NABONE TO NABONOSEAD

Report by Tom Ryan, CUNY, January 2021

In order to modernize and upgrade the NABONE Zooarchaeological Database (also referred to as the NABONE Zooarchaeological Analysis Package, hereafter as NABONE), the decision was made to migrate all existing data into the Strategic Environmental Archaeological Database (SEAD) and provide for ongoing data entry into the SEAD system. This alternative was chosen over the option of creating an entirely new, standalone NABONE system. The project is currently referred to by the working name NABONOSEAD. The current NABONE system consists of a Microsoft Access database, Excel spreadsheets for providing analytic output, and a coding manual (NABONE Manual). SEAD is a web accessible database using PostgreSQL. The initial motivation for this project was concern over the long-term sustainability of the Microsoft Access based system for zooarchaeological data. However, the project has expanded to incorporate other goals.

The aims of the NABONOSEAD project are to deliver sustainable upscaling of the Access zooarchaeological database, open access to the data online and through an API, and efficient and portable data storage. The move away from a Microsoft Access system will provide for a standardization of data entry (standardization of metadata) and help to prevent data entry errors. It will allow the NABONE data to be more easily linked to other systems, including the DataARC system, as well as giving the opportunity for structured development of common analysis tools, including more powerful tools using R and Python.

History of the NABONE System

Development of the NABONE system began in January 1997 at a meeting of 22 zooarchaeologists from Canada, US, UK, and Scandinavia specializing in North Atlantic collections in New York, as part of a US NSF-funded effort to improve data comparability and curation in this area (NABONE Read Me Notes). At this meeting, the decision was consciously made to allow each user to make their own modifications to the package to suit their own approaches (NABONE Read Me Notes). At the time this decision was made, the challenges these modifications made for data comparability were believed to be outweighed by the importance of reaching an agreement between different groups of end users. A team from the City University of New York working out of the Hunter Zooarchaeology Lab was charged with development and the difficult need to balance that desire for comparability with the competing desire for independence among the various zooarchaeologists. Ultimately, the largest group of users are researchers who started their careers in the Hunter lab. It must be considered that the compromises made with respect to comparability may have hindered broader adoption of the system rather than encouraged it. The ability to compare datasets with ease may have drawn in more users excited by this prospect. This balance remains a challenge for similar projects today, including NABONOSEAD: will researchers give up familiarity and some degree of control in order to gain the unfamiliar advantage of joining a broader community with access to significantly more data?

For NABONE, Microsoft Access was chosen for storing and updating of data. However, concern over size limitations led to the recommendation that a different Access DB was used for each project (NABONE Read Me Notes). Excel was chosen for the analytic output due to the capability for easier graph production (NABONE Read Me Notes). In the Access DB was a single table called NABONE into which all data would be entered, and which drove the included queries. This table contained no input controls, so the entered data needed to be carefully proofread to prevent data entry errors. Additionally, situations not covered in the manual led to ad hoc development of codes. And, as stated earlier, different users were encouraged to modify the system to their own uses. As of 2020, the system has had 9 editions with a 10th edition in progress. Among these various editions, each user has made their own individualized modifications to the system. At this time, no plans have been developed to change NABONE's use as a data entry system, which means at least for the time being, users will still be able to modify the system in non-standard ways. However, these non-standard practices will not be uploaded to the DataARC system. It is hoped that in the future the data entry system will be updated, and such changes can be managed through a more formal "change request" system, which will lead to a dialog with the users requesting modifications. Through such a formal system, modifications can be discussed with users, allowing the opportunity to present a variety of solutions, which can then be incorporated into a standardized data entry system. But, at the current time, the continued use of a legacy system, in addition to the already existing legacy data, present challenges as data is migrated into the new system.

The NABONE Access database contains a single table (essentially a flat file but implemented in Access) called NABONE. This table does not conform to data normalization standards. This lack of normalization can lead to confusion, particularly in areas of site structure and phasing. Additionally, some important information is not stored in the NABONE database at all but must instead be found in excavation reports or similar documents. The NABONE database also contains a set of queries which aid in creating the exports that are fed into the Excel spreadsheets for calculations. Each individual user often adds their own queries tailored to their interests. The NABONE table itself is not fully documented outside of the database itself, instead relying on the Description field of the Access database. Attached to this document is an Excel spreadsheet listing the relevant information contained in the NABONE system, its source, and its destination in SEAD.

