Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

Final Report and Additional Information About this Repository:

Because all of the information packed into this repository is public domain, has been published online, and concerns scientific data collection, the names and institutions of those who contributed is plainly evident in the material and its citations. However, because NOAA is a government organization, and WOLL is a focus project made by a small group of researchers affiliated with a few German universities, their needs and aims differ. Accordingly, NOAA relies on Globally Unique Identifiers (GUIDs) whereas WOLL lists the Direct Object Identifiers (DOIs) of their publications on the sidebar of all their files. A suitable middle ground for their incorporation into this repository would need to give credit to the information's creators (so not unique identifiers), while also allowing NOAA's ongoing data to be accounted for when it has not been published under a DOI. For that reason, direct academic data citation of the websites and articles included may be the best option to incorporate both NOAA's dataset and WOLL's repository under a single umbrella. Further investigation into this issue following feedback is necessary,

Concerning long-term preservation of the data in this repository, WOLL's repository and its apparent loss of data presents a real issue. Some of the files in the "Textile Tool Database" have inexplicably become unavailable—even their thumbnails disappeared—although they still exist in the list of files in that database. Those files could not be included in this repository. Perhaps it is due to *Edition*'s open-access platform being vulnerable to deletion of published

files. Thankfully, no special software is required to open any files from either NOAA's dataset or WOLL's repository—they are all either .csv, .xsls, .pdf, .xml. or .json file types.

Understandably, WOLL's information is protected under a Creative Commons License, yet no licensing information is provided for NOAA's Hawaiian monk seal telemetry dataset. The dataset's additional metadata table offers that use "for analysis or publication is expected to be conducted in collaboration with and proper attribution to NOAA scientists responsible for data collection." Because of that, a Creative Commons License on the repository would be suitable, assuming that telemetry data of monk seals would be used responsibly.

While none of the data tables in this repository contain identifiable information from the researchers, the metadata files in both the WOLL repository and the NOAA dataset do. However, there is no need to adjust the privacy when combining the two into a repository, as the work of the scientists and researchers behind the data should ethically be attached to its fair and responsible use. Such is the express wish of the publishers, and so no measures in confidentiality need to be taken when republishing public information with good intentions for fair use.

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

A Thorough Investigation of "Hawaiian Monk Seal Telemetry Tag Deployments", Published by NOAA's Protected Species Division Hawaiian Monk Seal Research Program to *Data.gov*

URL: https://catalog.data.gov/dataset/hmsrp-hawaiian-monk-seal-telemetry-tag-deployments

As part of the United States Department of Commerce, the scientific and regulatory agency National Oceanic and Atmospheric Association (NOAA) collects meteorological and oceanographic data in order to forecast weather and predict climate conditions. Its agency is split into six branches: the National Environmental Satellite, Data, and Information Service, the National Marine Fisheries Service, the National Ocean Service, the National Weather Service, the Office of Marine & Aviation Operations, and the Office of Oceanic & Atmospheric Research. Each branch has its own dedication, evident from the title.

Protected through the OPEN Government Data Act, and codified in Title II of the Foundations for Evidence-Based Policymaking Act, the distribution and display of United States federal data within the General Services Administration's repository *Data.gov* undoubtedly involves bodies such as NOAA. But it was Ph.D. laureate Charles Littnan, fellow of the National Marine Fisheries Service and director of the Protected Species Division, who contributed the dataset "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments" to *Data.gov* on October 20, 2022. Its package includes geospatial coordinates of "animal-mounted cell phone and other telemetry tag" measurements from Hawaii's native seal population. Its data displays the locations

of seals that were tracked, and the data which was recorded, downloaded, and lightly processed before being compiled into .CSV files.

As federal data, key stakeholders in NOAA's "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments"dataset include the United States government, who are highly invested in the restoration of population counts of Hawaiian monk seals—today considered to be one of the most endangered seal species in the world. As current population counts are down to 1,570 seals total (one third of historic values), NOAA's website notes that "monk seals are protected under the Endangered Species Act, the Marine Mammal Protection Act, and State of Hawai'i law", as well as being one of their "Species in the Spotlight" (NOAA Fisheries). For that, it could be argued that the State of Hawaii and its tourism industry also have a stake in population data of Hawaiian monk seals. In fact, NOAA themselves have stake in Littnan's telemetry data, as a current uptick in numbers of Hawaiian monk seals is a result of their "recovery efforts" and careful monitoring of the species' wellbeing (NOAA Fisheries). Students who use the data for anything from mapmaking to essay writing have stake in the availability of the "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments" dataset too. Although less strongly, other relevant institutions may include the Environmental Protection Agency, or even international agencies like the United Nations, to whom global ecology data is relevant.

