | **2486** | **2227** |
| | | | |
| Within 49km | **1642** | **1414** | **1011** |

**Indigenous language areas map**

The map shows 2,523 coordinate points (**Figure 15**). Pin and numbered cluster icons can be zoomed in to reveal individual specimen data. The map is queryable using single or multiple DwC formatted data fields derived from the original "master workbook", Faunal_MOFB.xlsx, and key words. For example, by searching the waterBody field for "Puget Sound", 70 specimen records are returned on the map (**Figure 16**).

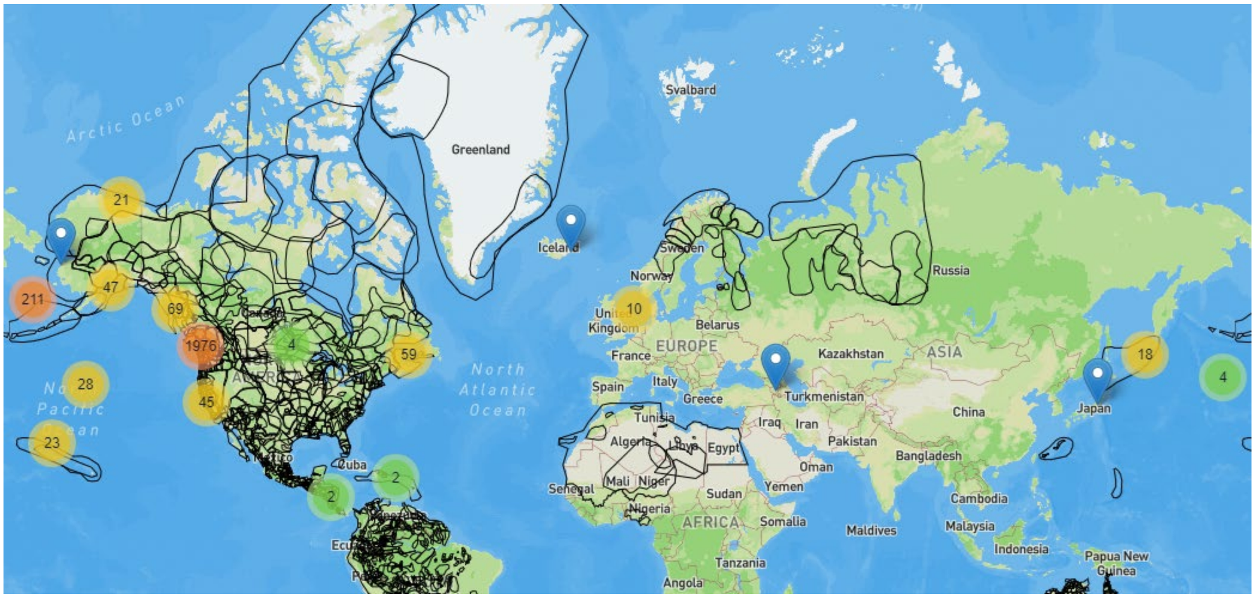



**Figure 15** Map of 2,523 specimens with geocoordinates with Indigenous language areas outlined in black from Native Land Digital. Location data from which geocoordinates could be derived is available for 86% (n=2,523) of the specimen records in the collection. Each specimen is indicated by a clickable blue teardrop pin on the map which is linked to the specimen record in the dataset. Specimens located near each other are grouped by the coloured circles with number figures.

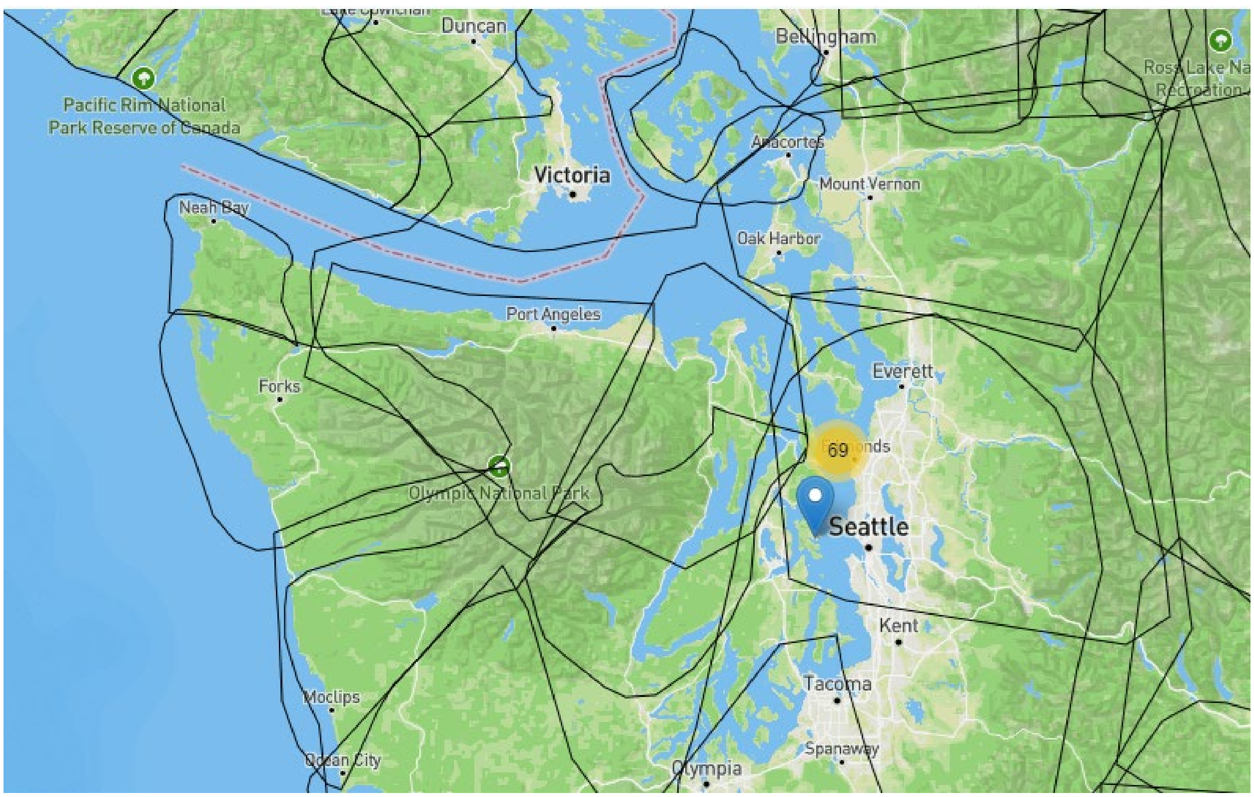



**Figure 16** Specimens collected from Puget Sound. Searching the waterBody field for “Puget Sound” results in 70 specimens visualized on the map.

To have a better understanding of languages represented by the specimens in the collection the Hakai geospatial team ran a Native Lands Digital API query using a simplified spreadsheet containing the catalogue numbers and manually determined latitude and longitude coordinates. This query produced 2,124 results (**Table 10**) and a simplified specimen map image (**Figure 17**). The language areas with the most numerous specimens (n=810) were from the SENĆOŦEN, Malchosen, Lekwungen, Semiahmoo, and T'Sou-ke dialects language area including greater Victoria, southern Vancouver Island and across the southern Gulf Islands to NW Washington. The next most numerous are 242 specimens from the Dakelh / Carrier (ᑕᗸᒡ) language area in interior British Columbia, followed by 196 specimens from Nuučaan̓uł / Nuu-chah-nulth language area on western Vancouver Island, 167 specimens from Hul'q'umi'num', Halq'eméylem, and hən̓q̓əmin̓əm̓ area on eastern central Vancouver Island and the Fraser Valley, and 134 from Éyɂáɂjuuthem; ʔayʔaǰuθəm area near Comox. A large number of specimens are from the Lushootseed language area in Washington State (119), and areas in Alaska including Unangam Tunuu / Aleut (87), and Łingít / Tlingit (58) language areas.

**Table 10** Number of UVicZL comparative specimens present across 58 Indigenous language areas returned from the Native Land Digital API representing 84% (2,124/2,523) of manually derived coordinates.

| Indigenous Language Orthographic / English (Syllabic) | Region | Number of specimens |
|---|---|---|
| **SENĆOŦEN; Malchosen; Lekwungen; Semiahmoo; T'Sou-ke** | Southern Vancouver Island | 810 |
| **Dakelh / Carrier (ᑕᗸᒡ)** | Central Interior BC | 242 |
| **Nuučaaṅuł / Nuu-chah-nulth** | Western Vancouver Island | 196 |
| **Hul'q'umi'num'; Halq'eméylem; həṅq̓əmiṅəm̓** | Vancouver Island & Fraser delta | 167 |
| **Éyɂáɂjuuthem; ʔayʔaǰuθəm / Comox** | Eastern Vancouver Island | 134 |
| **Lushootseed** | WA (Puget Sound) | 119 |
| **Unangam Tunuu / Aleut** | Aleutian Islands AK | 87 |
| **Łingít / Tlingit** | Coastal AK | 58 |
| **Secwepemcstín / Shuswap** | Interior BC | 38 |
| **Alutiiq / Sugpiaq** | Southwest Coastal Alaska | 32 |
| **Danezaa Zaageʔ / Beaver (ᑕᓀᑲ ᑲᕋ″)** | BC and AB | 19 |
| **Diitiidʔaatx̣ / Ditidaht** | Western Vancouver Island | 19 |
| **Inupiaq** | Northern Coastal AK | 19 |
| **Dakhóta / Eastern Dakota** | SK and MB | 18 |
| **Kashaya** | Northern Coastal CA | 13 |
| **Ktunaxa / Kootenay** | BC, AB and MT | 13 |
| **Ainu Itak** | Hokkaido Japan & Russian Federation | 12 |
| **Nsyilxcən; Ṅsyil̓xčṅ; Ṅsəl̓xčiṅ / Syilx Okanagan** | Interior BC | 12 |
| **X̱aad Kil; X̱aaydaa Kil / Haida** | Haida Gwaii BC | 10 |
| **qʷidiččaʔa·tx̌/ Makah** | Neah Bay WA | 9 |
| **Eyak** | Southern Coastal AK | 7 |
| **'Ōlelo Hawai'i** | Hawai'i | 7 |
| **Shoshoni** | ID, WY, NV, UT | 7 |
| **Tuwaduq / Twana** | Olympic Peninsula WA | 6 |
| **Koyukon** | Central AK | 5 |
| **Kwak̓wala** | NE Vancouver Island, Central Coast BC | 5 |
| **Dene Zágé' / Kaska** | Northern BC and YK | 4 |
| **Ipai** | Southern CA | 4 |
| **Columbia** | Central WA | 3 |
| **Hailhzaqvla** | Central Coast BC | 3 |
| **Northern Paiute** | Southern OR, Western NV | 3 |
| **Siuslaw** | Central Coast OR | 3 |
| **Sḵwx̱wú7mesh Sníchim / Squamish** | Squamish BC | 3 |
| **Sm'algya̱x / Tsimshian** | Central Coast BC | 3 |
| **Southern Pomo** | West Central CA | 3 |
| **Chorotega* (c/b other Central American)** | Costa Rica / Nicaragua | 2 |
| **Hiwatahia Taíno** | Puerto Rico | 2 |
| **Plains Cree / Nēhiyawēwin (ᓀᐦᐃᔭᐍᐏᐣ)** | AB and SK | 2 |
| **Sahtúot'ı̨nę Yatı̨́ / Northern Slavey** | NT | 2 |
| **Sallirmiutun [Siglitun]** | Coastal NT | 2 |
| **Southern Sierra Miwok* (c/b other CA location)** | Central CA | 2 |
| **Tillamook** | Central Coast OR | 2 |
| **Anishinaabemowin / Ojibwemowin** | Southern ON | 1 |
| **Central Yup'ik** | Western Coastal Alaska | 1 |
| **Chehalis** | Olympic Peninsula WA | 1 |
| **Dena'ina** | Southern Coastal AK | 1 |
| **Dän k'í / North Tutchone** | YK | 1 |
| **Erie* (s/b Anishinaanemowin)** | Bordering Lake Erie (US) | 1 |
| **Inland Łingít / Inland Tlingit** | BC and YK | 1 |
| **Konkow Maidu** | Northern Central CA | 1 |
| **Nakota** | AB | 1 |
| **Nuxalk / Bella Coola** | Central Coast BC | 1 |
| **Ohlone** | West Central CA | 1 |
| **Qaxun** | Aleutian Islands AK | 1 |
| **Serrano** | South Central CA | 1 |
| **Tse'khene / Sekani** | Northern Interior BC | 1 |
| **Tłı̨chǫ Yatıì / Dogrib** | NT | 1 |
| **Upper Kuskokwim** | Southern AK | 1 |
| | | 2124 (84%) |

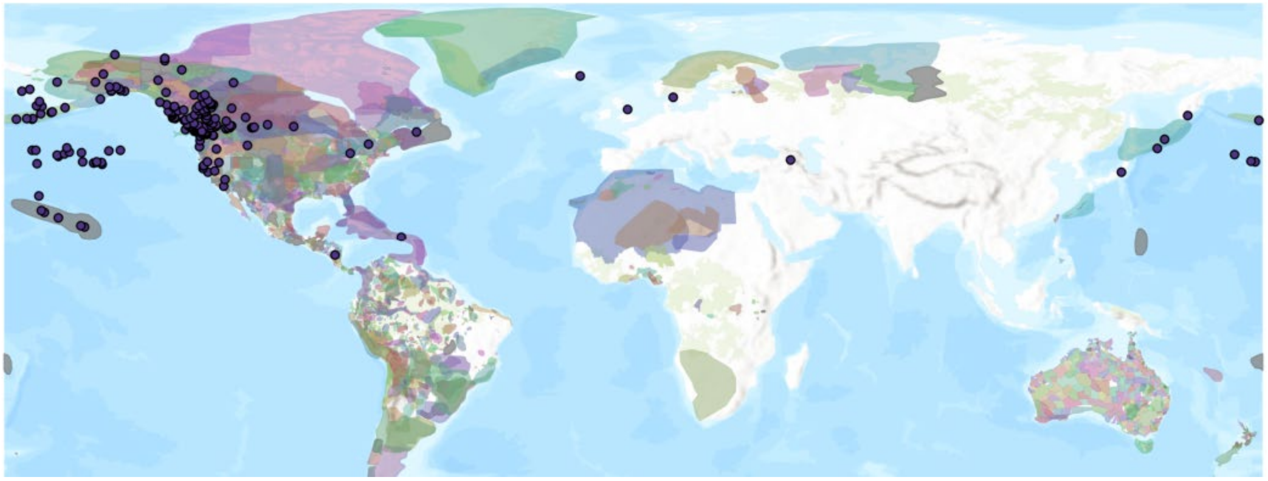


**Figure 17** Hakai Native Lands Digital API query for specimen locations based on manually derived coordinates from Mapcarta. Most specimens are from British Columbia and Washington State, as well as around the Aleutian Islands in Alaska.