This system is still in use today, primarily, as stated above, by researchers from the Hunter lab. This newly generated data still presents challenges to users who are not the original data creators. For those outside the NABONE community, this data is difficult to access and understand. In addition to the difficulties in accessing and using this data, the interests of the community itself are shifting. Archaeologists are increasing thinking about larger scale synthesis research as outlined in the 2014 article "Grand Challenges for Archaeology" (Kintigh et al., 2014).[*]

Storing data from individual sites in separate, differing databases creates barriers to participation in such synthesis projects. The original NABONE project envisioned a community of practice made up of zooarchaeologists specializing in the North Atlantic region. However, the "Grand Challenges" paper envisions a community of practice that includes not just zooarchaeologists, but rather all archaeologists and those working in other disciplines on related projects (e.g., geographers, ecologists). While at one time, the NABONE system conformed with how its users worked and what they were interested in, the gap between the infrastructure and practice is widening quickly. The ability to participate in this envisioned, broader community will likely act as an incentive to researchers to employ the new infrastructure despite its constraints on individual variation. The hope is that a community may be formed around the tool

---

[*] Kintigh, Keith W., Jeffrey H. Altschul, Mary C. Beaudry, Robert D. Drennan, Ann P. Kinzig, Timothy A. Kohler, W. Fredrick Limp et al. "Grand Challenges for Archaeology." Proceedings of the National Academy of Sciences 111, no. 3 (2014): 879-880.

instead of attempting to define the tool by the currently existing community, as was done in the case of the initial NABONE design.

Work Accomplished Under the DataARC Project

The NABONOSEAD Project was conceived with four implementation packages. Initially, these packages were intended to be consecutive stages. However, as discussed below, work on much of the third implementation package was deferred to a future phase of the project.

Implementation Package 1

During Implementation Package 1, the focus was on identifying problems and brainstorming potential solutions to those identified. Primarily, the SEAD structure was checked for compatibility with the NABONE structure, asking the following questions:

- Could SEAD hold NABONE data at the species level and allow for needed aggregations?
- Could SEAD map and store NABONE metadata correctly?
- Did SEAD have the capacity to hold the needed lookup/dimension data?
- Would SEAD allow the queries necessary to perform typical NABONE analyses without relying on the separate EXCEL templates?

As a conceptual step rather than a practical step, this first implementation package was completely relatively swiftly within days of the conception of the project in May 2018 and all raised questions were answered in the affirmative, with some caveats that did not prevent moving forward with the project. NABONE's species level data was analogous to the insect data around which SEAD was initially formulated. Bones could be described with a species, element, and count along with a location where the bone was found. Taphonomy (burning, butchery, gnawing, etc.) was also already accommodated within the SEAD system as modifications.

During this implementation step, it was noted that SEAD employed an ecocode system, which allowed for additional descriptive information about each species in the system. By using the existing ecocode system, it was hoped that the aggregations seen in the existing NABONE system could be recreated. It appears that all these aggregations cannot be replicated solely by relying on ecocodes but, as described below, the capabilities of NABONE could be expanded beyond what is currently, easily achievable in the NABONE system

However, it was also decided that some work would be left out of these four initial implementation packages. The system will need some modification in order to incorporate certain measures such as:

- Bone metrics
- Minimum Animal Units (MAU)
- Modified General Utility Index (MGUI)
- Density measures
- Tooth wear analysis.

These modifications will likely involve something analogous to the ecocode system, allowing the addition of descriptions of particular bone elements in the lookup tables. This data is primarily of interest to zooarchaeologists, so for the purposes of the DataArc project, with its goal of appealing to a broader slice of users, this step will be left for the future. Accommodating metrics, MAU, MGUI, density measures, and tooth wear will likely require a change to the underlying SEAD database, whereas other migration steps only required the addition of data into the existing structure. Meanwhile, the DataARC system can act as a guide to those underlying datasets with their included calculations.

Implementation Package 2

The goal of Implementation Package 2 was to import the lookup/dimension data from NABONE to SEAD, with a focus on taxa, bone elements, and taphonomy. Much of this work occurred during the first quarter of 2019. However, additions and modifications have been continuous with each new dataset loaded. This work required:

- Identifying a source for the data
- Creating distinct lists of lookup data
- Importing them into the SEAD system

The below table (Table 1) indicates the source of the lookup/dimension data in NABONE and its destination table and field in the SEAD system.

