NZRR: the next steps following the first workshop

David Orton and Eva Fairnell, University of York

Introduction

The workshop held at York on 16 May was an integral part of a Historic England (HE)-funded project with the aim of establishing the first phase of a coordinated National Zooarchaeological Reference Resource (NZRR), in order to improve the visibility and accessibility of existing reference collections and hence facilitate zooarchaeological research across all sectors. Hosted by the Archaeology Data Service (ADS) to ensure long-term sustainability and ease of updating, this resource will initially incorporate a central, searchable, database of zooarchaeological reference material held by a variety of stakeholders. On conclusion of the initial project, the lead institution (York) commits to undertake and fund annual updates to the database.

In addition, following the launch of the NZRR database, the zooarchaeology community will be consulted across all sectors, with the aim of establishing the requirements and priorities for digitisation of reference specimens. This will be used as the basis for future applications to various bodies, aimed at funding the development of a digital reference collection to be hosted as part of the NZRR, including downloadable files for offline use and printing as well as files suitable for online interrogation. The existence of the database will allow such applications to be made from a position of knowledge.

We thank the 44 people who attended the workshop. Sixteen universities, eight archaeological units, four museums and one wildlife trust, as well as HE, ADS and six independent researchers, were represented, some people with a foot in more than one camp.

The aims of this workshop were to:

- ascertain the features the sector requires of an online, searchable database of zooarchaeological reference collections
- develop a plan for achieving at least an initial version of this within the budget and timescale of the current project.

The workshop was very productive, with different views, ideas and comments shared and discussed. The aim of this document is to summarise the day's output (pages 4–21), and outline a plan for the next stage of the project (pages 2–3 plus the attached spreadsheet), thus fulfilling the aims stated above. This document is being disseminated to both those who were at the workshop and those who have expressed an interest in the project but were unable to attend.

The next steps

Given the consensus that the model proposed (Figure 1) was appropriate and feasible, and the pointers provided by the discussions on minimum data needed per collection, there are several tasks that both the project and curators can start to address.

In conjunction with the ADS where appropriate, the project coordinator and data manager need to resolve the:

- format in which curators should aim to provide their data
- format in which the project will supply the data to the ADS
- format by which searches will be carried out on the web interface
- format by which search results will be returned on the web interface
- source(s) to use to standardise taxonomy.

Curators of collections need to resolve:

- how to filter their data so that only appropriate specimens will be included in the data submitted
- how to organise their data ready for collation and upload by the data managers
- any legal issues, such as ownership and Defra licensing https://www.gov.uk/government/publications/licence-to-possess-plants-and-animals-for-scientific-purposes.

Some of these tasks can be addressed concurrently by the project and curators.

Participation

If you would like your collection to be part of the NZRR, please let the project know by emailing David Orton david.orton@york.ac.uk or Eva Fairnell eva.fairnell@york.ac.uk by **12 August 2016**. **Data are not needed at this stage.**

The file associated with this document provides the provisional names of the data fields that will be used. These fields are a distillation of all the points raised during the course of the workshop (see the summary tables pages 6–15). It is anticipated that curators will be asked to supply data in an Excel-style spreadsheet using standardised headers for the columns.

Those fields indicated with an asterix in the attached file are the minimum data fields required (the collection-level data); those indicated by a dagger symbol are desirable data fields; those indicated by a double-dagger symbol are fish-specific; fields without an asterix or dagger symbol can be filled in but are not essential. Some detail provided, for example actual ages, may be distilled to true/false categories, to provide, for example, yes/no search returns, rather than the full range of all possible entries. In most cases, it is hoped that a curator can match the data fields within his/her database, however that is held, to a relevant data field name given in the spreadsheet.

Further instructions will be provided to curators once the deadline for registering has passed. The file attached is for guidance and planning, although it can be filled in now if a collection is already in

order. It is possible to provide only collection-level data (the fields marked with an asterix). When registering interest, please let David or Eva know how easily you will be able to provide the collection and specimen data in an appropriate format, i.e. whether you will need further assistance or guidance and how much time you might need.

If you have any questions, thoughts or comments, please do contact David or Eva.