The geocoder map has a separate facet interface[70] to visualize the specimen locations for Indigenous language areas using the query function as indicated in **Figure 18**. Each of the 58 language areas can be queried individually to see which specimens were collected there. When users click on a pin as shown in **Figure 19**, this reveals specimen data in greater detail and a field 'IndigenousName' where animal names from the Indigenous language area could be inserted. In this example, a pacific white sided dolphin specimen was collected from the Kwakwala Indigenous language area off northern Vancouver Island where the Indigenous name for dolphin is hatsawe'[71]

[70] https://zooarch.server.hakai.app/opencage-geocoded-native-lands/native-lands-language-facet-only?_facet_size=max

[71] https://www.firstvoices.com/kwakwala/search?q=dolphin&domain=both&types=word%2Cphrase%2Csong%2Cstory&sort=

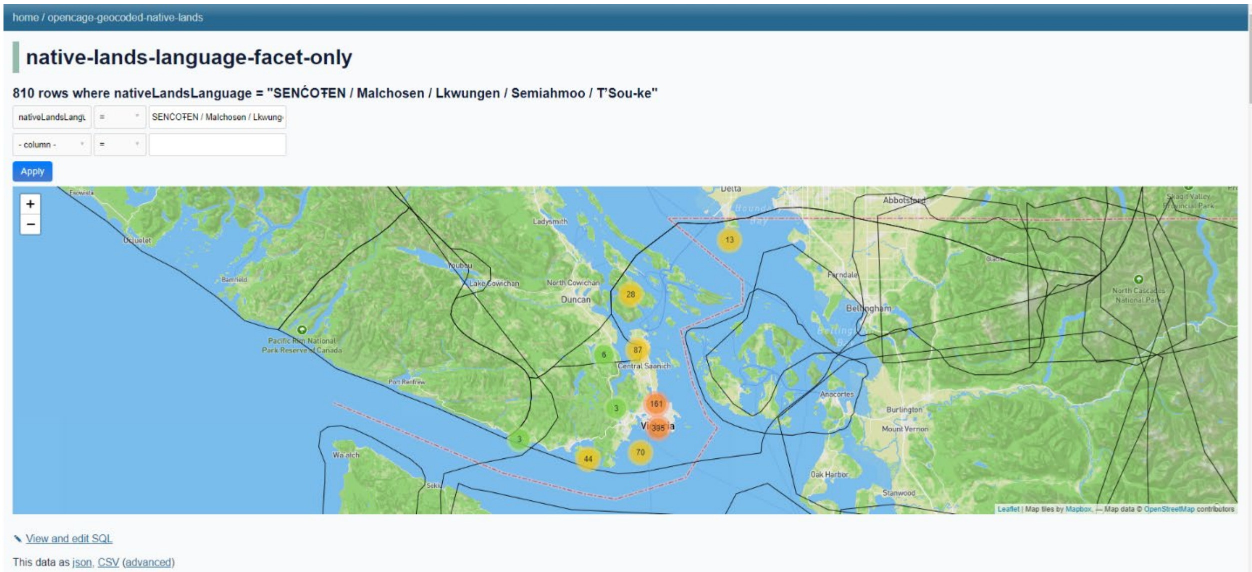



**Figure 18** An example of specimen numbers within a specific Indigenous language area.

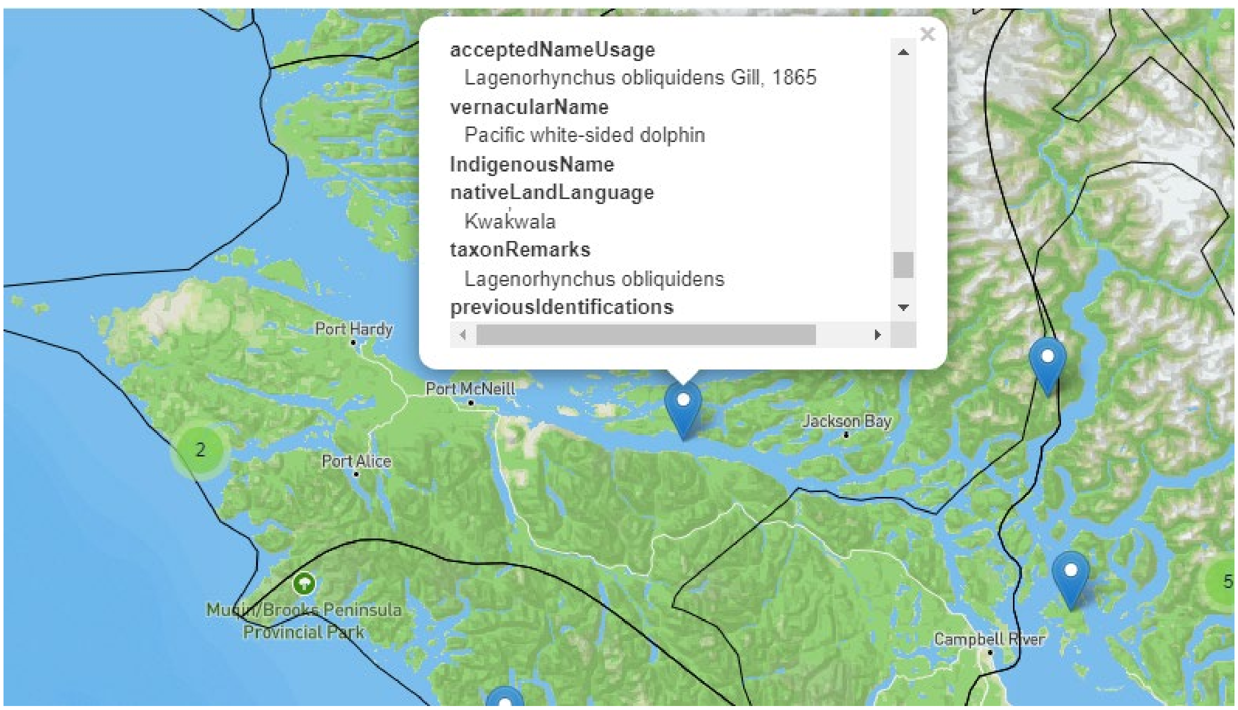



**Figure 19** Information window shows detail for specimen collected from this location. In this example the field 'IndigenousName' is shown as an example where the animal name, "hatsawe'"[13], from the Indigenous language area Kwak̓wala could be inserted.

By refining a query for specimens within a specific language area, a detailed listing of each animal can be downloaded as a CSV spreadsheet file. For example, by searching for animal specimens in the X̱aad Kil; X̱aaydaa Kil (Haida) language area, ten specimens are displayed on a map (**Figure 20**). The downloadable file generated by this query can be further refined to fields of interest. In this example, occurrenceID, scientificName, vernacularName, and nativeLandLanguage list the animals collected in this area. Animal names in the X̱aad Kil; X̱aaydaa Kil (Haida) language are added to the generated data (**Table 11**).


**Figure 20** X̱aad Kil / X̱aaydaa Kil (Haida) language area showing ten specimens from the native language facet map including an isolated specimen on the north end (blue teardrop).

**Table 11** X̱aad Kil / X̱aaydaa Kil (Haida) animal names from FirstVoices and Darwin Core terminology (occurrenceID, scientificName, and vernacularName) of the ten specimens shown in the X̱aad Kil / X̱aaydaa Kil (Haida) language area from Figure 20

| occurrenceID | X̱aad Kil / X̱aaydaa Kil (Haida) name[72] | scientificName | vernacularName |
|---|---|---|---|
| UVICZL-92.20 | daw gaatl'lxagaa.nga | *Buteo lagopus* (Pontoppidan, 1763) | Rough-legged hawk |
| UVICZL-92.21 | sgil ḵuuna.nga | *Melanitta fusca* (Linnaeus, 1758) | White-winged scoter |
| UVICZL-90.69 | iinang | *Clupea pallasii* Valenciennes, 1847 | Pacific Herring |
| UVICZL-84.46 | n/a | *Podothecus accipenserinus* (Tilesius, 1813) | Sturgeon Poacher |
| UVICZL-84.143 | k'aad | *Odocoileus hemionus* Rafinesque, 1817 | Sitka blacktail |
| UVICZL-03.29 | taan | *Ursus americanus* Pallas, 1780 | Black bear |
| UVICZL-02.29 | taan | *Ursus americanus* Pallas, 1780 | Black bear |
| UVICZL-02.28 | taan | *Ursus americanus* Pallas, 1780 | Black bear |
| UVICZL-00.64 | jiigwal awga | *Sorex vagrans* Osgood, 1901 | Wandering shrew |
| UVICZL-18.105 | sḵaats'ixuu | *Lottia alveus* (Conrad, 1831) | Pacific eelgrass limpet |

[72] https://www.firstvoices.com/hlgaagilda-xaayda-kil/

## 3.5 Discussion

### Animal taxonomies and Indigenous nomenclature

*Ethnobiology, Vernacular names, and Darwin Core extensions*

Current Western scientific practice focuses on taxonomy and prescribes Latin binomial naming conventions for species. This practice and the common English names given to animals has striking parallels to European colonization and toponyms, where declaring the "discovery" of "new" places [species] dictated the erasure and replacement of Indigenous placenames with European names (Trisos et al. 2021). Like conventional place names, scientific nomenclature rarely reflects local knowledge or the embedded history and landscape of a particular species (Gillman and Wright 2020, Franco 2021). Colonial "root taxonomies conceive of difference typologically and ignore questions of history" (Stahl 2020 p. 41). Whereas, Indigenous taxonomies encode ecological, morphological, utility value, and oral history knowledge in animal names that work to revitalize language and promote Indigenous food sovereignty (Gomes 2020, Franco 2021).

For the first time, the geocoordinates and resulting map produced provides a visual representation of the UVicZL regional collection spanning the Northwest Coast of North America and into the North Pacific. Using the Indigenous language areas derived from geolocation data, a future goal for the UVicZL collection will be to engage communities in sharing and linking Indigenous animal names to collection specimens as well as open-source audio clips of Indigenous animal names shared by community initiatives. Using the query function described in the results, such changes make it possible to open dialogues with Indigenous communities about some of the animals found in their territories. Indigenous animal names can be added to specimen records using this approach as seen in the example in **Table 11**. The DwC formatted dataset allows for multiple names under the "vernacular name" idiom which enables supporting multiple Indigenous nomenclatures across numerous Indigenous languages. The DwC vernacular names extension supports the description of common names that might be related to a species described in the main occurrence data file. This extension describes such properties as the language, location, and source of the common name to be recorded along with the publication of multiple common names for a single taxon as distinct rows in the extension file. Metadata allows for descriptions of "vernacular names", in this instance, the Indigenous language area, the language(s) spoken there, and potential animal names and sources of that shared information. In this regard, the ethical practice of recognizing data creators is fully documented as prescribed by OCAP®, FAIR and CARE principles.

*Local Contexts project*

In addition, the UVicZL collection has been registered as a project with Local Contexts[73], a global initiative that supports Indigenous communities with digital tools that can reassert cultural authority in heritage collections with machine readable data. A digital Open to Collaborate Engagement Notice (**Figure 21**) and a Biocultural Disclosure Notice (**Figure 22**) have been attached to the collection dataset and will be visible on the website we are developing. The Open to Collaborate Notice reaffirms our commitment to Indigenous collaboration, partnership, and engagement. The Biocultural Disclosure Notice recognizes the rights of Indigenous peoples to permission the use of information, collections, and data generated from the biodiversity resources associated with traditional lands, waters, and territories. The notice provides additional

[73] https://localcontextshub.org/projects/1c898f97-fa37-435b-843a-34752510e0a8

discoverability to facilitate Indigenous community application of Biocultural Labels to the collection as it progresses digitally. These labels are a viable avenue to pursue considering the challenging task of sourcing and verifying animal names from multiple nations where specimens have been collected. Local Contexts labels may also provide additional discoverability to Indigenous communities and scholars. These labels support CARE practices that necessitate the defining and recording of Indigenous providence and cultural metadata. Metadata about knowledge held by Indigenous peoples should also follow community protocols to define access and control (Farnel 2018). As the sharing of Indigenous data can conflict with the protection of this data, CARE practices and the use of digital labels can be avenues of restoring relationships and confirming cultural metadata. Local Contexts labels, which predate FAIR and CARE, are digital tags that support the application of these principles (Carroll et al. 2021).

**Figure 21** Open to Collaborate Notice from Local Contexts


**Figure 22** Biocultural Notice from Local Contexts

**Open collections with Indigenous language areas and animal names**

My research, adopting aspects of OCAP®, FAIR, and CARE Principles, assembled geospatial coordinates for over two thousand animal specimens in the UVicZL collection. This enables a specific connection to a crucial Indigenous aspect of data related to the collection: Indigenous language areas. This considerable effort represents a step towards improving access to Indigenous communities and scholars by creating a list of specimens across 58 Indigenous language areas and many more dialects. Making use of Indigenous data and metadata available through the open-source Indigenous led initiative, Native Land Digital, I was able to explore and identify which specimens in the UVicZL collection are associated with Indigenous language areas and build a foundation for community driven inclusion of Indigenous animal names. The associated transparency and accountability are steps toward decolonizing Indigenous data held by institutions. As recognized by OCAP®, Indigenous Nations have inherent rights to assert control and ownership of the data pertaining to them and the rights to identify, manage, and enhance the well-being of their communities. Specimens in the comparative collection are from First Nations territories and can have their origin recognized using Indigenous language areas and eventually nomenclature. As language is an essential aspect of Indigenous identity, Indigenous nomenclature requires development of appropriate metadata that has been accredited by Indigenous peoples who reside in those areas. Registering the project with Local Contexts can enable this enrichment of specimen data records. Additionally, if Indigenous language audio files are open sourced or directly provided by communities, these can now be linked as media files to specimen records. An approach that seeks the open collaboration of Indigenous communities rather than non-Indigenous compilation of information returns inherent data sovereignty to data creators.

