Arts and Humanities Research Council

Collaborative Doctoral Partnership

Historic England

The FAIRness of archaeological data:

an examination of bioarchaeological and Historic High Street datasets.

University of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council

Collaborative Doctoral Partnership

Historic England

Introduction

What are the FAIR data principles

Why be FAIR

Application in:

Bioarchaeology	Historic High Street

University of York

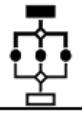
ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

The FAIR data principles

(Wright and Richards, 2020)

Findable	Persistent IDs 	Metadata schemas 	PIDs in metadata 	
Accessible	Communication protocols 	Harvestable metadata and Endpoints 	Open Access 	Repositories 
Interoperable	Metadata models 	Standardised file formats 	Ontologies 	Controlled vocabulary 
Reusable	Systematic documentation 	Community standards 	Detailed metadata 	Usage licence 

(Authors own)

UNIVERSITY of York

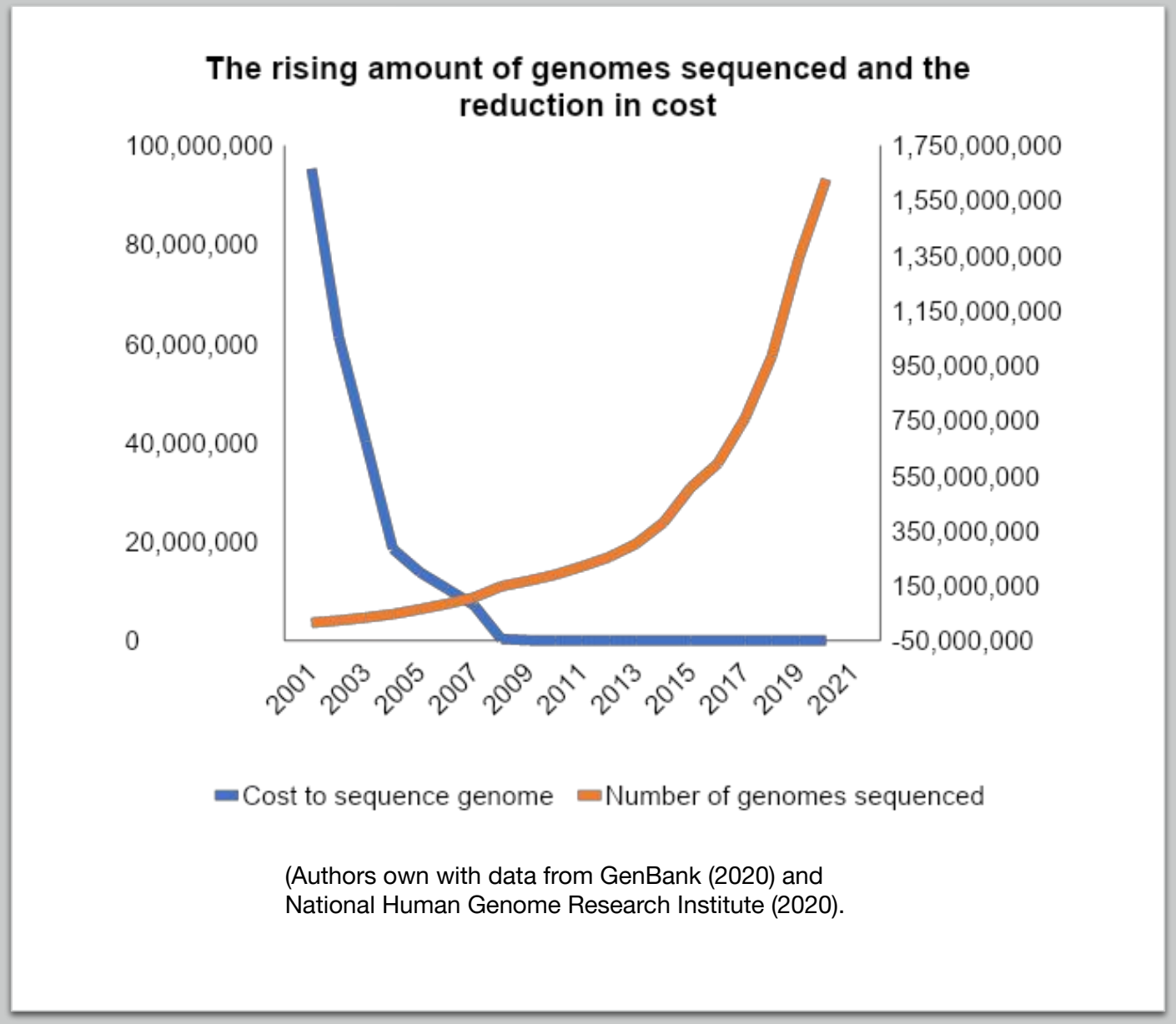
ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council