Summary of workshop

The day started with a presentation from David Orton (University of York and NZRR project coordinator) outlining various different ways in which an NZRR could work and concluding with an explanation and justification for the model proposed by the project. Figure 1 shows a schematic of this proposed model, and the Appendix shows the alternatives considered. The morning discussions considered (a) the extent to which an NZRR is needed; (b) what the best approach/structure should be; and (c) what collection-level data would need to be included as a minimum to meet the sector's requirement. The afternoon session began with a presentation from Polydora Baker and Fay Worley (HE) about the Access database they use to manage the extensive HE collection. Using a relational database, HE is able to track where each specimen is, how often it has been used, what condition it is in, etc. If a collection is to be promoted via the NZRR, then the more organised it is in-house, the easier it will be for curators to upload data to the NZRR and help potential end-users. The HE model provides an example of good practice, particularly for larger collections. Museum collections have their own database management systems, often built on commercially available platforms. The afternoon discussions centred on the minimum specimen-level data needed, and how to tackle concordance of taxonomy.

The attendees were overwhelmingly in favour of an NZRR, many stating that it would be a very useful resource, and the proposed model was supported. As the day progressed, there were conflicting views and suggestions regarding some issues, but all within the context of an agreed need for an NZRR. In the long-term, the model used now to start the NZRR can always be reviewed and refined, depending on uptake, demand, feedback and funding. If possible, the model used should include the flexibility for expansion of data fields in the future.

There should probably be different levels of data entry, from the most basic (essential) to more detailed (extra), so that collections need not be excluded because certain data cannot be provided easily, and to accommodate the peculiarities of individual collections. End-users can always contact curators for more information if the minimum data provided indicate that a collection holds specimens of potential interest.

The discussions inevitably overlapped in content. The following tables present themed summaries of the points made during the course of the day and further thoughts and comments arising. Many thanks to Andy Hammon, Jim Morris, Jacqui Mulville, Rebecca Nicholson, Naomi Sykes and Richard Thomas for facilitating the smaller group discussions, and to David Orton and Fay Worley for taking notes during the general discussions.