The accessibility and openness of the UVicZL collection contributes to improving archaeological practice and decolonization by recognizing specimen collection locations and creates a space for over 2,400 specimens and 600 species to be linked across 58 Indigenous languages animal names. This work seeks to interrupt the colonial practice of depicting Indigenous peoples as a vestige of the past, including how the discipline of anthropology has been complicit in aspects of erasure (Wolf 1982). Naming practices matter as do the historical processes that gave rise to their enactment are embedded within them. “An effective archaeology pays attention to how our archives and narratives operate in socially persuasive ways to reproduce and bolster particular ways of imagining pasts and authorizing futures” (Stahl 2020 p. 42). By historicizing our knowledge production practices, such as taxonomies, we challenge the tendency for analytical facilities such as the UVicZL to have “historical amnesia” (ibid p. 42). Inclusion of Indigenous languages and potential for animal names is educational and invokes an environmental presence in contemporary resource management for the communities where archaeological research is conducted. It also provides a foundation for future participation in how Indigenous animal names are used in archaeology including how these data are collected, stored, analyzed, curated, and used in the future (Gupta et al. 2023). Additionally, acknowledging Indigenous languages with location data can indicate the regional significance and specimen origin of animals and their Indigenous names. The now open UVicZL collection developed with open science principles provides a practical example of an “ecology of knowledges” approach, where Western taxonomies and Indigenous nomenclature can be connected in an emergent and liminal digital space. Indigenous communities have the potential to connect to this list of specimens and species collected from their landscapes and seascapes to make meaningful contributions to regional knowledge while stimulating learning of language and of the animals within their territories. Similarly, archaeologists can broaden their scope of disciplinary knowledge dissemination and improve ethical practice and methods by considering and integrating CARE into their workflows. By acknowledging the Indigenous languages and the potential to include Indigenous animal names, practitioners might better recognize the long-held relationships Indigenous peoples have animals. This research seeks to open and broaden a dialogue between open science and Indigenous knowledge. Including Indigenous languages alongside English verbatim and Linnean taxonomic names assists the broader process of decolonizing and addressing incommensurability of knowledge and creating new relationships with Indigenous peoples by developing new modes of collaboration, engagement, and potential partnership, particularly involving the legacy of data and collections originating on Indigenous land that encompass both colonial and precolonial histories.

# Chapter 4: Conclusions

Zooarchaeological comparative collection labs primarily act as morphological identification facilities supporting archaeological inquiry. These resources are the basis for understanding human animal interactions from the deep past and serve as regional biodiversity repositories. As such, they cross disciplinary boundaries advancing the development of baselines which can better assess past ecosystems, improve contemporary resource management, and aid in supporting Indigenous food sovereignty. My thesis has developed methods and tools to reveal the intrinsic knowledge held in collections and make it widely available using open science and ethical practices. In this chapter I review the research questions I asked in Chapter 1 and summarize my answers from Chapter 2 and 3. I then explore what further steps could be taken for open collections and conclude this chapter with implications for zooarchaeological methods, and Indigenous resource management, biodiversity, and conservation.

## 4.1 Review of Research Questions

In Chapter 2 I address my first question: "*Can open science and FAIR principles contribute to collection accessibility and broaden discoverability by innovating documentation methods and connecting with diverse audiences?*" I explore avenues to improve access to zooarchaeology collections and lab facilities, issues arising from collection arrangement, and ways to better present collection contents. I implement the "extended specimen' concept to enrich data with linked open data, Darwin Core (DwC) terminology, open source ontologies, and visual displays. Developing this set of actions and tools enables me to address my second question: "*How can open science tools like linked open data (LOD), data standards, osteological and lifestage terminology, and methods to display geospatial, occurrence, and biometrics broaden collection* *specimen data and use across disciplines?*" My research required considerable effort to digitize physical records and learn and implement open science data standards. This effort transformed the capacity of the UVicZL collection, originally based on a card catalogue, single paper records, and excel lists with less than five data fields, to become a rich and well cited online dataset with ~320,000 fields of FAIR data fully integrated among tens of thousands of global biodiversity datasets accessible on GBIF. My research has ensured that the UVicZL zooarchaeology reference collection now has publicly available and transparent granular details and documents, including a map of the lab, and accessible collection composition data that can be used to efficiently sort and identify a wide range of collection information which was not possible previously. This includes the number of species and searchable taxonomic categories, many of which had species name changes since the specimen collection and cataloguing dates. For instance, the collection now has an accurate count of known specimens (2,922), species (671), and also lists critical conservation information such as the IUCN Red List status. Also achieved is the ability to categorize the total number of complete and partial specimens and those with sex and lifestage data as researchers can now search for specimens that have elements listed, such as skulls or scales, or refine by sex and age. The use of vocabularies and linked data has improved collection interoperability and has expanded its future capacity for including more refined specimen osteological data such as specimen bone counts and both disciplinary and universal terminologies. Some specimens have links to 3D models and additional links can be generated as specimens are better inventoried, photographed, or scanned providing yet another visual resource to researchers unable to be physically present in the lab. As the nature of interdisciplinary research is rapidly changing archaeological collections and repositories can benefit from open ontologies as this accelerates the capacity to talk across disciplines.

The aggregation of primary specimen data into searchable and viewable results, including the original catalogue card image, provides detailed information for comparative archaeological specimen identification. Taxonomic links allow the collection to be searched by TSN from up to four sources including GBIF, ITIS, NCBI, and FishBase. UVicZl's visibility as a natural history collection and resource outside of anthropology and archaeology has been greatly expanded by publication in GBIF with 53 citation results to date. The search function, visualization, and enabled integration with other natural history collections discoverable through publication on GBIF and GRSciColl transforms the use of the collection from a small group of frequent users to a diverse audience of global researchers. Known users and discoverability to future users have been greatly expanded from a handful of zooarchaeologists in western North America to global biodiversity researchers in the ecology, biology, fishery science, and conservation fields. The momentum and interest in natural history collections adopting open science practices is increasing as historical records of occurrences and distribution provide vital information in unprecedented times where climate change, reconciliation, and ecosystem management require urgent solutions.

Additionally, this primary data includes information such as weights, donors, and collection locations. The geographic component of the specimen data is available as coordinate data and through repository and web content visualization which provides a regional geography of collection content. The process of using open science platforms which enable geospatial mapping by deriving occurrence coordinate data provided an excellent transdisciplinary learning opportunity and created a pathway for future collection geospatial refinement. In Chapter 3, I acknowledge historical inequities, look for solutions to incommensurable knowledge, and geocode specimen data to enable locations to be categorized with Indigenous language areas.

Here, I answer my third question, "How can collections adopt CARE principles and connect with Indigenous histories and communities to foster more equitable sharing and reuse of data to positively impact communities and the histories zooarchaeologists study in the present and future?" Indigenous languages are rich sources of biological knowledge that span millennia where animals, landscapes, and people are viewed as integral relationships. Indigenous-led open science projects such as Native Land Digital and Local Contexts connect the archival specimen data in the UVicZL collection with Indigenous languages to extend the knowledge of animals outside the Western paradigm and broaden the history and geography of British Columbia and the Pacific Northwest. This connection provides a venue for an "ecology of knowledge" to empower Indigenous communities to enrich scientific research. The knowledge generated can be used to revitalize Indigenous-led resource management systems and improve conservation efforts in regional and local ecosystems. Linguistic information about Indigenous animal names is spread across diverse sources, online portals, apps, and community dictionaries (e.g., First Voices). The UVicZL collection of 671 species spanning 58 Indigenous language areas is well positioned to become a foundational platform for recognizing this diverse range of ethnobiological knowledge and can facilitate researcher awareness, public education, and assist with Indigenous language revitalization.

## 4.2 Further Steps for Open Collections

This project demonstrates the potential to have archives such as reference collections, which are housed at many Canadian universities and internationally, to become better recognized by adopting open science practices to increase collection discoverability and accessibility. By increasing awareness and recognition of these collections, researchers can help to alleviate invisibility problems. Institutions and collections managers should be encouraged and supported to curate their collection data digitally, make collection composition and region known, and provide zooarchaeological directories with links to their collection data. Ideally, collection curation should involve steps to encompass the entire data lifecycle. Recording specific trait attributes, collection location including geocoordinates, and scientific and common nomenclature aids in the specimen extension process. Open collections that foster data annotation practices, contain "extended specimens". These specimens, in addition to trait attributes include taxonomy, imaging, research related to specimen identification, geolocation and regional data, local knowledge that support Indigenous communities such as Indigenous language, and biometric data. Adopting biodiversity standards such as the Darwin Core vocabulary allows for seamless publishing and promotes interoperability. Collections can be uploaded to multiple open access repositories expanding their influence across disciplines. Publishing facilitates creation of DOIs allowing collections and specimens to be explicitly publicly acknowledged and cited in faunal reports and research papers. As the digital collections domain continues to evolve, solutions to persistent identifiers for individual specimens will also emerge. The impact of comparative collections can be quantified through citation tracking. Digitized collections assigned a DOI, Google Scholar profile, or ORCID identifier[74] can enable this as can tracking by repositories in which they are published.

## 4.3 Implications

### Zooarchaeological methods

A key step in zooarchaeology is element and taxon identification. Interpretive analyses are impossible without first identifying archaeofaunal assemblages, as they rely on the data from this

[74] https://orcid.org/0000-0001-8282-7957

step. Identification practices infer analogies of relation and form to make biological inferences. Zooarchaeologists using contemporary comparative collection specimens make assumptions about archaeological specimens having similar form and function; therefore, identification practice standards deserve greater attention and are related to larger implications for science (Wolverton 2013). Indeed, varied identification practices and high rates of inter-analyst error contribute to the replicability crisis. Validity in identification depends on the practices of recording, describing, and sharing methods and data. Zooarchaeologists can only improve identification practices by providing methods and data that are available for critique and reuse. Practices that describe study limitations, explain inferences and assumptions, and clarify interpretation support improvements in standards. Faunal analysists should record collection methods, screen sizes, sampling techniques, identification protocols, and comparative collection use to reduce knowledge gaps. Beginning with recovery, procedures and sampling strategies employed have a direct effect on understanding archaeofaunal assemblage composition. For example, documenting smaller screen sizes that will recover taxa that would not have been represented otherwise, should be obligatory. Further documenting the sorting of remains, and the procedures for element and taxon identification, including comparative collection, individual specimen, and anatomical attributes, reduce bias and lend credibility to identity replication. Methods should describe detailed diagnostic criteria and definitive skeletal traits, and include morphometric attribute variability for independent evaluation (Lyman 2019). Developing open access, open source, trait identification databases that include taxon, skeletal element, diagnostic anatomical trait, and reference collection specimen data would improve archaeological assemblage identification methods (ibid).

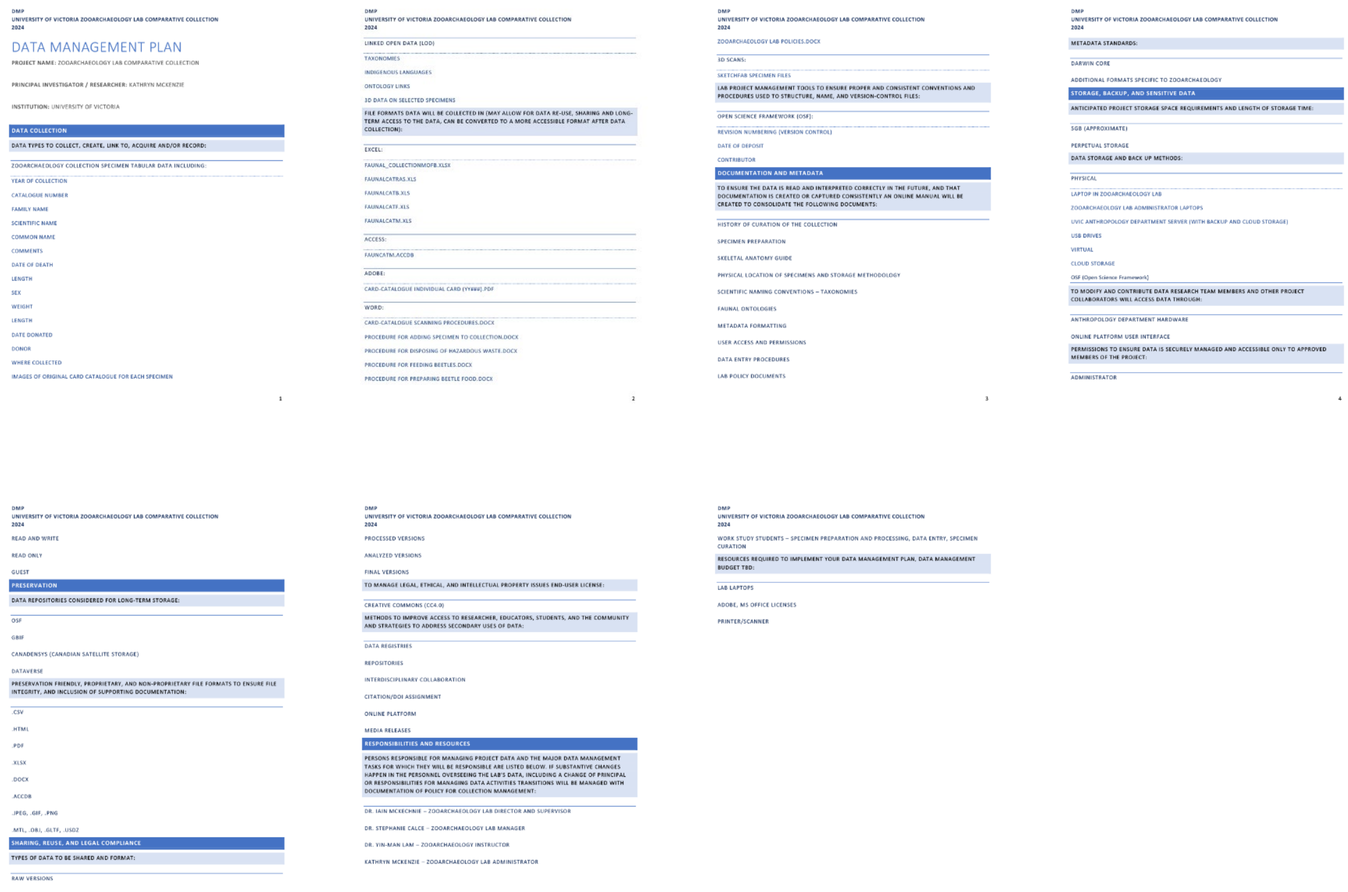
While there can never be absolute certainty about assemblage composition and the ecological implications associated with archaeofaunal remains, zooarchaeologists can increase confidence with thorough recording of identification methods and the collections and assumptions on which taxonomic assignments are based. Open science provides opportunities to facilitate documentation and improve data access, thereby improving standardized methods and taxonomic identification. Open access data journals, data aggregators, and research workflow platforms are tools available to zooarchaeologists for this purpose. Data use and reuse is encouraged through open data journals where zooarchaeologists can publish and access research data and supplemental methods. Analysts should acknowledge and cite collections and individual specimens used in their research and publishers should require collection data attribution. This practice improves "experimental replicability" and educates users about the identification process employed by faunal specialists (Lyman 2010, Gifford-Gonzalez 2018).