Collaborative Doctoral Partnership

Why be FAIR

- Archaeology is a destructive process (Oakley, 2005, 171; Pálsdóttir, 2019, 2)
- More and more data created (Green *et al.*, 2017, 180).
- Increase in misuse of PDF format (Evans and Moore, 2014, p. 124; Kansa et al. 2020, p. 45; Sobotkova, 2018, p. 121).



(Authors own with data from GenBank (2020) and National Human Genome Research Institute (2020).

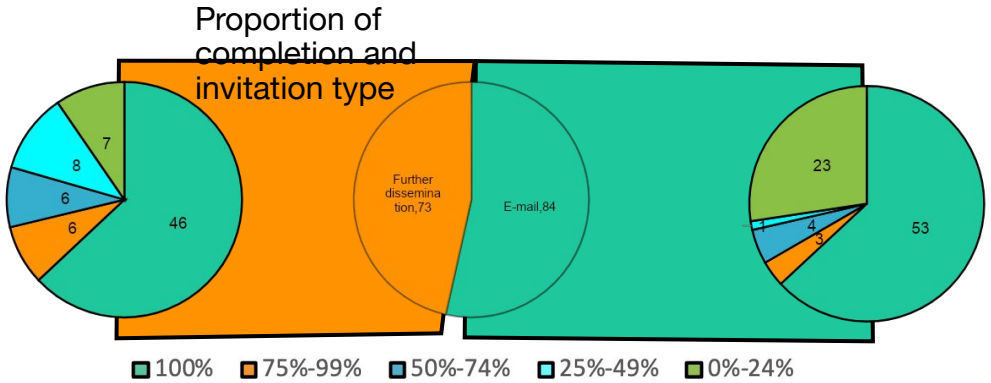
UNIVERSITY of York

Arts and Humanities Research Council

Collaborative Doctoral Partnership

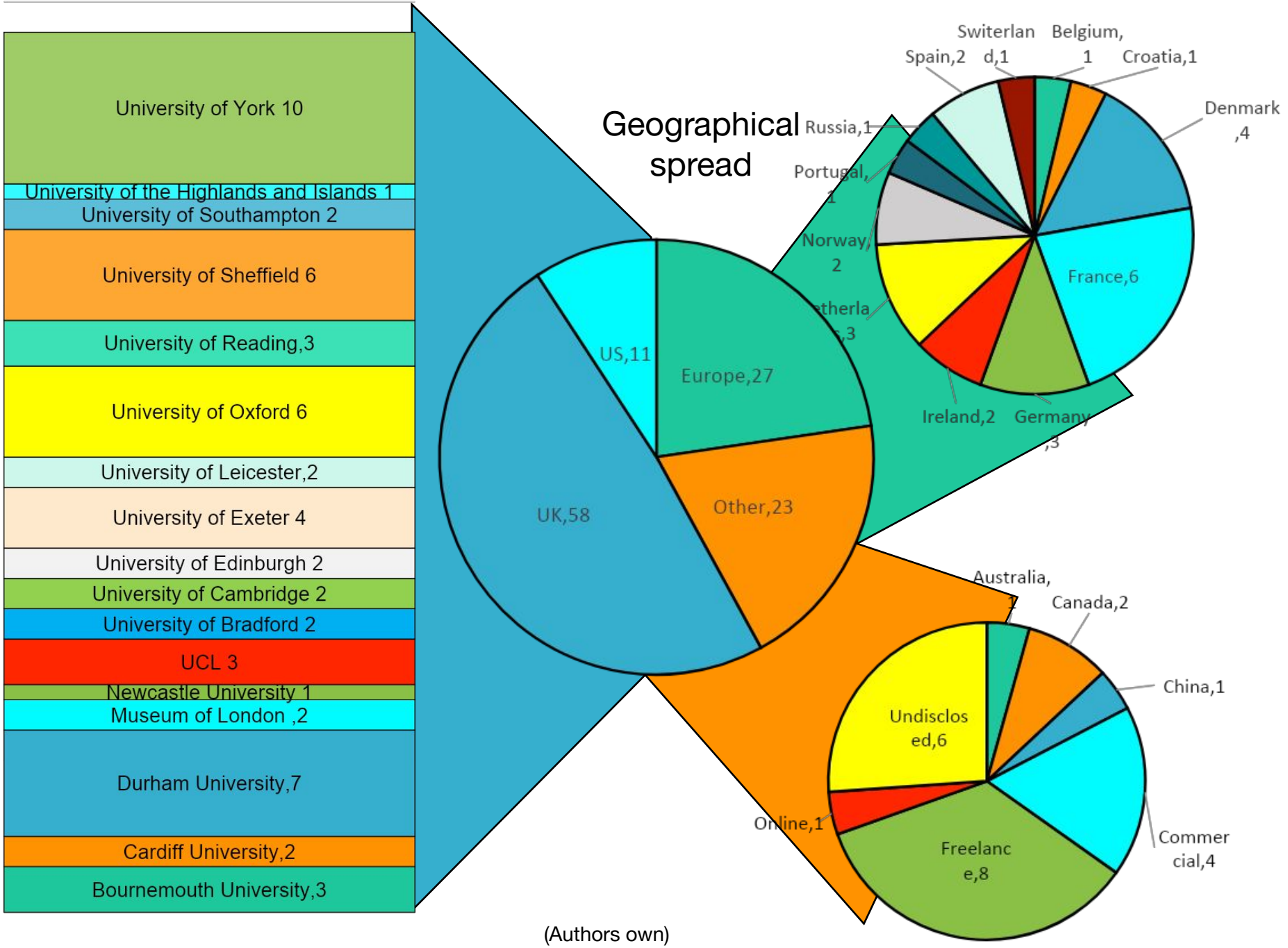
How FAIR is bioarchaeology

- Questionnaire
  - (154 responses)
  - Level of interactivity between specialisms.
  - How FAIR is bioarchaeology ?



(Authors own)

UNIVERSITY of York


University of York 10
University of the Highlands and Islands 1
University of Southampton 2
University of Sheffield 6
University of Reading,3
University of Oxford 6
University of Leicester,2
University of Exeter 4
University of Edinburgh 2
University of Cambridge 2
University of Bradford 2
UCL 3
Newcastle University 1
Museum of London ,2
Durham University,7
Cardiff University,2
Bournemouth University,3
Geographical spread
US,11
Europe,27
UK,58
Other,23
Switerlan d,1
Belgium, 1
Spain,2
Croatia,1
Denmark ,4
Russia,1
Portugal, 1
Norway, 2
France,6
Ireland,2
Germany
Australia, 1
Canada,2
China,1
Undisclos ed,6
Online,1
Freelanc e,8
Commer cial,4