Table 1- Source and Destination of Lookup Data

Origin	**NABONE Field**	**SEAD Destination Table**	**SEAD Destination Field**
Species Codes	From Manual*	tbl_taxa_tree_master	taxon_id
Bone Elements	From Manual*	tbl_abundance_elements	abundance_element_id
End	From Manual	tbl_modification_types	modification_type_id
Fragment Size	From Manual	tbl_modification_types	modification_type_id
Fusion State	From Manual	tbl_modification_types	modification_type_id
Butchery	From Manual	tbl_modification_types	modification_type_id
Burning	From Manual	tbl_modification_types	modification_type_id
Gnawing	From Manual	tbl_modification_types	modification_type_id
Age Estimate	From Manual	tbl_modification_types	modification_type_id
Sex	From Manual	tbl_modification_types	modification_type_id

**Additional data had to be derived from the relevant fields in the NABONE table in various databases due to non-documented additions to these fields.*

The lookup/dimension table had to be derived from the Word document manual rather than from tables in the existing NABONE database as such lookup tables are not present in the NABONE Access database. It is also worth noting that the manual did not include complete sets of all the lookup/dimension data required and additional data for these tables had to be extracted from the individual NABONE databases. Both the taxa and element tables have required additions as new data are imported into the system, because individual users have included less frequently occurring taxa that are found on their sites. The same has occurred with bone elements, though to a lesser degree.

The basic unit of recorded information in SEAD is the abundance held in tbl_abundances and joined to lookup tables as shown below (Table 2). For the purposes of NABONE data, these abundances are the individual bone records held in the NABONE table. These individual bones are then described by various fields in that individual record, connected, in the SEAD system, either directly or through intermediate tables, to the lookup/dimension tables. Because each bone can have multiple modifications, this information is connected through the abundance_id as a foreign key in tbl_abundance_modifications.

Table 2- tbl_abundances

Field	Description
abundance_id	Unique identifier for each abundance
taxon_id	Identifies the species, joins to tbl_taxa_tree_master
analysis_entity_id	Specifies the location (context) of the find, joins to tbl_analysis_entities
abundance_element_id	Identifies the bone element, joins to tbl_abundance_elements
abundance	Count of the bones – usually 1, can be greater if multiple bones fit same description
date_updated	Date data in row was added/modified

Initial focus was on adding taxa, bone elements, and taphonomy information, which characterized each individual bone in the NABONE system. This information is of general interest across disciplines and structures for defining and storing this information are already in use across many disciplines. Therefore, determining where and how to store this information in SEAD was relatively easy. The taxa and element information fit directly into SEAD's existing systems in taxa tree tables (prefixed with tbl_taxa_tree) and tbl_abundance_elements. Taphonomy was entered into the modifications system, consistent with how it was already used in the SEAD system. These modifications are connected to a single lookup table, called tbl_modification_types, with data as described in Table 3.

Table 3 – tbl_modification_types

Field	Data Stored	Notes
modification_type_id	Unique identifier	Joins tbl_abundance_modifications
modification_type_name	Name per NABONE manual	
modification_type_description	Groups the modifications per NABONE manual	
date_updated	Date added or modified	

A cumbersome aspect of these element and taxa system was the decision to use the tbl_record_types table to distinguish between bone elements for fish, birds, and mammals. The tbl_record_types is used in SEAD primarily to distinguish between data that is traditionally collected by different specialists. This decision was based on the original NABONE system manual, which had separate sections for bone elements from fish, birds, and mammals. However, during the transition from the paper manual, Excel spreadsheets, and Access databases to the normalized PostgreSQL SEAD database, this separation became cumbersome while serving no useful purpose. This choice has led to the unnecessary repetition in the table tbl_abundance_elements of those bone elements which occur in multiple species. The current plan is to remove this distinction between each species group, including each bone element only a single time i.e., mammal humerus and bird humerus will be collapsed into humerus. The initial process of loading legacy data was designed based on this system, so the wisest strategy was to forge ahead with those species distinctions in place. However, with the current break in loading legacy data, it is a priority to remove this distinction to allow easier analysis of the data.