**Indigenous resource management, biodiversity, and conservation**

Indigenous marine and terrestrial management systems incorporated transferred knowledge and were developed over time to respond to and create coastal dynamism. (Mathews and Turner 2017). These ecosystems, culturally and ecologically linked together, were managed through accumulated generational knowledge which became integral to Indigenous worldviews and the enduring maintenance of coastal ecosystems (ibid). The spiritual and economic complexity of NWC Indigenous peoples reflected by these systems are becoming more apparent as detrimental and unsustainable colonial systems continue to harm the environment and contribute to climate change. Colonial destruction and restriction of Indigenous lifeways have inhibited their ability to pass on key knowledge, assure food sovereignty, and maintain their identity (Mathews and Turner 2017, Atlas et al. 2021). In many regions a disruption in ecological balance and biodiversity has occurred in absence of Indigenous management (Bliege Bird and Nimmo 2018). Further archaeological research and use of technology in collaboration with Indigenous communities may help to fill in these gaps in knowledge and work to restore ancient coastal ecosystems (Mathews and Turner 2017). Relationship building and co-management working toward system recovery and reestablishment are critical steps in the ability of communities to reclaim control over their identity, food sovereignty, and improve species biodiversity (Mathews and Turner 2017, Atlas et al. 2021). Zooarchaeological research and comparative collection preservation, digitization, and the continuing acquisition of specimens as historical archives enables biodiversity baseline data research. Establishing regional baseline data derived from archaeological sites and the animals found within them can be influential to resource policy decisions (Frazier 2007, Lyman 1996). Sharing data for compilation and reuse creates opportunities to reevaluate and understand changes in biodiversity with humans as active participants.

# Appendix: Supplementary Information for Chapter 2

## 5.1 Data Management Plan



DMP
UNIVERSITY OF VICTORIA ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION
2024

DATA MANAGEMENT PLAN

PROJECT NAME: ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION

PRINCIPAL INVESTIGATOR / RESEARCHER: KATHRYN MCKENZIE

INSTITUTION: UNIVERSITY OF VICTORIA

DATA COLLECTION

DATA TYPES TO COLLECT, CREATE, LINK TO, ACQUIRE AND/OR RECORD:

ZOOARCHAEOLOGY COLLECTION SPECIMEN TABULAR DATA INCLUDING:

YEAR OF COLLECTION
CATALOGUE NUMBER
FAMILY NAME
SCIENTIFIC NAME
COMMON NAME
COMMENTS
DATE OF DEATH
LENGTH
SEX
WEIGHT
LENGTH
DATE DONATED
DONOR
WHERE COLLECTED
IMAGES OF ORIGINAL CARD CATALOGUE FOR EACH SPECIMEN

1

DMP
UNIVERSITY OF VICTORIA ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION
2024

LINKED OPEN DATA (LOD)

TAXONOMIES
INDIGENOUS LANGUAGES
ONTOLOGY LINKS
3D DATA ON SELECTED SPECIMENS

FILE FORMATS DATA WILL BE COLLECTED IN (MAY ALLOW FOR DATA RE-USE, SHARING AND LONG-TERM ACCESS TO THE DATA, CAN BE CONVERTED TO A MORE ACCESSIBLE FORMAT AFTER DATA COLLECTION):

EXCEL:

FAUNAL_COLLECTIONMOFB.XLSX
FAUNALCATRAS.XLS
FAUNALCATB.XLS
FAUNALCATF.XLS
FAUNALCATM.XLS

ACCESS:

FAUNCATM.ACCDB

ADOBE:

CARD-CATALOGUE INDIVIDUAL CARD (YY###).PDF

WORD:

CARD-CATALOGUE SCANNING PROCEDURES.DOCX
PROCEDURE FOR ADDING SPECIMEN TO COLLECTION.DOCX
PROCEDURE FOR DISPOSING OF HAZARDOUS WASTE.DOCX
PROCEDURE FOR FEEDING BEETLES.DOCX
PROCEDURE FOR PREPARING BEETLE FOOD.DOCX

2

DMP
UNIVERSITY OF VICTORIA ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION
2024

ZOOARCHAEOLOGY LAB POLICIES.DOCX

3D SCANS:

SKETCHFAB SPECIMEN FILES

LAB PROJECT MANAGEMENT TOOLS TO ENSURE PROPER AND CONSISTENT CONVENTIONS AND PROCEDURES USED TO STRUCTURE, NAME, AND VERSION-CONTROL FILES:

OPEN SCIENCE FRAMEWORK (OSF):

REVISION NUMBERING (VERSION CONTROL)
DATE OF DEPOSIT
CONTRIBUTOR

DOCUMENTATION AND METADATA

TO ENSURE THE DATA IS READ AND INTERPRETED CORRECTLY IN THE FUTURE, AND THAT DOCUMENTATION IS CREATED OR CAPTURED CONSISTENTLY AN ONLINE MANUAL WILL BE CREATED TO CONSOLIDATE THE FOLLOWING DOCUMENTS:

HISTORY OF CURATION OF THE COLLECTION
SPECIMEN PREPARATION
SKELETAL ANATOMY GUIDE
PHYSICAL LOCATION OF SPECIMENS AND STORAGE METHODOLOGY
SCIENTIFIC NAMING CONVENTIONS – TAXONOMIES
FAUNAL ONTOLOGIES
METADATA FORMATTING
USER ACCESS AND PERMISSIONS
DATA ENTRY PROCEDURES
LAB POLICY DOCUMENTS

3

DMP
UNIVERSITY OF VICTORIA ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION
2024

METADATA STANDARDS:

DARWIN CORE
ADDITIONAL FORMATS SPECIFIC TO ZOOARCHAEOLOGY

STORAGE, BACKUP, AND SENSITIVE DATA

ANTICIPATED PROJECT STORAGE SPACE REQUIREMENTS AND LENGTH OF STORAGE TIME:

5GB (APPROXIMATE)
PERPETUAL STORAGE

DATA STORAGE AND BACK UP METHODS:

PHYSICAL

LAPTOP IN ZOOARCHAEOLOGY LAB
ZOOARCHAEOLOGY LAB ADMINISTRATOR LAPTOPS
UVIC ANTHROPOLOGY DEPARTMENT SERVER (WITH BACKUP AND CLOUD STORAGE)
USB DRIVES
VIRTUAL
CLOUD STORAGE
OSF (Open Science Framework)

TO MODIFY AND CONTRIBUTE DATA RESEARCH TEAM MEMBERS AND OTHER PROJECT COLLABORATORS WILL ACCESS DATA THROUGH:

ANTHROPOLOGY DEPARTMENT HARDWARE
ONLINE PLATFORM USER INTERFACE

PERMISSIONS TO ENSURE DATA IS SECURELY MANAGED AND ACCESSIBLE ONLY TO APPROVED MEMBERS OF THE PROJECT:

ADMINISTRATOR

4

DMP
UNIVERSITY OF VICTORIA ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION
2024

READ AND WRITE
READ ONLY
GUEST

PRESERVATION

DATA REPOSITORIES CONSIDERED FOR LONG-TERM STORAGE:

OSF
GBIF
CANADENSYS (CANADIAN SATELLITE STORAGE)
DATAVERSE

PRESERVATION FRIENDLY, PROPRIETARY, AND NON-PROPRIETARY FILE FORMATS TO ENSURE FILE INTEGRITY, AND INCLUSION OF SUPPORTING DOCUMENTATION:

.CSV
.HTML
.PDF
.XLSX
.DOCX
.ACCDB
.JPEG, .GIF, .PNG
.MTL, .OBJ, .GLTF, .USDZ

SHARING, REUSE, AND LEGAL COMPLIANCE

TYPES OF DATA TO BE SHARED AND FORMAT:

RAW VERSIONS

5

DMP
UNIVERSITY OF VICTORIA ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION
2024

PROCESSED VERSIONS
ANALYZED VERSIONS
FINAL VERSIONS

TO MANAGE LEGAL, ETHICAL, AND INTELLECTUAL PROPERTY ISSUES END-USER LICENSE:

CREATIVE COMMONS (CC4.0)

METHODS TO IMPROVE ACCESS TO RESEARCHER, EDUCATORS, STUDENTS, AND THE COMMUNITY AND STRATEGIES TO ADDRESS SECONDARY USES OF DATA:

DATA REGISTRIES
REPOSITORIES
INTERDISCIPLINARY COLLABORATION
CITATION/DOI ASSIGNMENT
ONLINE PLATFORM
MEDIA RELEASES

RESPONSIBILITIES AND RESOURCES

PERSONS RESPONSIBLE FOR MANAGING PROJECT DATA AND THE MAJOR DATA MANAGEMENT TASKS FOR WHICH THEY WILL BE RESPONSIBLE ARE LISTED BELOW. IF SUBSTANTIVE CHANGES HAPPEN IN THE PERSONNEL OVERSEEING THE LAB'S DATA, INCLUDING A CHANGE OF PRINCIPAL OR RESPONSIBILITIES FOR MANAGING DATA ACTIVITIES TRANSITIONS WILL BE MANAGED WITH DOCUMENTATION OF POLICY FOR COLLECTION MANAGEMENT:

DR. IAIN MCKECHNIE – ZOOARCHAEOLOGY LAB DIRECTOR AND SUPERVISOR
DR. STEPHANIE CALCE – ZOOARCHAEOLOGY LAB MANAGER
DR. YIN-MAN LAM – ZOOARCHAEOLOGY INSTRUCTOR
KATHRYN MCKENZIE – ZOOARCHAEOLOGY LAB ADMINISTRATOR

6

DMP
UNIVERSITY OF VICTORIA ZOOARCHAEOLOGY LAB COMPARATIVE COLLECTION
2024

WORK STUDY STUDENTS – SPECIMEN PREPARATION AND PROCESSING, DATA ENTRY, SPECIMEN CURATION

RESOURCES REQUIRED TO IMPLEMENT YOUR DATA MANAGEMENT PLAN, DATA MANAGEMENT BUDGET TBD:

LAB LAPTOPS
ADOBE, MS OFFICE LICENSES
PRINTER/SCANNER

7

**Figure 23** Data Management Plan

## 5.2 R Code

```r
# Load required libraries
library(rgbif)
library(readxl)
library(writexl)
library(tidyverse)

# Set input file path
input_file <- "C:/Users/kmaca/OneDrive/Documents/Hakai/zooarch/data/ITIS.xlsx"

# Read the GBIF taxon ID numbers from the input Excel file
taxon_data <- read_excel(input_file)

#Make output path
output_file <- "C:/Users/kmaca/OneDrive/Documents/Hakai/zooarch/output_data/ITIS_modified.xlsx"

# Function to get ITIS TSN based on GBIF taxon ID
get_tsn <- function(TaxonName) {
  # Query ITIS using rgbif package
  query <- rgbif::name_usage(name = TaxonName, datasetKey = "9ca92552-f23a-41a8-a140-01abaa31c931")[["data"]] #This is the GBIF datasetKey for ITIS

  # Extract the TSN if available
  if (nrow(query) == 1) {
    tsn <- query$taxonID
  } else {
    tsn <- NA
  }

  return(tsn)
}

# Apply the get_tsn function to each GBIF taxon ID in the input data
taxon_data$TSN <- sapply(taxon_data$TaxonName, get_tsn)

# Write the modified data to the output Excel file
write_xlsx(taxon_data, output_file)
```

**Figure 24** GBIF ITIS TSN Rscript written to pull ITIS numbers from the GBIF API

## 5.3 Darwin Core Terms

**Table 12** Darwin Core terms, metadata description, and examples

| Term | Description | Example |
|---|---|---|
| type | The nature or genre of the resource | PhysicalObject |
| language | A language of the resource | en |
| license | A legal document giving official permission to do something with the resource | http://creativecommons.org/licenses/by/4.0/legalcode |
| rightsholder | A person or organization owning or managing rights over the resource | The Anthropology Department of the University of Victoria |
| accessRights | Information about who can access the resource or an indication of its security status | |
| bibliographicCitation | A bibliographic reference for the resource as a statement indicating how this record should be cited (attributed) when used | University of Victoria, Zooarchaeology Lab Collection, UVICZL-16.1 |
| collectionID | For physical specimens, the recommended best practice is to use an identifier from a collections registry such as the Global Registry of Biodiversity Repositories (http://grbio.org/) | https://www.gbif.org/grscicoll/collection/1cbebd16-1e98-417a-b1fe-531208312892 |
| institutionCode | The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record | UVIC |
| institutionID | An identifier for the institution having custody of the object(s) or information referred to in the record | http://grscicoll.org/institution/university-victoria |
| collectionCode | The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived | UVICZL |
| datasetName | The name identifying the data set from which the record was derived | UVICZL Mammals |
| basisOfRecord | Recommended best practice is to use the standard label of one of the Darwin Core classes | PreservedSpecimen |
| dynamicProperties | A list of additional measurements, facts, characteristics, or assertions about the record. Meant to provide a mechanism for structured content. Recommended best practice is to use a key:value encoding schema for a data interchange format such as JSON | https://localcontextshub.org/projects/1c898f97-fa37-435b-843a-34752510e0a8 |
| occurenceIDURL | | https://www.gbif.org/occurrence/3329069886 |