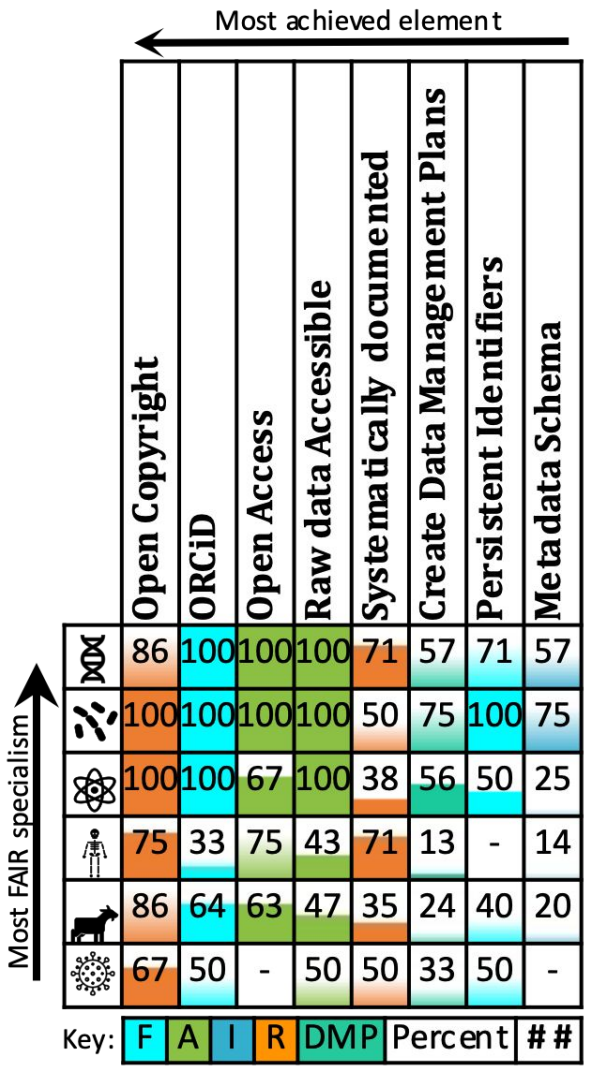
(Authors own)

Arts and Humanities Research Council

Collaborative Doctoral Partnership

Results

- Reuse of data is important
- Some extent of reuse present
- No standardised process
- Data is not FAIR



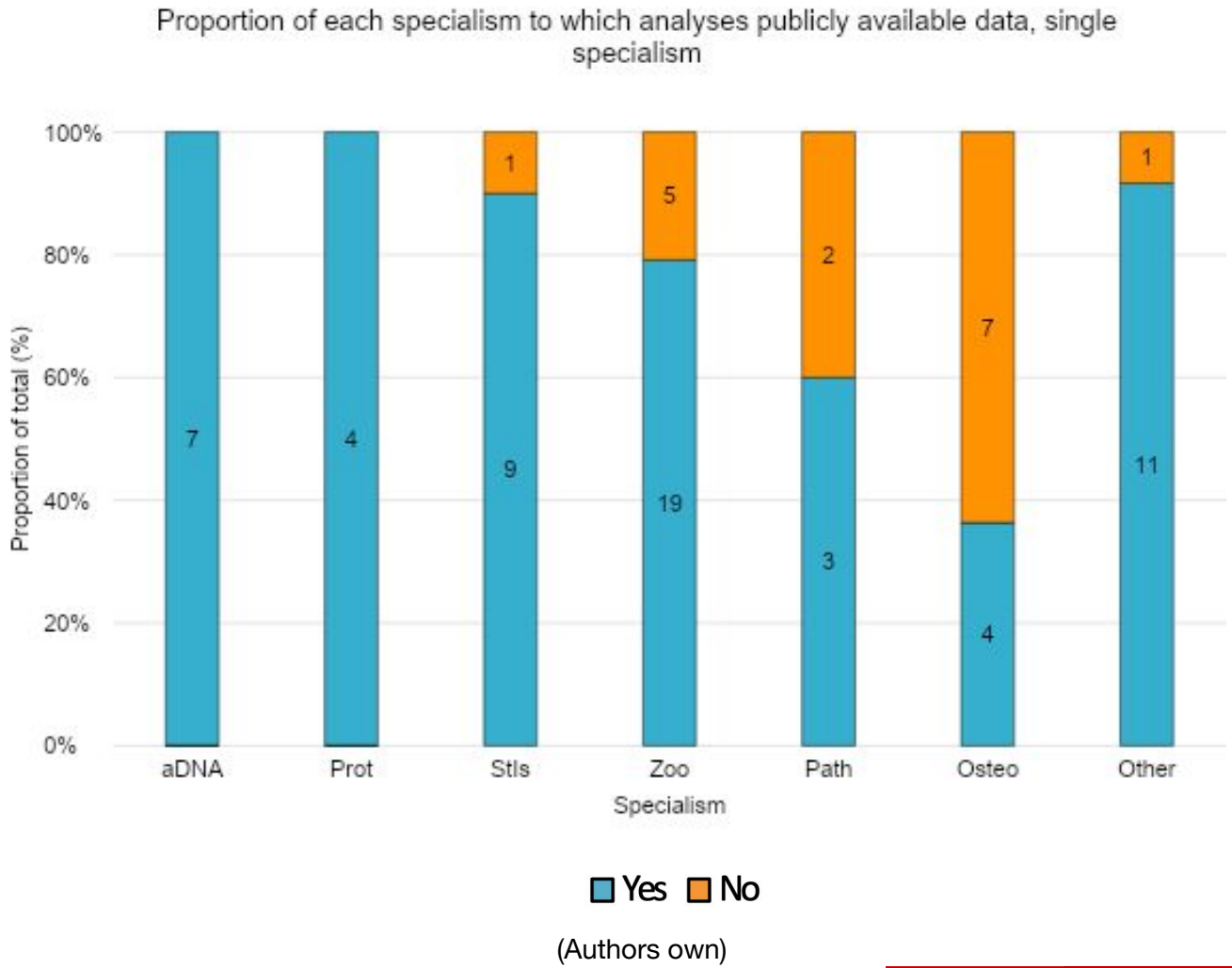
(Authors own)