In addition to the lookup data anticipated during the initial planning phase, during the data load process other data elements were found that required lookup data. These data were added in this course of the second implementation package. Much of this information was not documented in the NABONE manual nor did it exist in the NABONE databases or spreadsheets. Rather it had to be found in excavation reports. Methods by which each dataset was generated and dated were recorded as well as the sampling method used in each context. Another table is used to describe site locations, where multiple locations can be attached to a single site (e.g. Skútustaðir in Iceland and Northeast Region and Skútustaðahreppur). This additional lookup data is shown in Table 4.

Table 4 – Lookup Data Not in Manual or Database

Information	**Source**	**SEAD Destination Table**
Site Location	Published Excavation Reports	tbl_locations
Sampling Method	Published Excavation Reports	tbl_sample_types
Dataset Method	Published Excavation Reports	tbl_methods

The ecocode system that existed in SEAD had no direct parallel to any data existing in the NABONE system. Ecocodes were originally added to SEAD to specify preferred environments for the insect species recorded in SEAD (i.e in swamps, in meadows, etc.). These ecocodes were expanded to include behaviors of certain insects, indicating, for example, that they feed on a particular plant, or eat worms, or feed on a plant's inner parts. As a method for recording behaviors and environments, it was apparent that adding ecocode lookup data relevant to NABONE would expand the analytical possibilities of the system. This data had to be created based on perceived requirements and desires of the zooarchaeologists analyzing this data, rather than simply derived from the existing datasets. The ecocode system is also easily expandable as researchers contribute their ideas on what is of interest to them. The current ecocodes added for NABONE data are shown below (Table 5) with the definitions being self-explanatory.

Table 5 - Ecocode Definitions

ecocode_definition_id	**definition**	**ecocode_group_id**
161	Non-Migratory Bird	8
160	Migratory Bird	8
159	Marine Mammal	8
158	Freshwater Fish	8
157	Marine Fish	8
152	On fast ice	8
151	Wild	8
153	On floe ice	8
156	Non-Migratory Terrestrial	8
155	Fresh Water Migrant	8
154	Sea Bird	8
148	Aquatic	8
149	Terrestrial	8
150	Domestic	8

Plans include adding lookup data for species associations, describing how various species interact with each other. Incorporating these tables will allow the generation of food webs and use webs from the existing NABONE data. There is no timetable for when this lookup data will be added to the system.

Implementation Package 3

Implementation Package 3 was broadly planned with the goal of replacing the NABONE outputs, which are created by Excel spreadsheets fed by the data within the NABONE Access databases. The planned steps were:

- Map Excel formulae to SQL in PostgreSQL
- Create views
- Use SEAD infrastructure to create GeoJSON files for dataARC
- Use the SEAD API to expose those GeoJSON files.

However, this task proved too large and complex to be completed at this stage of the project. For an example of this complexity, Figure 1 below shows a section of the output relating to a single species, *Bos taurus*, from a single context at the Skálholt site. The Excel spreadsheets depend on subsets of data being exported from the Access database into those spreadsheets to give a category of results which are then combined into a single report. To provide these results in the NABONOSEAD database would have required, in addition to the reverse engineering of the code in both the Access database and Excel spreadsheets, involvement of the end users to understand the process they followed to produce their desired results. A simpler dataset was created for output, based on the ecocode system described above. The work on creating an appropriate export file occurred primarily during the summer of 2019. Figures 2 & 3 show an example of this simplified data output. Because the NABONE dataset based on the ecocode system is similar to the SEAD output, the SEAD infrastructure could successfully be used to output this data. No attempt has been made to expose this set with the SEAD API to date.

For the future, conversion of the Excel templates with their included formulae remains an important goal, without which the conversion will be of limited use to the data providers. While the hope is to deliver significantly more functionality, the base functionality of the existing NABONE system will have to be recreated before the current users of NABONE can convert to and rely on the new system.