As the dataset is still in progress, its contents are sparse. Currently, within the "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments", there is one bulk TAR Compressed File containing telemetry data in .CSV format and one .PDF file containing NOAA's Data

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

Management Plan for ongoing uploads. Besides that, there are two attached links, one explaining the dataset's internal Global Change Master Directory (GCMD) keywords, and the other providing a full metadata record. Because the two links serve only to provide context to users, I would argue that the "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments" dataset contains just two bulk files. As the files are in .CSV and .PDF form, no special software is required to open or manage the data, apart from standard services such as Microsoft Excel or Adobe Reader.

As listed below in *Data.gov*'s attached "Additional Metadata" table, there are no access constraints. However, the use constraint is specific to "analysis or publication [...] conducted in collaboration with and proper attribution to NOAA scientists responsible for data collection". This is understandable, because NOAA's data is representative of federal movements to monitor and improve Hawaiian monk seal populations. For a student working on a class project, it could be assumed that proper citation and credit to the research team would be sufficient evidence of "collaboration and proper attribution". Without any licenses, any recommendations are just that.

By clicking the link to a full metadata record, the "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments" dataset provides catalog details, such as: dates of publication, frequency of updates, geospatial location of its data collection, file types the dataset contains, contact information of the lead researcher and their affiliation, a lineage statement with data processing steps (including more contact information), data management and quality specifics, and a download link for the one .CSV file.

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

Solely because there is not much data with which to produce metadata—only the one .CSV of geospatial data and a .PDF of the Data Management Plan—the full metadata report is quite short, albeit comprehensive. Its internal catalog details, found in the full metadata report, are the standard for its metadata—including an item ID (28583), Globally Unique Identifier (gov.noaa.nmfs.inport:28583), and acknowledgement of its parent organization, the Pacific Islands Fisheries Science Center.

Due to its recent publication (February 23, 2018), I could not find any external papers citing this dataset when browsing databases like Academic Search Complete or Google Scholar. However, NOAA has used its own data in more recent publications, such as the "HMSRP Hawaiian Monk Seal Specimen Data" dataset (which includes physical specimens, collection information, status, storage locations, and laboratory results associated with individual specimens)" on *Data.gov,* dating to October 28th, 2022. In particular, that dataset includes "physical specimens, paper logs and Freezerworks database of all logged information on specimens collected from Hawaiian monk seals since 1975"—of which the .CSV file from the "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments" dataset is but one component.

Due to the fact that the "HMSRP Hawaiian Monk Seal Telemetry Tag Deployments" dataset is so small, and practically only contains a single bulk file, I would say its enrichment depends entirely on time and updates. User ability to access the files are strong, as they only utilize common formats such as .PDF and .CSV—but with such a narrow scope of focus, applicability to the layperson is low. Even still, the "HMSRP Hawaiian Monk Seal Telemetry

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

Tag Deployments" dataset makes great use of what little information it contains, and labels its metadata thoroughly and appropriately.

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

Exploring WOLL: A German Repository Titled "The Textile Revolution", Created by Exzellenzcluster Topoi 264 and Sourced From *re3data.org*

URL: http://repository.edition-topoi.org/collection/WOLL

Without a doubt, my choice to select the repository WOLL (also titled "The Textile Revolution") from *re3data.org* is rooted in my academic background in archaeological science alongside my own personal passion for textile arts. At home, I collect and prepare a variety of natural fibers for spinning and plying into yarn—followed by thrashing, dyeing and warping them to be made into complex fabrics on my handloom. It's an arduous process which teaches patience, lends self-accomplishment, connects artists to living culture—yet also provides a hands-on lesson in the human experience prior to industrialization and the mass-production of goods. Generally, among archaeologists, material culture is prized when its origin predates the advent or mention of itself in writing systems. As such, "The Textile Revolution" includes "archaeozoological and geo-archaeological types of evidence from material culture between 6500 - 1500 BC in the Near East and Europe attributing the innovation to the spread of sheep husbandry and wool processing" (WOLL). To find WOLL in the first place, simply enter into *re3data.org*'s keyword search the word "textiles", and select the matching result.