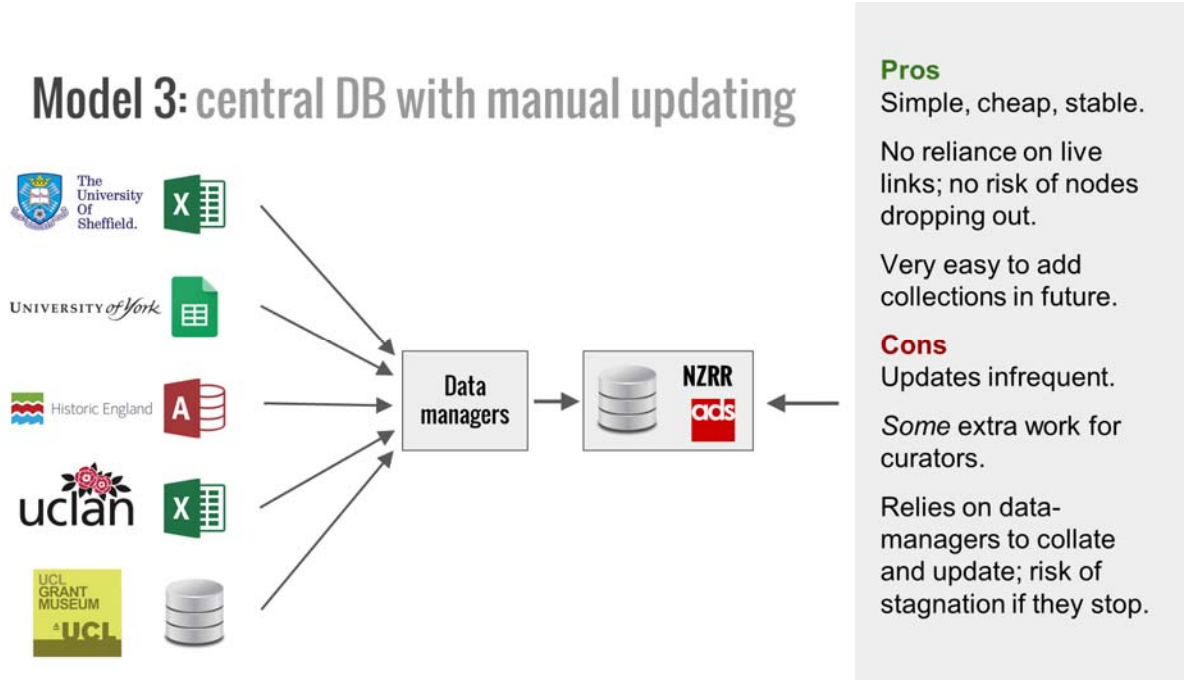


Figure 1: The model proposed by the project, to be coordinated by the University of York (project coordinator and data manager) and ADS (database host). NB: the five collections shown here are just examples selected from those already committed to the project.

Themed summaries

Positive feedback

Supporting comments	Further comments
Enhances the profile of zooarchaeological collections	
Promotes a hands-on approach to using collections	Even if, as hoped, the project leads to the scanning of images to fill in gaps in collections and fulfil any need for a virtual collection, handling specimens would always be important
Useful for people new to the profession: provides knowledge and access to less experienced end-users, overcoming patronage and encouraging relevant methodologies	
Promotes collaboration within the zooarchaeological community, and between regionally linked wider participants, for example museums and universities	
Cost-effective (time and money) for end-users: the closest resources can be found, rather than the most well known	
Makes it easier to find target specimens, for example more unusual species, a relatively large sample size of a species for targeted research, any species needed to verify an identification	
Could be a catalyst for new postgraduate research projects	
Useful for those with smaller resources, for example independent researchers and curators just setting up a collection	
Useful for curators to facilitate swaps and loans to fill in gaps in a collection	
Gives a collection a presence outside the holding institution, which could help combat 'orphaning' of collections if changes in staffing means there is no longer a curator	If a collection is registered with Natural History Near You http://natsca.org/NaturalHistoryNearYou, the Natural Sciences Collections Association (NatSCA) can potentially help if it is at risk
Raising the profile of a collection could help maintain/increase usage and so help justify curatorial costs through monitored usage	This could be a double-edged sword, the counterargument being the need to stop promoting something that incurs rising costs

Supporting comments	**Further comments**
Raising awareness of the collections could be of interest to other disciplines, for example forensic departments, veterinary schools, mammal societies, with the potential for collaboration	
Raising the profile of a collection could lead to investment in the development of local systems	
If successful, the model could be used by other special interest groups	
The proposed model is cheap and simple, and ADS has established connections and working practice within the world of archaeology	
The proposed model is the most realistic given the resources available, and also the most adaptable and future proof	
The proposed model makes the least demands on curators	

General issues

Issue	Possible solution and/or further thoughts
Start-up costs (time as well as money) could be an issue for curators	The cost (time/money/human resources) will vary depending on institutional set-up. The project needs to make the upload of data as straightforward as possible, and provide as much support as possible, to those institutions that require help organising their collection data. Student projects and work placements could be used to help curate collections and create up-to-date catalogues and databases
Commercial units may hold reference specimens but if no-one has the role of curator, no-one knows what the collection as a whole contains	Commercial participants were not initially expected to contribute collections to the database, although they are welcome to do so
Duplication of effort by curators	The project will do all it can to streamline the process, and some capacity has been allowed for the data manager to help curators. However, curators are encouraged to be able to sort their data to a greater or lesser extent, rather than just handing over unedited data. The project data manager may be able to help with setting up standard queries to automate data extraction for participants who use Access or similar software (e.g. the HE system as demonstrated by Fay Worley)
Need to be able to demonstrate to organisation funders and university managers that the project is of value and worth investment	
May be useful for postgraduates, but less so for undergraduates	
By raising profile, could be at risk of becoming a target for 'rationalisation' of an institution's time and money	Collaboration and research output will have to be obvious potential outcomes of the project
There could be different issues for public and private collections	The uploading and searching of data will have to be flexible enough to cope with different types of collection
To be really successful, the more significant collections, including museums, need to be part of the NZRR	
Presenting the right level of information will be crucial	