Figure 1 – Sample of Output from Skálholt

MEASURES OF ABUNDANCE							
		notes	*For Caprines Only!*	NISP	% NISP		
NISP	**883**	*Identified frags.*	Sheep	0	0.00		
NISP/Volume (DD)	**#DIV/0!**	*bones/cu cm*	Goat		0.00		
MNI (MAU max)	**88.50**	*simple MNI*	NISP/MNI	9.98			
MAU mean (RF)	**5.46**	*nean MAU= Perkins RF*	MNI % mean	6.25			
MAU Stand. Dev.	11.76		MNI% Stand Dev	13.36			
MAU sum	409.44		MNI % Coef.Var.	46.79			
MAU Coeff. Variation	46.42		MNI/Volume	#DIV/0!	*MNI/cu cm*		
TAPHONOMIC INDICATORS							
Bone Density Measures			**Dense/ Soft End Ratios**				
Binford mean density			D Fem/P Fem	1.52			
Binford mean D * MAU	7.43		D Hum/P Hum	1.31			
Binford (mean D * MAU)/NISP	0.008		D Tibia/P Tibia	0.23			
Binford Mean MGUI*MAU	179.04		Phalanx 1/Ph 3	6.50			
			Bone Density by Quartile				
Teeth % NISP	7.59			# non-zero	sum MAU		
Whole Long B % NISP	2.60		Most dense (1st)	20	359.77		
Long Bone Shaft % NISP	6.12		Dense (2nd)	17	102.99		
% MNI Indicators			Less dense (3rd)	18	78.65		
%MNI, P Hum	14.69		Least dense (4th)	20	15.72		
%MNI, D Hum	1.13				557.12	0.63	
%MNI, P Tibia	14.69		**Bone MGUI By Quartile**				
%MNI, D Tibia	3.39			# non-zero	sum MAU		
			Richest (1st)	19	7513.99		
FRAGMENTATION	Count NISP	%	Rich (2nd)	17	2843.45		
<1 cm	0.00	#DIV/0!	Less Rich (3rd)	21	691.821		
1-2 cm	0.00	#DIV/0!	Poorest (4th)	18	1483.50	mau MGUI/NISP	
2-5 cm	0.00	#DIV/0!		total	12532.8	14.19	
5-10 cm	0.00	#DIV/0!	**Density and MGUI % MAU Comparisons**				
>10 cm	0.00	#DIV/0!		density	MGUI		
total	0.00	#DIV/0!	Quartile Rank	%MAU	%MAU		
			1st	64.58	59.95		
AGE	Count NISP	%	2nd	18.49	22.69		
Neonatal	0	0.00	3rd	14.12	5.52		
Old	0	0.00	4th	2.82	11.84		
SKELETAL ELEMENT DISTRIBUTION			**Selected elements**				
	MAU count	Relative % MAU		NISP	MAU	% MAU	%MNI
All Cranial	149.50	35.48	Horn Core	0	0.00	0.00	0.00
Mandible	24.00	5.70	Mandible	48	24.00	10.95	27.12
Forequarter	81.50	19.34	Atlas	17	17.00	7.76	19.21
Vert & Ribs	39.00	9.25	Axis	9	9.00	4.11	10.17
Hindquarter	78.00	18.51	Scapula	36	18.00	8.21	20.34
Lower Forelimb	13.33	3.16	Pelvis	35	17.50	7.98	19.77
Lower Hindlimb	33.17	7.87	Hum Pro	26	13.00	5.93	14.69
Feet	2.92	0.69	Hum Dis	34	17.00	7.76	19.21
total MAU	421.42	100.00	Fem Pro	27	13.50	6.16	15.25
			Fem Dis	41	20.50	9.35	23.16
MARKING			Ulna	14	7.00	3.19	7.91
	#	%NISP	Rad Pro	24	12.00	5.48	13.56
Chewing	0.00	0.00	Rad Dis	12	6.00	2.74	6.78
Not Chewed	883.00	100.00	Tib Pro	26	13.00	5.93	14.69
Butchery	169.00	19.14	Tib Dis	6	3.00	1.37	3.39
No Butchery Marks	714.00	80.86	Calc	9	4.50	2.05	5.08
			Astrag	6	3.00	1.37	3.39
BURNING			Mtc Pro	10	5.00	2.28	5.65
	#	%NISP	Mtc Dis	5	2.50	1.14	2.82
White calcined	1.00	0.11	Mtt Pro	19	9.50	4.33	10.73
Black charred	1.00	0.11	Mtt Dis	6	3.00	1.37	3.39
Spotted scorched	145	16.42	Phalanges	28	1.17	0.53	1.32
Unburned	736.00	83.35	*sum*	438.00	219.17	100.00	