The collection scope of the repository WOLL is extremely narrow, and pertains to four research objects curated by the professional team devoted to "The Textile Revolution" project, from Exzellenzcluster Topoi. So, the repository is closed to general submissions—last updated in 2019—only having been contributed to by the researchers Ana Grabundzija, Hans Christian

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

Küchelmann, Martin Park, and Chiara Schoch.WOLL's first database is a documentation (description and data tables) of all archaeological artifacts pertaining to wool use sourced between Southeast Europe and West Asia between 3700-3200 BC (WOLL). The second is a pollen-specific database, followed by a bone-specific database and a database focused only on collected textile tools, all from the same region and period. Each of those databases includes a .JSON file of its metadata, as well as data tables and images with descriptions to bannerhead the dataset.

To maintain scholarly integrity, it should be noted that the open-access German platform Edition Topoi demands a peer review process to precede any data uploads. As per its internal standard, Edition Topoi lists the repository"The Textile Revolution" as part of a singular research project, with an end date, so no guidance is given to any potential data submitters regarding what a 'Submission Information Package' should contain, as everything has been pre-ordained by the research team whose project objective was to "comparatively investigate" existing data on assemblages in the manner they saw fit (WOLL). So, apart from the contact email edition@topoi.org being listed on *re3data.org*, and the repository's 'Overview' page's sidebar listing dedications to partnering institutions (the Freie Universität Berlin, specifically Excellence Cluster Topoi EXC 264, as well as the Deutsche Forschungsgemeinschaft, and the Deutsches Archäologisches Institut) combined with the fact that the website itself has a DOI (10.17171/2-17) for its "*pid* System", there is no venue for human assistance or consultation about additive submissions (WOLL). Because there is nothing left to be contributed to the

project, there is no requirement for metadata to be submitted in any given structure (or to any specific standard), although all four databases in the repository rely on .CSV data tables and .JSON metadata files. Simply because “The Textile Revolution” is open-access (along with its parent institution Edition Topoi), public dissemination of data without any restriction is possible, as long as resulting data use doesn’t violate their license for Creative Commons BY-NC-SA 3.0 DE (WOLL). To that end, logins are not even offered as a service on “The Textile Revolution” website—perhaps this is why *re3data.org* lists WOLL as a “data provider” only (re3data.org)!

The one access point for downloading data is through WOLL’s “Search” pathway, which presents all data as thumbnails below a search bar for queries. As well, there is a sidebar which offers a download of all data objects as a .CITE file. Actually, all files are alternatively available as .CITE files within the “The Textile Revolution” repository. While the metadata featured on WOLL adheres to only two standards—specifically, ‘CSV on the Web’ (CSVW) standards and relying on .JSON for data exchange—it is internally consistent with other metadata published through the German Research Foundation, listed as the Deutsche Forschungsgemeinschaft above. Due to the low amount of data within a highly controlled collection scope, there is no ‘Dissemination Information Package’ available, only direct downloads of .CSV and .JSON files for interested individuals willing to comply with the fair use conditions of a creative commons license. Seeing as WOOl is a scholarly project, that seems appropriate.

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

For such a tiny repository, WOLL is framed in an aesthetically pleasing manner (with images and accessible website navigation), organized well, and labeled thoroughly, which is something a textile artist—or artist or any sort—may appreciate in a sea of hard data.

Matthew Mayernik

LIS 545 B: Data Curation I

28 February 2023

Works Cited

Grabundzija, Ana, et al. “The Textile Revolution.” Edition | Topoi, Excellence Cluster Topoi, http://repository.edition-topoi.org/collection/WOLL.

“Hawaiian Monk Seal.” NOAA Fisheries, U.S. Department of Commerce, 15 Sept. 2022, https://www.fisheries.noaa.gov/species/hawaiian-monk-seal.

“Home.” Registry of Research Data Repositories, re3data, https://www.re3data.org/.

Littnan, Charles. “HMSRP Hawaiian Monk Seal Telemetry Tag Deployments.” *Data.gov*, U.S. General Services Administration, Technology Transformation Service, 28 Oct. 2022, https://catalog.data.gov/dataset/hmsrp-hawaiian-monk-seal-telemetry-tag-deployments.