Issue	Possible solution and/or further thoughts
The output needs to be tailored to the end-user	Users of a zooarchaeological collection are predominantly zooarchaeologists but other potential end-users need to be considered
If the NZRR is very successful at promoting collections, some curators could then be overwhelmed by interest	An access policy will be important for managing expectations of how easy and quickly a collection can be visited, i.e. to control access. Responding to requests and promoting a collection could also be combined as an activity for open days, for example. However, a more constant low-level interest is more likely than an inundation
Will collections be used more? Commercial zooarchaeologists rarely have the time to visit collections outside their unit	The NZRR has the potential to support commercial specialists' requests to visit collections by indicating the nearest resources
Curators have to be committed to the NZRR if they submit data, i.e. be prepared to respond to requests from end-users	
There could be curation/conservation issues if the NZRR results in increased use of a collection	
Updating contact details could be an issue as a result of staff changes	Pros and cons were discussed for providing actual names, but because, for example, some university departments don't have a departmental secretary to field enquiries, it is probably better to provide at least one if not two named contacts. Should those individuals move on, they may still be trackable and have the best idea of who to contact in their absence/stead
Should unprovenanced, unprepared, archaeological, etc., material be included? Should multi-individual species and specimens in teaching collections be included?	These should probably be decisions made by curators for their collections: the data provided for the NZRR need only be a subset of a complete collection, avoiding the advertisement of more dubious specimens
Smaller collections may be too 'embarrassed' to contribute data	No collection is necessarily too small, but smaller collections could perhaps combine with a nearby larger collection and upload one dataset between them
If there is a successful NZRR, are small local collections necessary? Could be detrimental to setting up a new collection	The more parts to the whole, the better the NZRR will be. Not having a collection and therefore not being part of the NZRR could be used as an argument to obtain funding to start a collection, particularly if there is a regional gap

Issue	**Possible solution and/or further thoughts**
Access policy, regarding who can use collection, when and where, is something not every collection will currently have in place	The NZRR should probably have a neutral policy on this, not least because it will vary from collection to collection: it will be up to the curators to state an access policy as required. The access policy should also provide any expectations on the part of the curator regarding how research data arising should be shared
Disability access: if a collection cannot be easily accessed by wheelchair, for example, this should be stated somewhere	

Legal issues

Issue	Possible solution and/or further thoughts
Licensing/legislation could be an issue, perhaps more so for private collections	The NZRR should perhaps have a disclaimer stating that it is not responsible for each collection, and should perhaps only take licensed collections
Unprovenanced protected species	Curators will have to filter their data to avoid the inclusion of specimens that cannot be freely used
If items have been loaned rather than donated, there could be conflict of ownership and insurance issues	Providing a subset of data rather than the whole collection provides curators the opportunity to remove any potential contentious specimens so that they are not promoted via the NZRR
Ownership within the NZRR	The NZRR will not 'own' any data: ownership will remain with the curator/collection holder
Ownership within individual collections	Some collections may contain privately owned specimens. This needs to be clearly stated somewhere so that if staff move, for example, there is no dispute over ownership
Acquisition source	The original donor of a specimen might not want to be named, and this could be a breach of data security. Museums will not publish the names of donors
Access policies, policies for research data, copyright, licensing, etc.	The project could provide pointers, for example web links, to guidance on these issues
Charging: private collections, for example, might have to charge commercial units for access	

Technical issues

Issue	Possible solution and/or further thoughts
The software used needs to work in a universal browser to avoid costs for those, for example freelancers, with no funding for updating software and who therefore may not be able to download specific formats	The model and host (ADS) proposed would not require downloading of software. The emphasis is on simplicity, robusticity and longevity for data input and curation
The project will have to work closely with the curators and ADS to make sure no incorrect assumptions are made at each stage	
Without live links to individual databases, regular updates will be needed	
Undertaking the updating will be an ongoing issue. Currently the model relies on David Orton, Eva Fairnell and the University of York, and their priorities could change over time	This is not an immediate issue because of the HE funding and University of York backing. By 2017, once the HE funding has ceased, there should be sufficient feedback on the uptake and usage of the NZRR to present a case, if necessary, for further funding for its promotion, expansion and curation. A collaborative approach between active users may be possible. Even if after a few years it became logistically difficult to update the data, a point -in-time database would still provide the zooarchaeological community with a starting point for tracking down specimens
Will there be long-term commitment after the final phase of the current project (a similar point to that above)? Because the priorities of academic departments can change significantly, long-term sustainability is a concern	Organisations such as Association for Environmental Archaeology (AEA) and International Council for Archaeozoology (ICAZ) could perhaps be approached once the success of the NZRR has been gauged
Risk of stagnation if named contacts drop-out (also similar to points above): because of the reliance on a number of people, bound to become historic at some stage	Ultimately, as above, even a point-in-time database provides a first step in finding a collection that holds species of interest
Updating as frequently as possible would be good, perhaps more than once a year	Not all collections are actively growing, so this is not a universal issue. However, if the NZRR is successful, the model could be developed to enable curators to upload their own updates
Updating will have to include collection-level data as well as specimen data, as whole collections may move/be divided up if, for example, a private curator retires and sells/donates his or her collection	