Figure 2 - Output from NABONOSEAD (Part 1)

physical_sample_id	domestic	wild	Marine Mammal	Marine Fis	Freshwater Fish	terrestrial	aquatic	Sea Bird	Non Migra	Fresh Wat	On floe ic	On fast ice
7838	0	12	0	0	0	0	0	0	0	0	0	0
8195	1	0	0	0	0	0	0	0	0	0	0	0
8290	0	4	0	0	0	0	0	0	0	0	0	0
7809	1	1	0	0	0	0	0	0	0	0	0	0
8251	2	3	0	0	0	0	0	0	0	0	0	0
7976	13	4	0	0	0	0	0	0	0	0	0	0
8154	61	980	0	24	781	0	0	0	0	0	0	0
7872	2	2	0	0	0	0	0	0	0	0	0	0
8684	11	31	30	0	0	0	0	0	0	0	0	0
8352	0	0	0	0	0	0	0	0	0	0	0	0
8773	0	36	3	0	0	0	0	0	0	0	0	0
7954	8	6	0	0	0	0	0	0	0	0	0	0
8115	4	2	0	0	0	0	0	0	0	0	0	0
8138	607	2734	2	255	1498	0	0	0	0	0	0	0
8178	10	20	0	0	0	0	0	0	0	0	0	0
8137	0	192	0	0	136	0	0	0	0	0	0	0
7864	1	3	0	1	0	0	0	0	0	0	0	0
7927	0	3	0	0	0	0	0	0	0	0	0	0
8298	0	0	0	0	0	0	0	0	0	0	0	0
7816	1	0	0	0	0	0	0	0	0	0	0	0
7870	3	32	0	2	0	0	0	0	0	0	0	0
8055	0	3	0	0	0	0	0	0	0	0	0	0
8069	9	8	0	0	0	0	0	0	0	0	0	0
7751	3	0	0	0	0	0	0	0	0	0	0	0
7790	0	0	0	0	0	0	0	0	0	0	0	0
8304	0	1	0	0	0	0	0	0	0	0	0	0
8822	1	20	13	0	0	0	0	0	0	0	0	0
8135	262	967	0	291	72	0	0	0	0	0	0	0
8233	102	134	0	5	23	0	0	0	0	0	0	0

Figure 3 - Output from NABONOSEAD (part 2)

location_r	site_id	site_name	latitude_c	longitude_	age_older	age_young	age_name	age_abbreviation	sample_gr	sample_ty	sample_name
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	900	I-II	HOF_Phase_I-II	1426	16	427
Iceland	1293	Hofstaðir	65.608	-17.1638			No Data	HOF_Phase_No Data	1463	16	1679
Iceland	1302	Sveigakot	65.51543	-17.0288	1020	900	II	SVG_AU_II	1473	16	7
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	900	I-II	HOF_Phase_I-II	1425	16	2403
Iceland	1302	Sveigakot	65.51543	-17.0288	1020	900	II	SVG_AU_II	1467	16	NO SU
Iceland	1293	Hofstaðir	65.608	-17.1638	950	900	II	HOF_Phase_II	1439	16	1317
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	950	I	HOF_Phase_I	1459	16	7e
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	950	I	HOF_Phase_I	1430	16	367
Greenland	1288	Brattahlið (E29N	61.15076	-45.502	700	650	Phase IV	BNF_Phase_IV	1496	16	67
Iceland	1302	Sveigakot	65.51543	-17.0288	1080	1020	I	SVG_AU_I	1475	16	2177
Greenland	1310	Qorlortorsuaq	60.76791	-45.2312			Phase 2	E74_PH2	1501	16	12
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	95	VII-VIII	HOF_Phase_VII-VIII	1434	16	219
Iceland	1293	Hofstaðir	65.608	-17.1638	900	100	IV-VII	HOF_Phase_IV-VII	1457	16	297
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	950	I	HOF_Phase_I	1459	16	6hk
Iceland	1293	Hofstaðir	65.608	-17.1638			No Data	HOF_Phase_No Data	1461	16	1829
Iceland	1293	Hofstaðir	65.608	-17.1638	950	900	II	HOF_Phase_II	1459	16	6H
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	950	I	HOF_Phase_I	1430	16	288
Iceland	1293	Hofstaðir	65.608	-17.1638	950	900	II	HOF_Phase_II	1432	16	566
Iceland	1302	Sveigakot	65.51543	-17.0288	900	840	III	SVG_AU_III	1473	16	1144
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	900	I-II	HOF_Phase_I-II	1425	16	591
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	950	I	HOF_Phase_I	1430	16	324
Iceland	1293	Hofstaðir	65.608	-17.1638			UNSTRATIF	HOF_Phase_UNSTRATIFIED	1452	16	UNSTRATIFIED
Iceland	1293	Hofstaðir	65.608	-17.1638	950	900	II	HOF_Phase_II	1454	16	1137
Iceland	1293	Hofstaðir	65.608	-17.1638	900	895	III	HOF_Phase_III	1424	16	160
Iceland	1293	Hofstaðir	65.608	-17.1638	1020	900	I-II	HOF_Phase_I-II	1425	16	2054
Iceland	1302	Sveigakot	65.51543	-17.0288	1080	1020	I	SVG_AU_I	1473	16	1463
Greenland	1311	Tatsip Ataa	60.80679	-45.5298	1050	1150	Phase 2	E172_2	1504	16	10
Iceland	1293	Hofstaðir	65.608	-17.1638	950	900	II	HOF_Phase_II	1459	16	6f
Iceland	1302	Sveigakot	65.51543	-17.0288	1080	1020	I	SVG_AU_I	1467	16	8