| Term | Description | Example |
| --- | --- | --- |
| occurrenceID | An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique | UVICZL-01.47 |
| catalogNumber | An identifier (preferably unique) for the record within the data set or collection | 08.63 |
| recordedBy | A list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence. The primary collector or observer, especially one who applies a personal identifier (recordNumber), should be listed first | Becky Wigen |
| sex original | The sex of the biological individual(s) represented in the Occurrence | Sex recorded in original excel spreadsheet or catelogue card |
| Sex | The sex of the biological individual(s) represented in the Occurrence | Male tDAR Faunal Sex Ontology doi:10.6067/XCV87H1GP8 |
| lifeStage original | The age class or life stage of the biological individual(s) at the time the Occurrence was recorded | Lifestage recorded in original excel spreadsheet or catalogue card |
| lifeStage | The age class or life stage of the biological individual(s) at the time the Occurrence was recorded | Adult tDAR Faunal Age Ontology doi:10.6067/XCV8PZ574Q |
| preparations original | A list (concatenated and separated) of preparations and preservation methods for a specimen | skull \| skeleton \| scale \| shell |
| preparations | A list (concatenated and separated) of preparations and preservation methods for a specimen | skull tDAR Faunal Element Ontology doi:10.6067/XCV8HQ3XDT \| skull OLS Uber anatomy ontology http://purl.obolibrary.org/obo/UBERON_0003129 |
| disposition | The current state of a specimen with respect to the collection identified in collectionCode or collectionID | in collection, missing, loaned to |
| associatedMedia | A list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of media associated with the Occurrence | osf.io/catalogueCardGUID |
| associatedReferences | A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the Occurrence | http://www.nature.com/articles/s41598-020-71574-x |
| NCBIIDrefURL | A list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of genetic sequence information associated with the Occurrence | NCBI:txid9615 |
| otherCatalogNumbers | A list (concatenated and separated) of previous or alternate fully qualified catalog numbers or other human-used identifiers for the same Occurrence, whether in the current or any other data set or collection | 08/63 |
| occurrenceRemarks original | Comments or notes about the Occurrence | Remarks from recorded in original excel spreadsheet |

| Term | Description | Example |
| --- | --- | --- |
| occurrenceRemarks | Comments or notes about the Occurrence | from Lake Washington (freshwater), length: 17cm, weight: 152.5g, donated: L.S. Smith (Fall '88), processed: Sept. 19/89 |
| eventRemarks | Comments or notes about the Event | Date of death \| Date entered collection |
| eventDate | The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context | 2000-07-30 |
| verbatimEventDate | The verbatim original representation of the date and time information for an Event | 1995-07-30 |
| year | The four-digit year in which the Event occurred, according to the Common Era Calendar | 1995 |
| month | The integer month in which the Event occurred | 7 |
| day | The integer day of the month on which the Event occurred | 30 |
| continent | The name of the continent in which the Location occurs | North America |
| waterbody | The name of the water body in which the Location occurs | Pacific Ocean |
| islandGroup | The name of the island group in which the Location occurs | Broken Islands |
| Island | The name of the island on or near which the Location occurs | Vancouver Island |
| country | The name of the country or major administrative unit in which the Location occurs | Canada |
| countryCode | The standard code for the country in which the Location occurs | CA |
| stateProvince | The name of the next smaller administrative region than country (state, province, canton, department, region, etc.) in which the Location occurs | BC |
| municipality | The full, unabbreviated name of the next smaller administrative region than county (city, municipality, etc.) in which the Location occurs. Do not use this term for a nearby named place that does not contain the actual location | Saanich |
| locality | The specific description of the place. Less specific geographic information can be provided in other geographic terms (higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description | Long Beach |
| verbatimLocality | The original textual description of the place | Spectacle Lake, Vancouver Island, BC |
| locationRemarks | Comments or notes about the Location | north side of Diana Island |

| Term | Description | Example |
|---|---|---|
| verbatimCoordinates | The verbatim original spatial coordinates of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem. | 42º30'N 161º23'W |
| decimalLatitude | The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive. | -41.09 |
| decimalLongitude | The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive. | -121.17 |
| coordinateUncertaintyInMeters | The horizontal distance (in meters) from the given dwc:decimalLatitude and dwc:decimalLongitude describing the smallest circle containing the whole of the dcterms:Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term | 1000 |
| coordinatePrecision | A decimal representation of the precision of the coordinates given in the dwc:decimalLatitude and dwc:decimalLongitude | |
| georeferencedBy | A list (concatenated and separated) of names of people, groups, or organizations who determined the georeference (spatial representation) for the dcterms:Location | steve.vandervalk@hakai.org \| Hakai Institute of the Tula Foundation |
| georeferenceSources | A list (concatenated and separated) of maps, gazetteers, or other resources used to georeference the dcterms:Location, described specifically enough to allow anyone in the future to use the same resources | Ubaldino, M (MITRE Corporation), 'OpenSextant Xponents: Geotagging Toolkit for World-wide Geography', 2019. https://opensextant.github.io/Xponents/ |
| georeferenceRemarks | Notes or comments about the spatial description determination, explaining assumptions made in addition or opposition to the those formalized in the method referred to in dwc:georeferenceProtocol | |
| identifiedBy | A list (concatenated and separated) of names of people, groups, or organizations who assigned the Taxon to the subject | Becky Wigen |
| identificationRemarks | Comments or notes about the Identification | Identified by B. Wigen from BCPM skeletons mainly on basis of size |
| GBIFtaxonIDURL | GBIF taxon identifier link | https://www.gbif.org/species/2433458 |
| GBIFtaxonID | GBIF taxon identifier TSN | 2433458 |
| ITIStaxonID | ITIS taxon identifier TSN | 173783 |

| Term | Description | Example |
|---|---|---|
| taxonID | An identifier for the set of taxon information (data associated with the Taxon class). May be a global unique identifier or an identifier specific to the data set | ITIS TSN: 173783 |
| fishbaseIDURL | FishBase taxon identifier link | http://www.fishbase.se/summary/3813 |
| fishbaseID | FishBase taxon identifier TSN | 3813 |
| canonicalName | The verbatim scientific name | (genus + specificEpithet) Enophrys bison |
| scientificName | The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the IdentificationQualifier term | (genus + specificEpithet + scientificNameAuthorship) Enophrys bison (Girard, 1854) |
| kingdom | The full scientific name of the kingdom in which the taxon is classified | Animalia |
| phylum | The full scientific name of the phylum or division in which the taxon is classified | Chordata |
| class | The full scientific name of the class in which the taxon is classified | Mammalia |
| order | The full scientific name of the order in which the taxon is classified | Carnivora |
| family | The full scientific name of the family in which the taxon is classified | Canidae |
| genus | The full scientific name of the genus in which the taxon is classified | Canis |
| subgenus | The full scientific name of the subgenus in which the taxon is classified. Values should include the genus to avoid homonym confusion | |
| specificEpithet | The name of the first or species epithet of the scientificName | familiaris |
| infraspecificEpithet | The name of the lowest or terminal infraspecific epithet of the dwc:scientificName, excluding any rank designation | |
| taxonRank | The taxonomic rank of the most specific name in the scientificName | species |
| scientificNameAuthorship | The authorship information for the scientificName formatted according to the conventions of the applicable nomenclaturalCode | Linnaeus, 1758 |
| taxonomicStatus | The status of the use of the dwc:scientificName as a label for a taxon. Requires taxonomic opinion to define the scope of a dwc:Taxon. Rules of priority then are used to define the taxonomic status of the nomenclature contained in that scope, combined with the experts opinion. It must be linked to a specific taxonomic reference that defines the concept | accepted |

| Term | Description | Example |
|---|---|---|
| acceptedNameUsage | The full name, with authorship and date information if known, of the currently valid (zoological) or accepted (botanical) dwc:Taxon | Enophrys bison (Girard, 1854) |
| vernacularName | A common or vernacular name | Domestic dog |
| IndigenousName | Indigenous Language Name | |
| nativeLandLanguage | Indigenous Language Area | Coast Salish |
| taxonRemarks | Comments or notes about the taxon or name | recorded in catalogue card as "Canis sp." |
| previousIdentifications | A list (concatenated and separated) of previous assignments of names to the Organism | recorded in catalogue card as "Eumentopias jubata" |
| measurementRemarks | Comments or notes accompanying the MeasurementOrFact | tail missing |
| measurementType | The nature of the measurement, fact, characteristic, or assertion | Weight, Length |
| measurementUnit | The units associated with the measurementValue | g, cm, kg |
| measurementValue | The value of the measurement, fact, characteristic, or assertion | 375, 45, 2.5 |
| weight | Measurements from catalogue cards or original excel spreadsheet | 6.25g |
| length | Measurements from catalogue cards or original excel spreadsheet | 125cm |
| width | Measurements from catalogue cards or original excel spreadsheet | 55mm |
| totalLength | Measurements from catalogue cards or original excel spreadsheet | 22 |
| forkLength | Measurements from catalogue cards or original excel spreadsheet | 21 |
| standardLength | Measurements from catalogue cards or original excel spreadsheet | 19.5 |
| wing | Measurements from catalogue cards or original excel spreadsheet | |
| tarsus | Measurements from catalogue cards or original excel spreadsheet | |
| culmen | Measurements from catalogue cards or original excel spreadsheet | |
| diameter | Measurements from catalogue cards or original excel spreadsheet | |
| catalogueYearYy | catalogue card year (yy) | 04 |
| catalogueNumberNn | catalogue card number | 14 |
| catalogueYearYyyy | catalogue card year (yyyy) | 2004 |

| Term | Description | Example |
|---|---|---|
| | | non-DwC description |
| | | DwC description |

# A Compendium and Glossary of Terms for Open Collections

## 6.1 FAIR Principles

FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles define how data vocabularies, tools, resources and infrastructures should be characterized to facilitate data discoverability and reuse (Wilkinson et al. 2016). These principles enhance data reuse by facilitating improved data discovery through metadata standardization. The ease of data discovery and compatibility with other sources supports new downstream standalone research and the combining of primary datasets for meta-analyses (Kansa 2015, Curty et al. 2017, Faniel et al. 2018, Miller et al. 2020). Findable data is assigned globally unique and persistent identifiers, often termed as GUIDs or PURLs. A globally unique identifier is exclusive to your data and created by a registry service. These registry services guarantee the uniqueness of the identifier and the resolvability of the persistent link by actively maintaining web links. Datasets and their components deposited to repositories with this capability receive GUIDs which contain an internet link such as an URL that resolves to a webpage containing the data itself or its defined concept. Findability is enhanced by richly described metadata included in the identifier of the data they describe. Persistent identifiers use a free and open standardized communication protocol to aid accessibility as users can simply “click a link” to retrieve (meta)data. If datasets are no longer available or maintained online, their metadata should continue to be persistent as it remains valuable for future research planning and replication. Interoperable data uses language that has syntax and grammar precisely defined for ease of exchange and interpretation. An example of this language would be the Darwin Core (DwC) schema. These controlled vocabularies allow people or machines to find, access, interoperate, and reuse data. (Meta)data interoperability is also improved by linking (meta)data with contextual information as a reference. Information in other cited datasets or complimentary information enriches (meta)data. Expanding reusability dictates that (meta)data contain a ‘plurality’ of attributes by including as much accurate and relevant information as possible. A clear and accessible data use license also improves data reusability as does detailed provenance including how to cite it, and a workflow description in machine-readable format. Finally, for optimal reusability data should be contextual and follow best practices for the community or discipline it represents. This is especially true when the rights of Indigenous peoples conflict with open data practices. Non-indigenous institutions and researchers must look for data frameworks across the data life cycle that consider ecological, spiritual, and cultural protocols (Jennings et al. 2023).

**Table 13** FAIR Principles and Objectives. Modified from Australian Research Data Commons (ARDC) FAIR Data Self Assessment Tool(https://ardc.edu.au/resource/fair-data-self-assessment-tool/)

| FAIR Principle | Objectives |
|---|---|
| ***Findable*** | |
| *The dataset should have identifiers assigned* | Identifiers are essential for identifying, finding, retrieving, linking and citing datasets. A Web address (URL) can be used to specify the online location of a resource but over time URLs tend to change which leads to broken links to the data. To be useful, identifiers need to be persistent and unique. Digital Object Identifiers (DOIs) and other persistent identifiers (PIDs) provide a permanent citable reference to a particular dataset. The DOI is a permanent fixed reference to the dataset no matter where it is located online and enables citation and citation metrics.<br><br>Services to create a persistent identifier are often offered by your affiliated institution or the repository you are using to describe your data. |
| *The dataset identifier should be included in all metadata records/files describing the data* | The identifier (preferably a persistent identifier e.g., DOI PURL ARK or Handle) needs to be clearly stated in the metadata record describing the data collection, and also in any associated data files or metadata. |
| *The data should be described with metadata* | Comprehensive metadata records will include descriptive content that facilitates discovery, access and reuse of the data being described. While there is no 'one size fits all' list, comprehensive metadata should include:<br><br>• a globally unique persistent identifier e.g. a DOI<br>• a title<br>• related people, i.e. the data creator or custodian<br>• how to access the data and file formats<br>• a description of how the data were created and how to interpret the data subject or keywords<br>• citation information that clearly indicates how the data should be cited<br>• a machine-readable data licence<br>• provenance and contextual information such as:<br>    •<br>    ○ links to related publications, projects, services and software<br>    ○ methodology and processes involved in data production<br>• spatial and temporal coverage (if relevant)<br>• object-level data description<br><br>Providing metadata in a standard schema allows it to be read and used by machines as well as humans. |
| *The metadata record should be located in a repository or registry* | A rich metadata description alone does not ensure a dataset's 'findability' on the internet; the dataset needs to be registered or indexed in a searchable resource, such as a generalist, domain-specific, or institutional data repository or registry. Ideally these repositories/registries are indexed by search engines such as Google and/or Google Scholar. |
| ***Accessible*** | |
| *The data should be accessible* | Not all data that is discoverable can be freely accessed. Often there are embargoes, access controls, and access permissions associated with data due to a variety of issues such as privacy and commercial interests. Even with all these issues much sensitive data can be shared. Many other issues that could be perceived as blockers to sharing data may be overcome. |