Issue	Possible solution and/or further thoughts
The format the data have to be supplied in could be an issue, as not everyone is comfortable with, for example, .csv files	The current model has a data manager who can take curator-friendly formats such as Excel and convert them as necessary to ADS's preferred format
Local work may be needed to get a collection in order so that it can contribute to the NZRR	One role of the current model's data manager is to help with this, although how much per collection will depend on demand. Student projects and work placements could help locally
Inconsistencies in underlying data need to be resolved, and commercial units may need financial support for this	Standardisation for collection data input and end-user output will be facilitated by the project
Alternative models suggested for the NZRR were a wiki, or the model used by tDAR	The model proposed for the NZRR was in general considered to be a good, strong model. Because of its simplicity, the proposed NZRR is probably the most robust, but tDAR could be a viable alternative that addresses the issue of financial sustainability. However, not everyone is familiar with tDAR: essentially curators maintain their own databases as they wish, and a 'wrapper' is used to standardise them. It may not currently be a universally accessible alternative but could be a model worth pursuing if the NZRR proves to be a successful concept and further development is undertaken
SurveyMonkey could perhaps be used for people to submit collection-level details to the NZRR: it automatically groups data and is easy to fill in	
The Natural History Near You web page http://natsca.org/NaturalHistoryNearYou provides a good model for minimum collection-level data input	
A standard online form may help data submission, particularly for updates, which could go direct to the ADS rather than via the project staff	
Yes/no categories should be used as much as possible	
Should there be a minimum range of fields for a record to exist?	

Issue	Possible solution and/or further thoughts
Need to be able to indicate the presence of further data within a collection without the need to include it in the NZRR	
Needs to be capacity for expansion/amendment/evolution of the NZRR in the future	The NZRR could/should be built so that it could be used for other types of collection, not just vertebrates
End-users should be able to browse by higher classification but also by collection, location of collections, subset of collection(s), specimens	The use of drop-down menus as search filters should be considered
There is a difference between what fields end-users use to filter searches and the data held in the NZRR. The standardisation needed for each will be different	Standardisation may not be needed for both, and/or the standardisation process will be different for curators and end-users (i.e. the format of the data in and data out). But too much standardisation may also limit inventiveness
One means of submitting data will be needed so that the project does not have to spend time filtering data fields and formats	
Usage needs to be monitored by each collection and/or the NZRR	Google Analytics could be used as a model for providing feedback to curators
Can the search results be downloaded?	
Can search terms be used on multiple fields at once, and/or multiple values within a field?	

Taxonomy issues

Issues and suggestions
FishBase fishbase.org provides a good basis for taxonomy, but is very complex
There can be numerous common names for one species, for example for many fish species
There are pros and cons to free text search versus drop-down search
The project could find the most convenient source of taxonomy and apply it to the data submitted, stating the authorities used on the opening page
End-users can be expected to do a little research: finding the binomial of the target search species using the sources of nomenclature cited by the NZRR should not be a problem
English names could be dropped, in light of the point above
The NZRR could be linked to an online source, for example Species 2000 sp2000.org, or use the taxonomy HE has adopted
Any decisions made about taxonomy should be fed back to curators
The Natural History Museum (NHM) has comprehensive taxonomic data http://www.nhm.ac.uk/our-science/data/uk-species.html
Could use a master list against which the curators' lists are mapped (similar to tDAR)
The International Commission on Zoological Nomenclature iczn.com could be a help