UNIVERSITY of York

Current reuse of data: between specialisms

(Authors own)

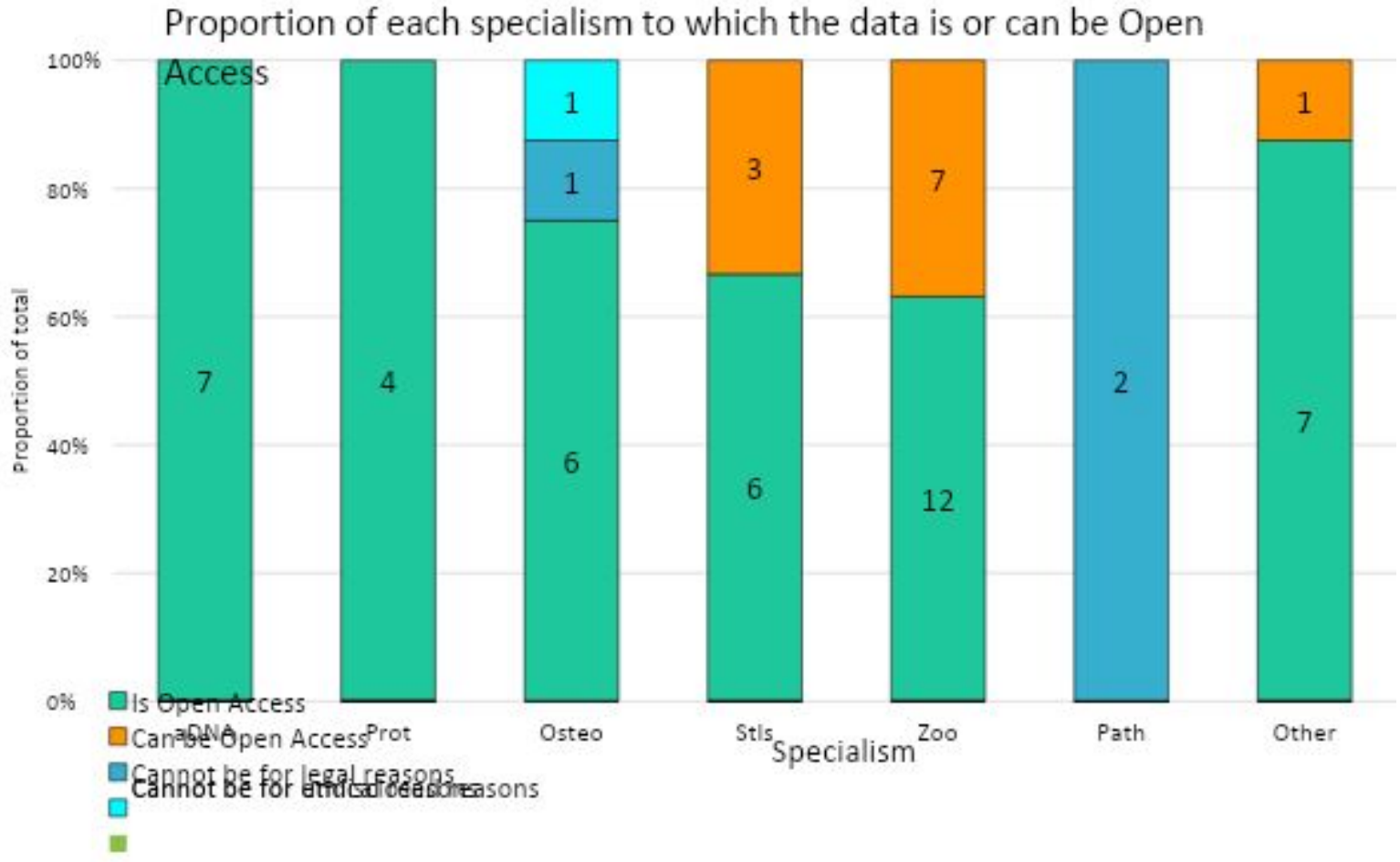
Reuse of data



(Authors own)

Open Access

Historic England



(Authors own)

ads ARCHAEOLOGY DATA SERVICE

Data type

	PDF	.XLSX	CSV	JPEG	RAW	FastQ	BAM	Other
aDNA	14%	29%	14%	14%	0%	86%	71%	43%
Osteoarchaeology	67%	33%	17%	33%	0%	0%	0%	92%
Paleopathology	60%	0%	0%	40%	0%	0%	0%	60%
Pealeoproteomics	0%	0%	50%	0%	100%	0%	0%	75%
StableIsotopes	80%	70%	40%	20%	0%	0%	0%	100%
Zooarchaeology	56%	36%	32%	36%	0%	0%	0%	72%
Other	67%	42%	25%	8%	0%	0%	0%	100%

(Authors own)

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

Results summary

>50% participants	✓
50% participants	-

F	Persistent identifiers	✓	-		✓	-	
	ORCiDs	✓		-	✓	✓	✓
A	Open Access	✓	✓		✓	✓	✓
	Raw accessible	✓		-	✓	✓	
I	Metadata	✓			✓		
	Data type	FASTQ	PDF	PDF	RAW	PDF	PDF
	Level of process	Raw	Fully	Partly	Raw	Fully	Fully
R	Copyright (none)	✓	✓	✓	✓	✓	✓
	Syst. Doc.	✓	✓	-	-		
	Data reused	✓	✓	✓	-	✓	✓
DMP	Create	✓			✓	✓	

(Authors own)

UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

Interoperable datasets

- Rise of “Grey Literature” and misuse of PDF format
- NLP and NER can unlock this data (Brandsen *et al.*, 2021)
- NLP – Processing of textual documents
- NER – Recognition and classifying of terms (Richards *et al.*, 2011)

University of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

Previous works

(Richards *et al.* 2011; Tudhope *et al.* 2011; May *et al.* 2012; Binding and Tudhope 2016; Talboom 2017)

Archaeotools 2007	NLP Guided search Machine learning
STAR 2007	Ontology GATE Developer
STELLAR 2010	
SENESCHAL 2013	Controlled vocabulary
Zooarch Entity Search 2017	Basis of project

UNIVERSITY of York

ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

Archaeology Data Service

(Richards 1997, 1057; Wright 2019)

26 YEARS

FREELY DISSEMINATE DIGITAL RESOURCES MADE BY RESEARCH

NATIONAL ARCHIVE FOR ARCHAEOLOGICAL DATA

OVER 1 MILLION FILES

UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

(Keily 2017)