| ***FAIR Principle*** | ***Objectives*** |
|---|---|
| *The data should be available online without requiring specialised protocols or tools once access has been approved* | Ideally users would like to retrieve appropriate internet content directly and unhindered once they have located it. Internet protocols (e.g. http and ftp) define rules and conventions for communication between devices, and tools and services are available to facilitate this process, e.g. APIs.<br><br>Protocols<br>HTTP (Hypertext Transfer Protocol) is the set of rules for transferring files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web. As soon as a Web user opens their Web browser, the user is indirectly making use of HTTP.<br>FTP (File Transfer Protocol) is a standard Internet protocol for transmitting files between computers on the Internet over TCP/IP connections.<br>API (Application Programming Interface): When information is made available in any machine readable format, it becomes possible to make that information directly available to programs that request that information over the web. An API is the way this information is made directly available to other machines. |
| *The metadata record should be available even if the data is no longer available* | Effort is often required to maintain data resources online which can often lead to it being neglected especially over long periods of time. This often leads to broken links between the metadata and the data. Having at least a description of the data in the form of a metadata record means there is a record of the data's existence allowing the possibility of someone tracking it down. |
| ***Interoperable*** | |
| *The data (file) format(s) should be interoperable* | When selecting file formats for archiving, the formats should ideally be:<br>Non-proprietary<br>Unencrypted<br>Uncompressed<br>In common usage by the research community<br>Adherent to an open, documented standard<br>• Interoperable among diverse platforms and applications<br>• Fully published and available royalty-free<br>• Fully and independently implementable by multiple software providers on multiple platforms without any intellectual property restrictions for necessary technology<br>• Developed and maintained by an open standards organization with a well-defined inclusive process for evolution of the standard.<br>File format examples:<br>Structured, open standard, machine-readable format e.g. (text) PDF/A, HTML, Plain text, (images) TIFF, JPEG 2000, GIF, (audio) MP3, AIFF, WAVE, (video) MOV, MPEG, AVI, (Tabular data) CSV<br>structured, open standard, non-machine-readable format, e.g. PDF, HTML, JPG<br>Proprietary format, e.g. doc (Word), .xls (Excel), .ppt (PowerPoint) |
| *The data elements should be defined by interoperable vocabularies/ontologies/tagging schemas* | Schema standards are schemas that have gone through a formal validation process by a standards organisation, such as the International Standards Organisation (ISO) or an equivalent body such as the Dublin Core Metadata Initiative (DCMI); or, commonly used and consistently applied metadata schemas that are well documented, endorsed, and maintained can also become 'de-facto standards'.<br>Resolvable global identifiers - An identifier is any label used to name something uniquely (whether online or offline). URLs are an example of an identifier. So are serial numbers, and personal names. A persistent identifier is guaranteed to be managed and kept up to date over a defined time period. |

| ***FAIR Principle*** | ***Objectives*** |
|---|---|
| *The metadata should be linked to other data and metadata (to enhance context and clearly indicate relationships)* | The goal is to create as many meaningful linkages as possible between (meta)data resources to enrich the contextual knowledge about the data, balanced against the time/energy involved in making a good data model.<br><br>Machine-readable (meta)data<br>(Meta)data in a format that can be automatically read and processed by a computer, such as CSV, JSON, XML etc. For more information: Open Data Handbook<br>Linked Data<br>Linked Data (also known as Linking Data) can be applied to improve the exploitation of the "Web of data." The expression refers to the publishing of structured data in a way that typed links are created between data from different sources to provide a higher level of usability. By using Linked Data, it is possible to find other, related data. Structured data should meet four requirements to be called Linked Data:<br>URIs should be assigned to all entities of the dataset.<br>HTTP URIs are required to ensure that all entities can be referenced and cited by users and user agents.<br>Entities should be described using standard formats such as RDF/XML.<br>Links should be created to other, related entity URIs.<br>All data that fulfil these requirements and are released for the public are called Linked Open Data (LOD). |
| ***Reusable*** | |
| *A license/usage rights should be attached to the data* | If data is not licensed no-one else can use it. Applying a Creative Commons licence to your data is a simple way to ensure that your data can be reused. The less restrictive the licence, the more that can be done with the data.<br>To make it easy for the Web to know when a work is available under a particular license, a "machine readable" version of the license provides a summary of the key freedoms and obligations written into a format that software systems, search engines, and other kinds of technology can understand.<br>Example from the Creative Commons License Chooser :<br>What appears in the metadata:<br>This work is licensed under a Creative Commons Attribution 4.0 International License. |
| *Provenance information should be captured to facilitate data reuse* | Data provenance is used to document where a piece of data comes from and the process and methodology by which it is produced. It is important to confirm the authenticity of data enabling trust, credibility and reproducibility. This is becoming increasingly important, especially in the eScience community where research is data intensive and often involves complex data transformations and procedures. |

## 6.2 Open access methods and tools

### Data management plans (DMP)

A DMP resolves uncertainties regarding data storage, publishing, discoverability, metadata formatting, data retrieval and data integration. A DMP requires routine review and revision and is, as such, a living document. It is becoming an increasingly commonplace standard to create project DMPs as they are required by many granting agencies and public and private funders to release funds (Kansa 2015, Faniel et al. 2018, Colella et al. 2020). The DMP should specify storage requirements, repository choices, data backup routines, hardware, and user access and permissions. Data life cycle and project long-term sustainability of the project is a critical aspect of a DMP. A DMP should consider that providers may not be permanent and data preservation standards evolve. Sharing, reuse, and legal compliance policies should also be stated, as well as administrator and user responsibilities and resources.

### Metadata standard – Darwin Core

Darwin Core (DwC) was developed by the historically named Taxonomic Databases Working Group, known today as Biodiversity Information Standards (TDWG). This scientific and educational not-for-profit was established to foster worldwide collaboration among the creators, managers, and users of biodiversity information with a goal of broadly spreading efficacious knowledge about global biological organism heritage. DwC has a fundamental role in the sharing, use, and reuse of open-access biodiversity data and describes billions of species occurrence records available through the Global Biodiversity Information Facility (GBIF) data repository. DwC provides open-standard metadata vocabularies using common language to publish and share biodiversity data. It allows data to be interoperable and integrative.

**Linked data**

Linked Open Data (LOD), or more simply linked data, can be used to describe and link data and terminologies between datasets. A key aspect of LOD is the use of stable web Uniform Resource Indicators (URIs) that address and locate resources on the internet, such as websites, files, or services (Kansa et al. 2020). Uniform Resource Locators (URLs) and Uniform Resource Names (URNs) are types of URIs that are typically composed of various components, including the protocol (e.g., http, https), domain name, path, and query parameters. URLs are used by web browsers and other applications to retrieve or interact with resources that may not remain stable over time. To maintain stability and be able to resolve to a URL that may have moved to a new location, persistent identifiers are required. Some examples of stable URIs include Digital Object Identifiers (DOIs), Persistent URLs (PURLs), and Archival Resource Keys (ARKs). A Unique Uniform Identifier (UUID) and a GUID (Globally Unique Identifier), also persistent identifiers, are types of URNs.

*Taxonomies and genomics*

There exist challenges inherent in the study of biodiversity due to the incredibly diverse array of life forms, estimated to include millions of species. This diversity spans various taxonomic groups, from microscopic bacteria to large mammals, and includes organisms in terrestrial, aquatic, and marine environments. No single, exhaustive list of all life on Earth exists and documenting and cataloging life is a monumental undertaking. The task of creating a single comprehensive list is compounded by taxonomic complexity as taxonomists continually revise classifications based on new discoveries and advancements in molecular techniques (Thiele et al. 2021). This is made even more complex by evolution as a continuous process and new discoveries in genetics creating challenges in species recognition. Taxonomic changes, additions, and revisions are ongoing, making it challenging to maintain a static and complete list. Additionally, new species are regularly discovered, particularly in less-explored regions, deep ocean environments, and microhabitats. The discovery of new species adds to the challenge of keeping a definitive list up to date. Furthermore, different taxonomic authorities may have variations in classification and nomenclature. Disagreements and debates within the taxonomic community regarding the status and placement of certain species can lead to variations in different lists. There have been calls to set strict yet flexible rules depending on hierarchical level. Efforts to address these challenges are ongoing through initiatives such as the Catalogue of Life (CoL) expert curated global checklist of species (Hobern et al. 2021). These projects aim to compile and integrate taxonomic information from various sources. However, achieving a complete and up-to-date list of all life on Earth remains a dynamic and ongoing process that requires constant updating, understanding of historic naming, synonyms, and regional differences in common names. reflecting the evolving nature of biodiversity science and taxonomy. These issues have consequences for conservation, illegal trade, and invasive pest control (Garnett and Christidis 2017, Conix et al. 2021, Thomson et al. 2021).

International Taxonomic Information System (ITIS) is specifically designed to provide authoritative taxonomic information. It is a partnership of U.S. federal agencies and specifically focuses on North American taxa but also includes a substantial amount of information on species from other regions. Researchers working primarily with North American species or those who appreciate ITIS's taxonomic rigor may choose ITIS for its regional expertise. The GBIF repository is a global network of participant countries, organizations, and institutions that contribute and provide free and open access biodiversity data from various sources, including museums, herbaria, research institutions, and citizen science projects. Over 2.6 billion occurrence records, 92,000 datasets, 2,100+ publishing institutions and over 10,000 peer-reviewed papers are accessible through GBIF (gbif.org 02/08/2024). GBIF is primarily a global biodiversity data aggregator that focuses on collecting and disseminating information about species occurrences, geographical distributions, and related data. While it includes taxonomic information, its main emphasis is on biodiversity data derived from a variety of sources.

National Center for Biotechnology Information (NCBI) provides phylogenetic classification and taxonomic nomenclature with links to internal genomic and protein databases and external taxonomic sources. NCBI is a crucial player in the field of biology and bioinformatics, and it plays a significant role in providing and organizing biological information. NCBI is a part of the United States National Library of Medicine (NLM), which is a branch of the National Institutes of Health (NIH). NCBI is not a taxonomic database like ITIS and GBIF, but it is closely related due to its role in managing and providing access to a wide range of biological data, including genomic, molecular, and taxonomic information. NCBI hosts and maintains various genomic and molecular databases, such as GenBank, which is a comprehensive collection of DNA sequences. This information includes genomic data for a vast number of species. Taxonomic information is often associated with these genomic sequences, linking molecular and taxonomic data. NCBI hosts a taxonomy database, which provides a standardized classification of living organisms. It includes names, synonyms, and common names, along with information about the hierarchical relationships among different taxonomic groups. The taxonomy database is often used as a reference for taxonomic information in molecular biology and genomics. NCBI collaborates with other international databases and platforms, including GBIF and ITIS. This collaboration helps ensure consistency and interoperability of taxonomic information across different databases. For example, NCBI

Taxonomy is often referenced in GenBank entries, linking molecular data with taxonomic classifications. NCBI provides a variety of bioinformatics tools and resources that researchers use to analyze and interpret genomic and molecular data. These tools may include functionalities related to taxonomy, phylogenetics, and evolutionary analysis. While NCBI is not a dedicated taxonomic platform like ITIS or GBIF, it plays a critical role in the broader landscape of biological information. NCBI provides a wealth of genomic and molecular data, and its taxonomy database contributes to the overall understanding of the taxonomic relationships among living organisms.

Fishbase is a global species database that focuses specifically on fish species information, including taxonomy, distribution, ecology, and behavior. It is a collaborative project involving partnerships with institutions and researchers worldwide such as fisheries management and aquatic biologists and relies on contributions from this community to continuously update and expand its database. FishBase covers both freshwater and marine fish species. It provides detailed taxonomic information on fish species, including accepted names, synonyms, common names, and hierarchical classifications. Indigenous language names are sourced from scholarly publications. It includes interactive tools, maps, and graphical representations to present data on aspects of fish biology, morphological characteristics, ecology, and fisheries-related information.

*Ontologies and terminology*

An ontology is a structured framework that represents relationships among concepts and procedural knowledge (Kintigh 2006). An ontology is a valuable tool for handling both structural and semantic differences between datasets, facilitating better integration and understanding in data-related contexts. Developing an ontology serves two main functions, first to resolve syntactic differences and second to accommodate different semantics. Syntactic differences refer to variations in the structure or format of data. An ontology helps resolve these differences, making it possible to integrate or compare datasets that may have different syntax. For example, consider two datasets that store information about dates. One dataset might represent dates in the format "MM/DD/YYYY," while the other uses "YYYY-MM-DD." An ontology could provide a standard way to reconcile these formats for consistency. Semantics involve the meaning of terms or concepts. Different datasets may use different terms or concepts for the same idea. An ontology helps accommodate these varying semantics by providing a common understanding or mapping between them. For example, in one dataset, "subadult" might be used, while another dataset refers to the same entity as "juvenile." An ontology could define a relationship between these terms, ensuring that they are understood as equivalent in a unified manner. An ontology can often be structured as a “treelike hierarchy of concepts with increasing specificity” (Kintigh et al. 2018). In this hierarchy, general concepts are represented at higher levels, and as you move down the hierarchy, the concepts become more specific. This organization allows for a clear representation of relationships and dependencies between concepts. At the top of the hierarchy, you have a root node representing a broad and general concept. This could be the overarching domain or main theme of the ontology. The hierarchy then branches out into intermediate nodes, each representing more specific subcategories or aspects of the general concept. These nodes become more focused as you move down the hierarchy. At the bottom of the hierarchy, you have leaf nodes, which represent the most specific concepts. These are the details or instances related to the intermediate nodes. The specificity increases as you move from the root to the leaves. The links or edges between nodes represent relationships between concepts. Parent nodes are connected to child nodes, indicating a broader-to-narrower relationship. This structure helps capture the semantic relationships and dependencies within the knowledge domain.

**Figure 25** displays faunal age as the broadest concept (root), and as you move down the hierarchy, you encounter more specific concepts. The specificity increases from the top to intermediate (Subadult) and leaf nodes the bottom (Juvenile, Immature, Fetal or Neonate) of the hierarchy. This kind of hierarchical structure aids in organizing knowledge, facilitating navigation, and understanding relationships between different concepts within the ontology.

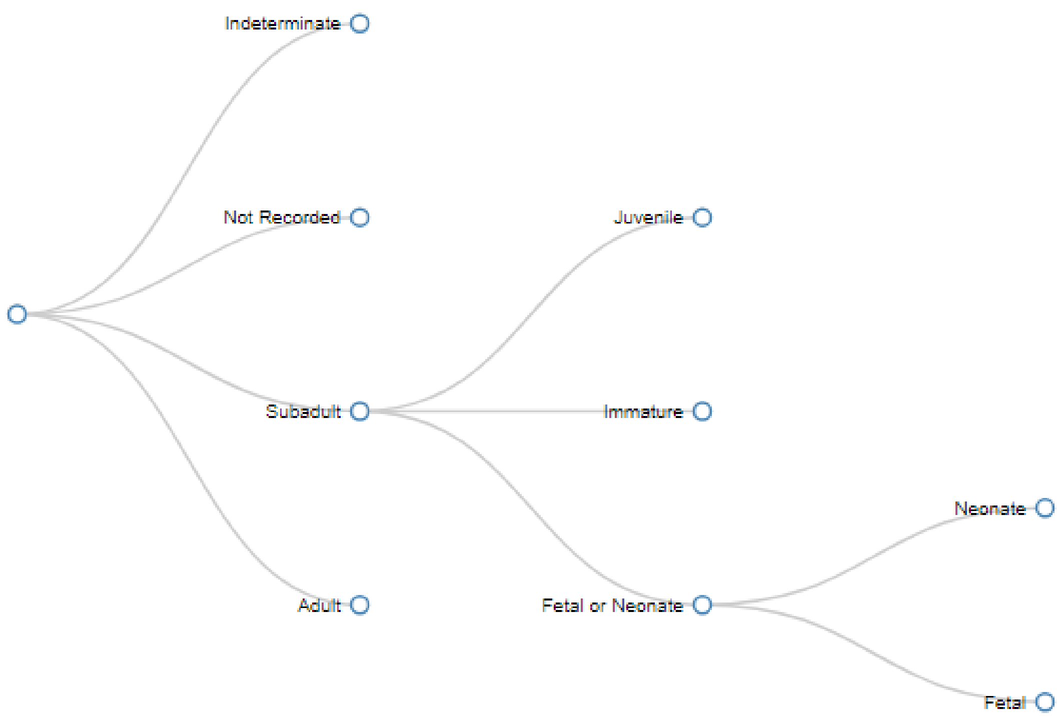



**Figure 25** tDAR Faunal Age Ontology depicting hierarchical levels (Source - doi:10.6067/XCV8PZ574Q)

Ontologies are available from the Digital Archaeological Record (tDAR)[75] and UBERON[76], which is part of the OLS or Ontology Lookup Service. Both tDAR and UBERON have persistent identifiers to link terminology. tDAR uses DOIs to link each ontology, whereas UBERON uses

[75] https://core.tdar.org/ontology/
[76] https://www.ebi.ac.uk/ols4/ontologies/uberon

PURLs. Both provide a stable and persistent reference to a specific resource, even if the URL itself changes or becomes inaccessible.