Minimum collection-level data

Data field	Points for curators to consider	Comments
A unique code is needed for each collection		The Collections Trust web site http://www.collectionstrust.org.uk/collections-link/museum-development-support/mda-codes should be checked to make sure a code does not already exist
Name of the collection		
Address of the collection	The actual address, for example 2nd floor of humanities department, rather than the institutional address, for example The University, University Road, Town	
Contact details	Ideally provide a specific name and email address, with a back-up name and email address	The call for updates would be directed to the same contact information, so if there are staff changes people leaving a post may have to field queries until the NZRR is updated. The pros and cons of using a departmental email will vary depending on the institution
Access policy	Bench fees to be charged? Any fair use restrictions? Any destructive sampling possible? Any copyright and licensing restrictions? Physical accessibility of the collection: space, layout, disabled access, visiting hours, etc. How much notice needed before a collection can be visited? Whether certain material can be brought in (for example to avoid pest contamination)	This will probably have to be a free text field, in order to accommodate specific collection requirements: each institution is likely to have subtly different details. Museums tend to stipulate quarantine conditions and processing periods. Should include the possibility of linking to full access policy online. Bench fees could be indicated by yes/no, but no need to provide actual fees charged
Collection summary	Whether further data are available on the collection specimens, what condition they are in, the provenance of the collection and its contents, whether the data within the NZRR is an abridged or truncated version of a collection's complete database	This should probably be another free text field, to allow a qualitative summary of the collection to be presented, for example how many birds/mammals/fish, etc., what proportion of the whole collection is included in the NZRR, what proportion of unprepared, poorly provenanced

Data field	Points for curators to consider	Comments
		and/or archaeological material is present
Type of organisation	What type of institution holds the collection	This could be a drop-down choice or be included in the Collection summary
Date last updated	A time stamp for each collection when submitted rather than the NZRR as a whole	
Other collections held		The focus of this project is vertebrate zooarchaeology collections, but many institutions may also hold other collections of interest, for example malacology, human osteology. A yes/no option could provide the opportunity for these to be mentioned
References		Any citations to research that has used the collection
Acknowledgements	If specimens from a collection are used, how should the collection be acknowledged/referenced, if necessary?	

Minimum specimen-level data to be provided by the curator

Data field	**Comment**
Id number	The original collection's unique id for each specimen
Genus	This is definitely required: higher classifications are inherently based on this, and can be provided by the NZRR rather than the curator
Species	Ideally this should be a separate field, so that a whole genus can be searched for easily
Common name	In theory this is not needed because the genus and species names are provided, but will help the data manager check concordance for the Latin name, and is more user friendly for end-users
Sex	
Age	Adult/subadult/juvenile/foetal/neonate/and/or specific age: broad categories would probably be easier than specific ages, with an option to indicate whether the specific age of a specimen is known
Completeness	Could have a coding system or a simple descriptor, or free text. Cranial/axial rather than percentage? Exclude single elements?
Cast	This could be very important, as many collections could have a cast of the same donor species so a search would return duplicates