Crossrail

200 ARCHAEOLOGISTS
OVER 10,000 ARTEFACTS

42 KM NEW TUNNELS AT
A DEPTH OF 30-40M

CLOSED EXAMPLE OF
DOCUMENTS

University of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

Methods

Document selection

Document annotation

Evaluation

Reliable,
time saving,
accessible,
useful and
reuse again

UNIVERSITY of York

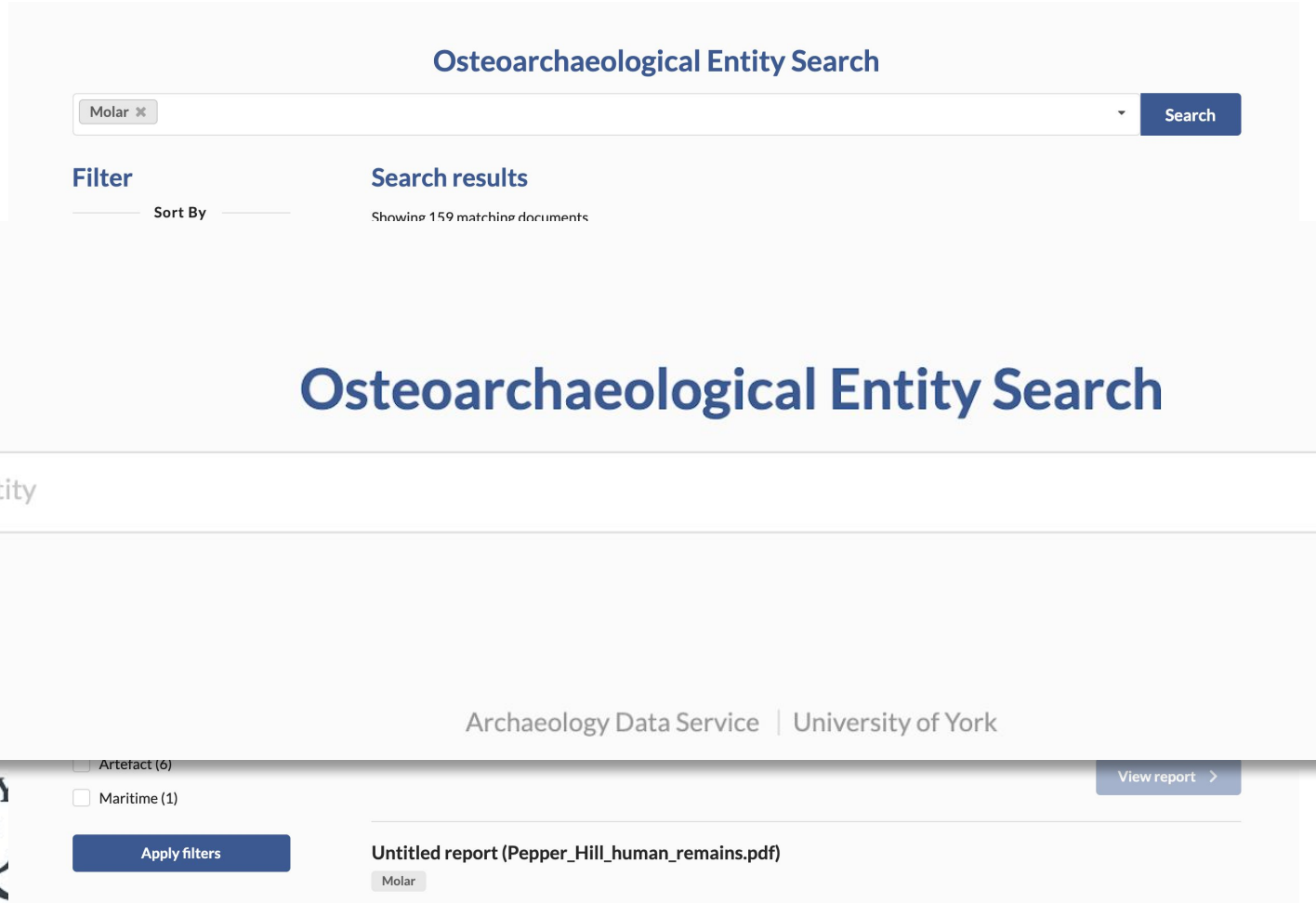
ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council