Implementation Package 4

Implementation Package 4 involves the importation of NABONE data into the new NABONOSEAD system. This stage has proved significantly more time consuming than initially anticipated due to variability among the many NABONE Access databases. The work completed to date occurred in several stages in the spring of 2019, the fall of 2019, and the summer of 2020. The first step of the process involves transferring data related to the location and physical structure of each site as well as connecting bibliographic information to the site and its datasets as detailed in Table 6. There is no central repository for this data and it must generally be pulled together from site reports which are not mentioned in the NABONE databases i.e., the connection made explicit in SEAD must be inferred from information in the databases and the reports.

Table 6 – Structure and bibliographic Information

Data Description	**Destination**	**Notes**
Site information	tbl_sites	Joins to location table
Creator/Excavator/Researcher	tbl_biblio, tbl_contacts, tbl_dataset_contacts	Usually the same in all 3 tables but can be different
Dataset description	tbl_datasets	Joins to creator information tables

Capturing the physical structure of the site proved to be far more complicated than originally planned. The fields in NABONE which store this information are used inconsistently from researcher to researcher and from site to site depending on the complexity of that site. It is almost universally true that the SU field in the NABONE database represents the archaeological context, the finest gradation of site structure found in the NABONE database. Above this level, the variations become more frequent. Two fields (Unit and AU) are used to hold higher level structure if it exists. Generally, Unit represents a distinct site area (a trench, for example) while AU (Analytical Unit) is used for phasing information. However, information expected in one of the fields occasionally appears in the other. For simpler sites, no information may appear in the fields at all. Additionally, an SU might appear in more than one Unit in the NABONE data, which is difficult to accommodate in SEAD (and is likely a data entry error). In some instances, the information expected to be found in Unit or AU had to be found in site reports instead of the database itself. Unfortunately, it must be admitted that these irregularities increase the chance of mistakes or misinterpretations in the data loading process. Where the data fits expectations, it is transferred as shown in Table 7.

Table 7 – Site Structure

NABONE Field	**SEAD Destination**	**Notes**
SU	tbl_physical_samples	For NABONE Data SU appears in both fields, but this is not a system requirement.
SU	tbl_analysis_entities	
Unit	tbl_sample_groups	
AU	tbl_relative_dates	Joins tbl_relative_ages and tbl_analysis_entities