Extra specimen-level data to be provided by the curator

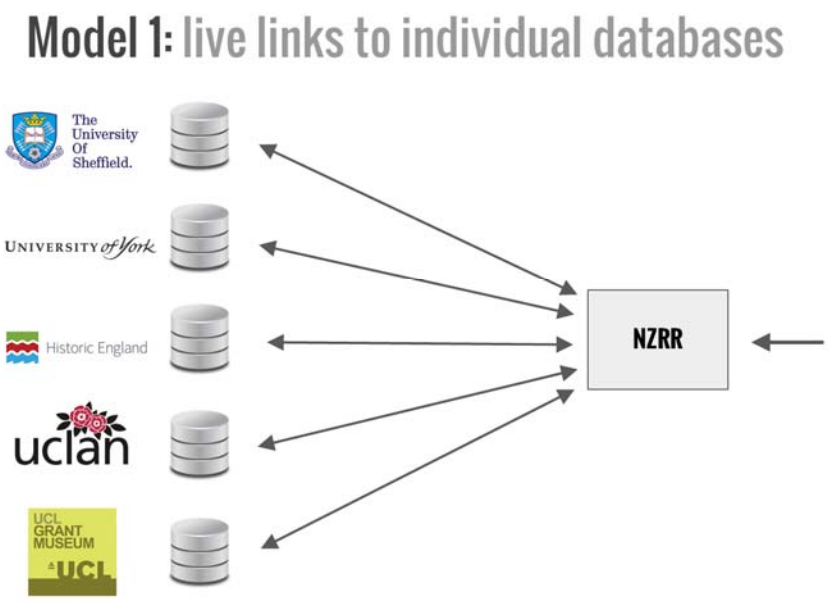
Data field	**Comment**
Breed/Sub-species	
Zoo-bred	
Lab-bred	
Wild	
Modern	
Archaeological	
Historic	
Identification criteria	
Articulated/partially articulated/ disarticulated	
Pathological	Option to flag whether pathology present
Locality	Where the specimen is from. Possibly tiered: county, and/or further detail. Would be useful for isotope research
Metrics: size type/value/TL (for fish)	It may be easier to have a yes/no option for whether carcass metrics, including carcass mass, are available
Non-metric traits	
Biomolecular data	
Teeth present	
Previously sampled	
Prep notes	Whether enzymes were used in the prep of the carcass could make a difference to suitability for isotope analysis, etc.
Index collection	
Elements elsewhere	This could be important to avoid double counting: many collections in the past have taken the right or left side of the same animal
Available for loan	
Ownership	
Collection notes	
Digital data: images	Links or references to 2D or 3D images that already exist
References	Any citations to previously published results

Extra specimen-level data to be available for the end-user

Data field	**Comment**
Class	Could be a look-up table
Order	
Family	
Mammal/bird/fish/herp	More intuitively useful than class?
Synonyms	

Appendices

The relative cost for each model reduces significantly from model 1 to 2 to 3.



Pros

Updates: 'live' data; seamless for curators once up and running.

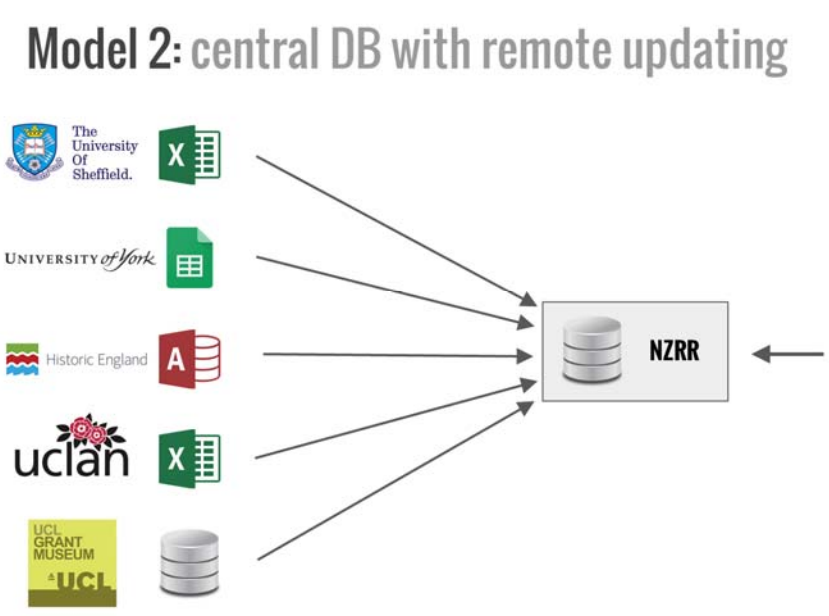
Cons

Set up: custom data-base at each node ⇒ huge developer & curator time demands.

Relies on local IT set-up

Stability: needs IT support at centre and at each node. If a node goes down, access lost.

Expandability: hard to add new nodes.



Pros

Updates: curators *can* update continuously.

Stability: no reliance on live links; no risk of nodes dropping out.

Relatively easy to add collections in future.

Cons

Duplicates work for curators.

Relies on curators to update as needed.

Need to train curators, deal with staff changes

If you require an alternative accessible version of this document (for instance in audio, Braille or large print) please contact our Customer Services Department:
Telephone: 0370 333 0607
Fax: 01793 414926
Textphone: 0800 015 0516
E-mail: customers@HistoricEngland.org.uk