**Workflow platforms and storage**

*Open Science Framework (OSF)*

Choosing a project management platform for research data should consider secure storage, user access permissions, and version control. Open Science Framework (OSF) is one such platform designed to support open and collaborative research practices. OSF provides user access restrictions, data storage and document version control. Each document uploaded receives a GUID from OSF. Public projects on OSF typically align with the FAIR (Findable, Accessible, Interoperable, Reusable) principles and promote transparency and openness in research. OSF public projects align with findability by using clear and descriptive titles and metadata, making them easily discoverable through search engines and within the OSF platform. Additionally, OSF assigns unique and persistent identifiers to projects, ensuring their long-term findability. Public projects on OSF are openly accessible to anyone without the need for an account or login. This aligns with the accessibility aspect of the FAIR principles. Research data, materials, and documents shared within OSF projects are typically available for download or viewing, promoting easy access. OSF allows researchers to add standardized metadata to their projects, facilitating interoperability with other research platforms and data repositories. OSF integrates with various research tools and services, such as external services to manage citations and storage, making it easier to link and work with other research-related software and systems. OSF promotes reusability by providing a structured platform for organizing research materials, data, and documentation, making them more accessible and reusable by other researchers. It enhances transparency and reproducibility by offering version control, allowing researchers to track changes and updates to their projects. Researchers can specify licensing terms for their projects, indicating how others can reuse their work while adhering to legal and ethical considerations. OSF public projects are designed to align with the FAIR principles by providing a user-friendly, transparent, and openly accessible platform for researchers to share, collaborate on, and disseminate their work. By following these principles, OSF contributes to the promotion of open science and the broader research community's ability to discover, access, and build upon research findings.

*GitHub*

GitHub is a platform primarily focused on code version control and collaboration within software development. While it's not specifically oriented toward research data management, it does have features that can align with certain aspects of the FAIR principles when used for open-source projects and collaborative research. GitHub provides a platform where researchers and developers can create and host repositories for their projects, making them easily findable through search engines and within the GitHub platform. Project README files typically contain information about the project's purpose, usage, and documentation, aiding in findability. GitHub allows users to create both public and private repositories. Public repositories are openly accessible to anyone, while private ones are restricted to authorized collaborators, ensuring various levels of access. GitHub's issue tracking system allows collaborators to report and discuss issues and improvements, enhancing accessibility for project contributors. GitHub uses Git, a widely adopted version control system, making it interoperable with a range of development tools and platforms. GitHub offers an Application Programming Interface (API) and integrates with various third-party tools, enhancing interoperability with other research and development systems. Open-source code hosted on GitHub is often reusable by other developers and researchers, fostering collaboration and reuse. Developers can specify licenses for their projects on GitHub, indicating how others can reuse the code while adhering to legal and ethical considerations.

**Data repositories and registries**

*Canadensys, GBIF, and GrSciColl*

Canadensys, a biodiversity data-sharing network and repository operating from the University of Montreal Biodiversity Centre, digitizes, publishes, and georeferences biological collections. It serves as a platform for various institutions, researchers, and individuals to share, access, and use biodiversity data from Canada. The repository is powered by the GBIF Integrated Publishing Toolkit (ITP), an open-source web application developed by GBIF and customized by Canadensys, used to publish and register datasets. The ITP maps Occurrence dataset file DwC fields to confirm proper formatting and, once confirmed, publishes to the Canadensys and GBIF repositories. Extension files, such as a Measurement or Facts (MoF) file, and a Simple Multimedia file, are used to add additional data to specimens that cannot be entered in a single occurrence core record field due to formatting restrictions. A separate row, linked to the specimen's unique occurrenceID DwC field, created for each measurement type ensures a complete description for specimens with multiple measurements. The Simple Multimedia file, also linked to the occurrence file by the occurrenceID, contains images. Both Canadensys and GBIF upload data in a standardized DwC format, making it easier to integrate and compare information across different sources to improve understanding of biodiversity patterns, and support global conservation initiatives.

GRSciColl and GBIF are interconnected in the realm of biodiversity data management, but they serve different primary functions. GRSciColl, a registry that documents various biological, geological, and other natural science collections held by museums, herbaria, research institutions, and similar establishments, focuses on cataloging and maintaining information about scientific collections worldwide. This information includes data about the specimens, their origins, and the institutions holding them. Conversely, GBIF is a global network that works on aggregating and providing access to biodiversity data from various sources worldwide. It aims to make this data freely available for scientific research, conservation, and sustainable development. GBIF aggregates data from multiple sources, including scientific collections, field observations, and other biodiversity-related datasets. GRSciColl provides visual summary of data from individual collections, which are one of the many sources of biodiversity data aggregated by GBIF. GRSciColl's information about the existence and locations of scientific collections helps GBIF in identifying and accessing these collections' data. When institutions register their collections in GRSciColl, it further enhances the visibility and accessibility of these collections' data through GBIF's platform.

# References

Ames, K. 1991. The Archaeology of the Longue Durée: Temporal and Spatial Scale in the Evolution of Social Complexity on the Southern Northwest Coast. Antiquity.

Anderson, D. G. 2024. Modern archaeological research and the importance of information management. Proceedings of the National Academy of Sciences 121:e2403792121.

Armstrong, C. G., A. C. Shoemaker, I. McKechnie, A. Ekblom, P. Szabó, P. J. Lane, A. C. McAlvay, O. J. Boles, S. Walshaw, N. Petek, K. S. Gibbons, E. Q. Morales, E. N. Anderson, A. Ibragimow, G. Podruczny, J. C. Vamosi, T. Marks-Block, J. K. LeCompte, S. Awâsis, C. Nabess, P. Sinclair, and C. L. Crumley. 2017. Anthropological contributions to historical ecology: 50 questions, infinite prospects. PLOS ONE 12:e0171883.

Armstrong, C. G., and J. R. Veteto. 2015. Historical Ecology and Ethnobiology: Applied Research for Environmental Conservation and Social Justice. Ethnobiology Letters 6:5–7.

Atlas, W. I., N. C. Ban, J. W. Moore, A. M. Tuohy, S. Greening, A. J. Reid, N. Morven, E. White, W. G. Housty, J. A. Housty, C. N. Service, L. Greba, S. Harrison, C. Sharpe, K. I. R. Butts, W. M. Shepert, E. Sweeney-Bergen, D. Macintyre, M. R. Sloat, and K. Connors. 2021. Indigenous Systems of Management for Culturally and Ecologically Resilient Pacific Salmon (Oncorhynchus spp.) Fisheries. BioScience 71:186–204.

Balée, W. L. 1998. Advances in Historical Ecology. Columbia University Press.

Beller, E. E., L. McClenachan, E. S. Zavaleta, and L. G. Larsen. 2020. Past forward: Recommendations from historical ecology for ecosystem management. Global Ecology and Conservation 21:e00836.

Betts, M. W., H. D. G. Maschner, C. D. Schou, R. Schlader, J. Holmes, N. Clement, and M. Smuin. 2011. Virtual zooarchaeology: building a web-based reference collection of northern vertebrates for archaeofaunal research and education. Journal of Archaeological Science 38:755.e1-755.e9.

Bingham, H., M. Doudin, L. Weatherdon, K. Despot-Belmonte, F. Wetzel, Q. Groom, E. Lewis, E. Regan, W. Appeltans, A. Güntsch, P. Mergen, D. Agosti, L. Penev, A. Hoffmann, H. Saarenmaa, G. Geller, K. Kim, H. Kim, A.-S. Archambeau, C. Häuser, D. Schmeller, I. Geijzendorffer, A. G. Camacho, C. Guerra, T. Robertson, V. Runnel, N. Valland, and C. Martin. 2017. The Biodiversity Informatics Landscape: Elements, Connections and Opportunities. Research Ideas and Outcomes 3:e14059.

Blackburn, D. C., D. M. Boyer, J. A. Gray, J. Winchester, J. M. Bates, S. L. Baumgart, E. Braker, D. Coldren, K. W. Conway, A. D. Rabosky, N. de la Sancha, C. B. Dillman, J. L. Dunnum, C. M. Early, B. W. Frable, M. W. Gage, J. Hanken, J. A. Maisano, B. D. Marks, K. P. Maslenikov, J. E. McCormack, R. S. Nagesan, G. G. Pandelis, H. L. Prestridge, D. L. Rabosky, Z. S. Randall, M. B. Robbins, L. A. Scheinberg, C. L. Spencer, A. P. Summers, L. Tapanila, C. W. Thompson, L. Tornabene, G. J. Watkins-Colwell, L. J. Welton, the oVert Project Team, and E. L. Stanley. 2024. Increasing the impact of vertebrate scientific collections through 3D imaging: The openVertebrate (oVert) Thematic Collections Network. BioScience:biad120.

Blaser, M. 2016. Is Another Cosmopolitics Possible? Cultural Anthropology 31:545–570.

Bliege Bird, R., and D. Nimmo. 2018. Restore the lost ecological functions of people. Nature Ecology & Evolution 2:1050–1052.

Bromham, L., R. Dinnage, H. Skirgård, A. Ritchie, M. Cardillo, F. Meakins, S. Greenhill, and X. Hua. 2022. Global predictors of language endangerment and the future of linguistic diversity. Nature Ecology & Evolution 6:163–173.

Brown, S., K. Douka, M. J. Collins, and K. K. Richter. 2021. On the standardization of ZooMS nomenclature. Journal of Proteomics 235:104041.

Carroll, S. R., E. Herczog, M. Hudson, K. Russell, and S. Stall. 2021. Operationalizing the CARE and FAIR Principles for Indigenous data futures. Scientific Data 8:108.

Colella, J. P., R. B. Stephens, M. L. Campbell, B. A. Kohli, D. J. Parsons, and B. S. Mclean. 2020. The Open-Specimen Movement. BioScience.

Conix, S., S. T. Garnett, K. R. Thiele, L. Christidis, P. P. van Dijk, O. S. Bánki, S. K. Barik, J. S. Buckeridge, M. J. Costello, D. Hobern, P. M. Kirk, A. Lien, S. Nikolaeva, R. L. Pyle, S. A. Thomson, Z.-Q. Zhang, and F. E. Zachos. 2021. Towards a global list of accepted species III. Independence and stakeholder inclusion. Organisms Diversity & Evolution 21:631–643.

Cook, J. A., S. V. Edwards, E. A. Lacey, R. P. Guralnick, P. S. Soltis, D. E. Soltis, C. K. Welch, K. C. Bell, K. E. Galbreath, C. Himes, J. M. Allen, T. A. Heath, A. C. Carnaval, K. L. Cooper, M. Liu, J. Hanken, and S. Ickert-Bond. 2014. Natural History Collections as Emerging Resources for Innovative Education. BioScience 64:725–734.

Curty, R. G., K. Crowston, A. Specht, B. W. Grant, and E. D. Dalton. 2017. Attitudes and norms affecting scientists' data reuse. PLoS ONE 12.

Driver, J. C. 2011. Identification, Classification and Zooarchaeology. Ethnobiology Letters 2:19–39.

Driver, J. C., K. Bovy, V. L. Butler, K. D. Lupo, R. L. Lyman, and C. Otaola. 2011. Identification, Classification and Zooarchaeology (Featured Reprint & Invited Comments). Ethnobiology Letters 2:19–39.

van Erp, M., R. Hensel, D. Ceolin, and M. van der Meij. 2015. Georeferencing Animal Specimen Datasets. Transactions in GIS 19:563–581.

Faniel, I. M., A. Austin, E. Kansa, S. W. Kansa, P. France, J. Jacobs, R. Boytner, and E. Yakel. 2018. Beyond the Archive: Bridging Data Creation and Reuse in Archaeology. Advances in Archaeological Practice 6:105–116.

Farnel, S. 2018. Metadata as Data: Exploring Ethical Metadata Sharing and Access for Indigenous Resources Through OCAP Principles. Proceedings of the Annual Conference of CAIS / Actes du congrès annuel de l'ACSI.

Federhen, S. 2015. Type material in the NCBI Taxonomy Database. Nucleic Acids Research 43:D1086–D1098.

Fordham, D. A., S. T. Jackson, S. C. Brown, B. Huntley, B. W. Brook, D. Dahl-Jensen, M. T. P. Gilbert, B. L. Otto-Bliesner, A. Svensson, S. Theodoridis, J. M. Wilmshurst, J. C. Buettel, E. Canteri, M. McDowell, L. Orlando, J. A. Pilowsky, C. Rahbek, and D. Nogues-Bravo. 2020. Using paleo-archives to safeguard biodiversity under climate change. Science 369:eabc5654.

Fossile, T., J. Ferreira, D. da R. Bandeira, S. Dias-da-Silva, and A. C. Colonese. 2020. Integrating zooarchaeology in the conservation of coastal-marine ecosystems in Brazil. Quaternary International 545:38–44.

Franco, F. M. 2021. Intellectualist premise of folk names support their restoration in formal taxonomy. Ethnobotany Research and Applications 21:1–3.

Frazier, J. 2007. Sustainable use of wildlife: The view from archaeozoology. Journal for Nature Conservation 15:163–173.

Froese, R., and D. Pauly, editors. 2024. FishBase. World Wide Web electronic publication.