Once the structure of the site is in place and the bibliographic information is recorded, the actual bone data can be loaded into SEAD. It is in this step of the process where individual variability causes the greatest difficulties. First, the species of each recorded bone must be determined. NABONE uses a code-based system with codes assigned to many of the species encountered in the North Atlantic. This code system is not exhaustive, however, and the recommendation is for the person entering data to use the full scientific name where a code does not exist. For the sake of brevity, people regularly make up their own non-standard codes. These codes can often be deciphered by referring to other information in the record, consulting with the original creator, or another zooarchaeologist familiar with the system. But again, this process is prone to error and misinterpretation, even when the original data creator can be consulted. In cases where the original species is indecipherable, it is simply entered as an error in the system. Once the species has been properly converted so that it can be entered into SEAD, the bone element must be addressed. This process tends to be more straightforward than the species as the list of possible bone elements is much shorter than the list of possible species and thus there is less reason for a researcher to create their own code. However, both bone element and species contain data entry errors (simple typos) that must be corrected. If the bone element is indecipherable, it too is captured as an error. Even if the bone element is not usable, the species information can still be used. The count of bones is simply transferred to a count field in SEAD. See Table 2 above for field mapping information. It is worth noting that these individual variations will cause some data loss even when the researcher is interested in only their individual site. However, for the purposes of synthetic research, the types of "grand challenge" projects that are become more common, these individual variations can be prohibitively confusing, preventing sites from being included in these larger studies.

After data is added to tbl_abundances, the taphonomy associated with these bones (abundances) is loaded. In the NABONE table, the taphonomy is stored in a sequence of fields each representing a single facet of the process (End, Burning, Gnawing, etc.). In SEAD, these individual fields are all combined into a single table, tbl_abundance_modifications, and connected to the lookup table tbl_modification_types (Table 3 above).

Implementation Package 4 included plans to automate this data load process. However, because of the complexities described above, such a process has not yet been developed. While some routines have been created to handle small parts of the larger process, each stage requires a great deal of manual intervention, sometimes including consultations with the original data creators or other zooarchaeologists. When this occurs, not only is there a break in the process from a technical standpoint, but there is also often a temporal break, waiting for communications back and forth.

Future Implementation Stages

Future implementation stages were also discussed during planning, some of which must happen going forward and others which will not be happening for the foreseeable future. In the former category are a plan for the storage and management of the NABONOSEAD system. The system will require a hosting location and an administrator. A plan for quality control beyond the efforts of the person converting the data will have to be developed as well. In the latter category is the design and development of a data entry system. For now, data entry will still be handled through the existing Access DB, hopefully with some improvements to control the quality of data entered into the system. At some point, an interface independent of DataARC will likely also be developed, to give zooarchaeologists views of their data that are of specific interest to themselves but may not reach a wider audience.

Additionally, as discussed above, some steps were moved from their originally planned implementation stage to some future stage. And other ideas for the future arose during the course of the project. These include:

- Additional lookup/dimension data to support MAU, MGUI, and density measures
- Bone Metrics
- Tooth wear analysis
- Additional ecocodes developed
- Species associations for food webs and use webs
- Removal of record types from bone element data
- Queries to replace the NABONE Excel spreadsheets
- Continued loading of data from Access databases
- Loading of older data from non-Access sources

While no time frame has yet been specified for these future steps, some of them, particularly those needed to recreate the outputs of the current NABONE system, are high priority. Without those steps the data will not be useful to zooarchaeologists, although it retains some value for other archaeologists and those undertaking synthesis research. Additionally, the outputs of the dataARC system will allow others to find NABONE data and perhaps seek access more directly rather than through the existing interface.

Summary

The NABONE system was developed over two decades ago, driven by the types of research zooarchaeologists undertook. As research goals have changed, the NABONE system no longer meets those needs. A system more accommodating to synthesis research is required as the discipline moves forward. Such a system must start with the consolidation of all NABO data in a single data source rather than spread over dozens of individualized databases.

Much progress was made in moving data into the single database based on the SEAD system with the data of a dozen sites moved into this new system. The geographic range of this data covers Greenland, Iceland, and the Faroe Islands. However, the difficulty in creating a general process to load this data into the new system was one of the most important barriers during development. Because the NABONE system has gone through multiple iterations over the years and is used by multiple researchers all able to modify the base system, extensive human involvement is required to make sense of all the variations in data and format. This condition means that adding legacy data to the system will continue to be a time-consuming process.

Also important to note, the more general information such as species was easier to move to the new system than information of more specific interest in zooarchaeology, such as bone densities and tooth wear analysis. While the ease of adding this general information is positive in that it is likely to be off wider interest and use in synthesis research, it is important to find a way to accommodate the information of specific interest to drive those specialists to include their data in the system, thereby making it available to that wider community.