Garnett, S. T., and L. Christidis. 2017. Taxonomy anarchy hampers conservation. Nature 546:25–27.

Gauvreau, A., D. Fedje, A. Dyck, Q. Mackie, C. F. G. Hebda, K. Holmes, Q. Y. E. White, D. W. Housty, Ĝ. R. Housty, and D. McLaren. 2023. Geo-archaeology and Haíłzaqv oral history: Long-term human investment and resource use at EkTb-9, Triquet Island, N̓úláw̓itxˇv Tribal Area, Central Coast, British Columbia, Canada. Journal of Archaeological Science: Reports 49:103884.

Gifford-Gonzalez, D. 2018. An Introduction to Zooarchaeology. Springer International Publishing.

Gillman, L. N., and S. D. Wright. 2020. Restoring indigenous names in taxonomy. Communications Biology 3:1–3.

Gobalet, K. W. 2001. A Critique of Faunal Analysis; Inconsistency among Experts in Blind Tests. Journal of Archaeological Science 28:377–386.

Gomes, N. 2020. Reclaiming Native Hawaiian Knowledge Represented in Bird Taxonomies. Ethnobiology Letters 11:30–43.

Graeber, D. 2015. Radical alterity is just another way of saying “reality”: A reply to Eduardo Viveiros de Castro. HAU: Journal of Ethnographic Theory 5:1–41.

Gupta, A., and J. Ferguson. 1997. Beyond “Culture”: Space, Identity, and the Politics of Difference. Pages 33–51 *in* A. Gupta and J. Ferguson, editors. Culture, Power, Place: Explorations in Critical Anthropology.

Gupta, A., and J. Stoolman. 2022. Decolonizing US anthropology. American Anthropologist n/a.

Gupta, N., A. Martindale, K. Supernant, and M. Elvidge. 2023. The CARE Principles and the Reuse, Sharing, and Curation of Indigenous Data in Canadian Archaeology. Advances in Archaeological Practice 11:76–89.

Hakai Institute, and Tula Foundation. 2023, August 11. zooarch/python/setup.py at geocoding · HakaiInstitute/zooarch. https://github.com/HakaiInstitute/zooarch/blob/geocoding/python/setup.py.

Hardesty, D. L. 2007. Perspectives on Global-Change Archaeology. American Anthropologist 109:1–7.

Hedrick, B. P., J. M. Heberling, E. K. Meineke, K. G. Turner, C. J. Grassa, D. S. Park, J. Kennedy, J. A. Clarke, J. A. Cook, D. C. Blackburn, S. V. Edwards, and C. C. Davis. 2020. Digitization and the Future of Natural History Collections. BioScience 70:243–251.

Henare, A., M. Holbraad, and S. Wastell. 2007. Introduction: Thinking Through Things. Page Thinking through things: Theorizing artefacts ethnographically. Routledge, New York.

Hillis, D., R. Gustas, D. Pauly, W. W. L. Cheung, A. K. Salomon, and I. McKechnie. 2022. A palaeothermometer of ancient Indigenous fisheries reveals increases in mean temperature of the catch over five millennia. Environmental Biology of Fishes 105:1381–1397.

Hobern, D., S. K. Barik, L. Christidis, S. T.Garnett, P. Kirk, T. M. Orrell, T. Pape, R. L. Pyle, K. R. Thiele, F. E. Zachos, and O. Bánki. 2021. Towards a global list of accepted species VI: The Catalogue of Life checklist. Organisms Diversity & Evolution 21:677–690.

Hollesen, J. 2022. Climate change and the loss of archaeological sites and landscapes: a global perspective. Antiquity:1–14.

Jennings, L., T. Anderson, A. Martinez, R. Sterling, D. D. Chavez, I. Garba, M. Hudson, N. A. Garrison, and S. R. Carroll. 2023. Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. Nature Ecology & Evolution:1–5.

Johnson, K. R., I. F. P. Owens, and The Global Collection Group. 2023. A global approach for natural history museum collections. Science 379:1192–1194.

Kansa, S. W. 2015. Using Linked Open Data to Improve Data Reuse in Zooarchaeology. Ethnobiology Letters 6:224–231.

Kansa, S. W., L. Atici, E. C. Kansa, and R. H. Meadow. 2020. Archaeological Analysis in the Information Age: Guidelines for Maximizing the Reach, Comprehensiveness, and Longevity of Data. Advances in Archaeological Practice 8:40–52.

Kintigh, K. 2006. The Promise and Challenge of Archaeological Data Integration. American Antiquity 71:567–578.

Kintigh, K. W., K. A. Spielmann, A. Brin, K. S. Candan, T. C. Clark, and M. Peeples. 2018. Data Integration in the Service of Synthetic Research. Advances in Archaeological Practice 6:30–41.

LeFebvre, M. J. 2022, November 16. ChronoAge related archaeo question.

Lyman, R. L. 1996. Applied zooarchaeology: The relevance of faunai analysis to wildlife management. World Archaeology 28:110–125.

Lyman, R. L. 2010. Paleozoology's Dependence on Natural History Collections. Journal of Ethnobiology 30:126–136.

Lyman, R. L. 2017. Paleoenvironmental Reconstruction from Faunal Remains: Ecological Basics and Analytical Assumptions. Journal of Archaeological Research 25:315–371.

Lyman, R. L. 2019. Assumptions and Protocol of the Taxonomic Identification of Faunal Remains in Zooarchaeology: a North American Perspective. Journal of Archaeological Method and Theory 26:1376–1438.

Mabry, M. E., F. Zapata, D. L. Paul, P. M. O'Connor, P. S. Soltis, D. C. Blackburn, and N. B. Simmons. 2022. Monographs as a nexus for building extended specimen networks using persistent identifiers. Bulletin of the Society of Systematic Biologists 1.

Mackie, Q., D. Fedje, D. McLaren, and D. McLaren. 2018. Archaeology and Sea Level Change on the British Columbia Coast. Canadian Journal of Archaeology/Journal canadien d'archéologie 42:074–091.

Martindale, A., and G. P. Nicholas. 2014. Archaeology as Federated Knowledge. Canadian Journal of Archaeology / Journal Canadien d'Archéologie 38:434–465.

Marwick, B., J. d'Alpoim Guedes, C. M. Barton, L. A. Bates, M. Baxter, A. Bevan, E. A. Bollwerk, R. K. Bocinsky, T. Brughmans, A. K. Carter, C. Conrad, D. A. Contreras, S. Costa, E. R. Crema, A. Daggett, B. Davies, B. L. Drake, T. S. Dye, P. France, R. Fullagar, D. Giusti, S. Graham, M. D. Harris, J. Hawks, S. Health, D. Huffer, E. C. Kansa, S. W. Kansa, M. E. Madsen, J. Melcher, J. Negre, F. D. Neiman, R. Opitz, D. C. Orton, P. Przstupa, M. Raviele, J. Riel-Savatore, P. Riris, I. Romanowska, J. Smith, N. Strupler, I. I. Ullah, H. G. Van Vlack, N. VanValkenburgh, E. C. Watrall, C. Webster, J. Wells, J. Winters, and C. D. Wren. 2017. Open science in archaeology. SAA Archaeological Record 17:8–14.

Mathews, D. L., and N. J. Turner. 2017. Chapter 9 - Ocean Cultures: Northwest Coast Ecosystems and Indigenous Management Systems. Pages 169–206 *in* P. S. Levin and M. R. Poe, editors. Conservation for the Anthropocene Ocean. Academic Press.

McKechnie, I., and M. L. Moss. 2016. Meta-analysis in zooarchaeology expands perspectives on Indigenous fisheries of the Northwest Coast of North America. Journal of Archaeological Science: Reports 8:470–485.

McKechnie, I., M. L. Moss, and S. J. Crockford. 2020. Domestic dogs and wild canids on the Northwest Coast of North America: Animal husbandry in a region without agriculture? Journal of Anthropological Archaeology 60:101209.

McKenzie, K., and I. McKechnie. 2024. UVic Zooarchaelogy Lab (UVicZL) Data Management. OSF.

McLaren, D., D. Fedje, Q. Mackie, L. G. Davis, J. Erlandson, A. Gauvreau, and C. Vogelaar. 2020. Late Pleistocene Archaeological Discovery Models on the Pacific Coast of North America. PaleoAmerica 6:43–63.

McLaren, D., F. Rahemtulla, G. (Elroy White), and D. Fedje. 2015. Prerogatives, Sea Level, and the Strength of Persistent Places: Archaeological Evidence for Long-Term Occupation of the Central Coast of British Columbia. BC Studies: The British Columbian Quarterly.

McMillan, A., and I. McKechnie. 2015. Investigating Indigenous Adaptations to British Columbia’s Exposed Outer Coast: Introduction to These Outer Shores. BC Studies:3–20.

Meighan, C. W., D. M. Pendergast, B. K. Swartz, and M. D. Wissler. 1958. Ecological Interpretation in Archaeology: Part I. American Antiquity 24:1–23.

Miller, S. E., L. N. Barrow, S. M. Ehlman, J. A. Goodheart, S. E. Greiman, H. L. Lutz, T. M. Misiewicz, S. M. Smith, M. Tan, C. J. Thawley, J. A. Cook, and J. E. Light. 2020. Building Natural History Collections for the Twenty-First Century and Beyond. BioScience 70:674–687.

Monfils, A. K., E. R. Krimmel, J. M. Bates, J. E. Bauer, M. W. Belitz, B. C. Cahill, A. M. Caywood, N. S. Cobb, J. B. Colby, S. A. Ellis, D. M. Krejsa, T. D. Levine, T. D. Marsico, T. J. Mayfield-Meyer, J. A. Miller-Camp, R. M. (Gil) Nelson, M. A. Phillips, M. A. Revelez, D. R. Roberts, R. A. Singer, and J. M. Zaspel. 2020. Regional Collections Are an Essential Component of Biodiversity Research Infrastructure. BioScience 70:1045–1047.

Niven, L., T. E. Steele, H. Finke, T. Gernat, and J.-J. Hublin. 2009. Virtual skeletons: using a structured light scanner to create a 3D faunal comparative collection. Journal of Archaeological Science 36:2018–2023.

Orton, D. C. 2016. Archaeology as a tool for understanding past marine resource use and its impact. Pages 47–69 *in* K. Schwerdtner Máñez and B. Poulsen, editors. Perspectives on Oceans Past. Springer, New York.

Peres, T. M. 2010. Methodological Issues in Zooarchaeology. Pages 15–36 *in* A. M. VanDerwarker and T. M. Peres, editors. Integrating Zooarchaeology and Paleoethnobotany: A Consideration of Issues, Methods, and Cases. Springer, New York, NY.

Peterson, A. T., C. Cicero, and J. Wieczorek. 2005. Free and Open Access to Bird Specimen Data: Why? The Auk 122:987–990.

Reid, A. J., D. A. McGregor, A. K. Menzies, L. E. Eckert, C. M. Febria, and J. N. Popp. 2024. Ecological research ‘in a good way’ means ethical and equitable relationships with Indigenous Peoples and Lands. Nature Ecology & Evolution:1–4.

Richter, K. K., M. C. Codlin, M. Seabrook, and C. Warinner. 2022. A primer for ZooMS applications in archaeology. Proceedings of the National Academy of Sciences 119:e2109323119.

Rick, T. C. 2023. Coastal Archaeology and Historical Ecology for a Changing Planet. Journal of Anthropological Research 79:153–175.

Salomon, A. K., D. K. Okamoto, K̲ii’iljuus Barbara J. Wilson, hiininaasim Tommy Happynook, null Wickaninnish, wiicuckum A. Mack, S. H. Allan Davidson, G. Guujaaw, W. W. H. L. Humchitt, T. M. Happynook, weiwimtaeek C. Cox, H. F. Gillette, n’yasim S. Christiansen, D. Dragon, H. M. Kobluk, L. C. Lee, M. T. Tinker, J. J. Silver, D. Armitage, I. McKechnie, A. MacNeil, D. Hillis, E.-K. Muhl, E. J. Gregr, C. J. C. Commander, and A. Augustine. 2023. Disrupting and diversifying the values, voices and governance principles that shape biodiversity science and management. Philosophical Transactions of the Royal Society B: Biological Sciences 378:20220196.

Santos, B. de S. 2015. Epistemologies of the South: Justice Against Epistemicide. Routledge, New York.

Schnarch, B. 2004. Ownership, Control, Access, and Possession (OCAP) or Self-Determination Applied to Research: A Critical Analysis of Contemporary First Nations Research and Some Options for First Nations Communities. International Journal of Indigenous Health 1:80–95.

Sholts, S. B., J. A. Bell, and T. C. Rick. 2016. Ecce Homo: Science and Society Need Anthropological Collections. Trends in Ecology & Evolution 31:580–583.

Stahl, A. B. 2020. Assembling “Effective Archaeologies” toward Equitable Futures. American Anthropologist 122:37–50.

Thiele, K. R., S. Conix, R. L. Pyle, S. K. Barik, L. Christidis, M. J. Costello, P. P. van Dijk, P. Kirk, A. Lien, S. A. Thomson, F. E. Zachos, Z.-Q. Zhang, and S. T. Garnett. 2021. Towards a global list of accepted species I. Why taxonomists sometimes disagree, and why this matters. Organisms Diversity & Evolution 21:615–622.

Thomson, S. A., K. Thiele, S. Conix, L. Christidis, M. J. Costello, D. Hobern, S. Nikolaeva, R. L. Pyle, P. P. van Dijk, H. Weaver, F. E. Zachos, Z.-Q. Zhang, and S. T. Garnett. 2021. Towards a global list of accepted species II. Consequences of inadequate taxonomic list governance. Organisms Diversity & Evolution 21:623–630.

Trisos, C. H., J. Auerbach, and M. Katti. 2021. Decoloniality and anti-oppressive practices for a more ethical ecology. Nature Ecology & Evolution 5:1205–1212.

Trouillot, M.-R. 2003. Anthropology and the Savage Slot: The Poetics and Politics of Otherness. Pages 7–28 *in* M.-R. Trouillot, editor. Global Transformations: Anthropology and the Modern World. Palgrave Macmillan US, New York.

Wiley, A. E., H. F. James, and P. H. Ostrom. 2017. Emerging Techniques for Isotope Studies of Avian Ecology 1. Page The Extended Specimen: Emerging Frontiers in Collections-based Ornithological Research. CRC Press.

Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. ’t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van

Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3:160018.

Wolf, E. R. 1982. Europe and the people without history. University of California Press, Berkeley.

Wolverton, S. 2013. Data Quality in Zooarchaeological Faunal Identification. Journal of Archaeological Method and Theory 20:381–396.