Collaborative Doctoral Partnership

Historic England

How to use the tool



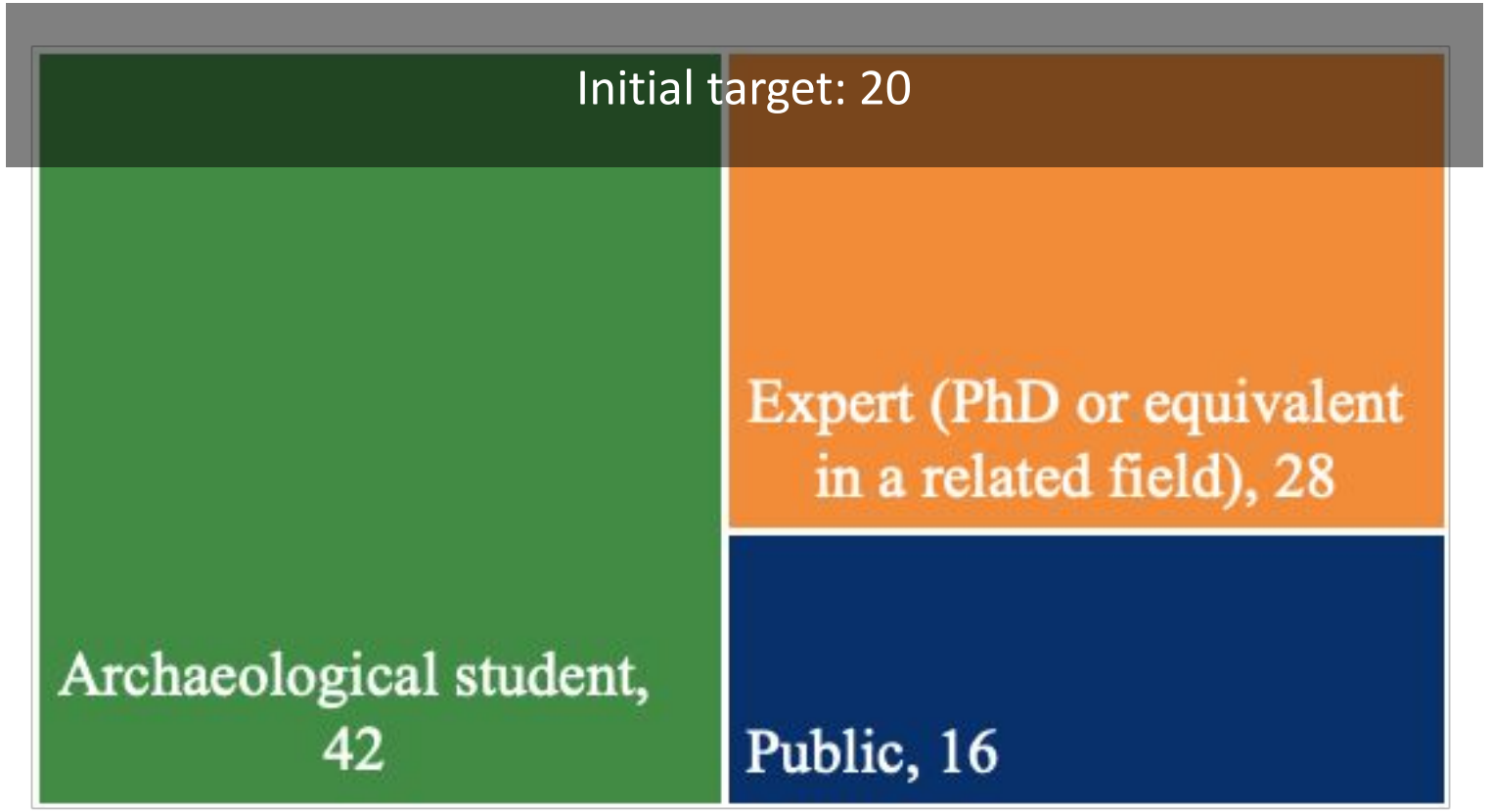
UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

Participants

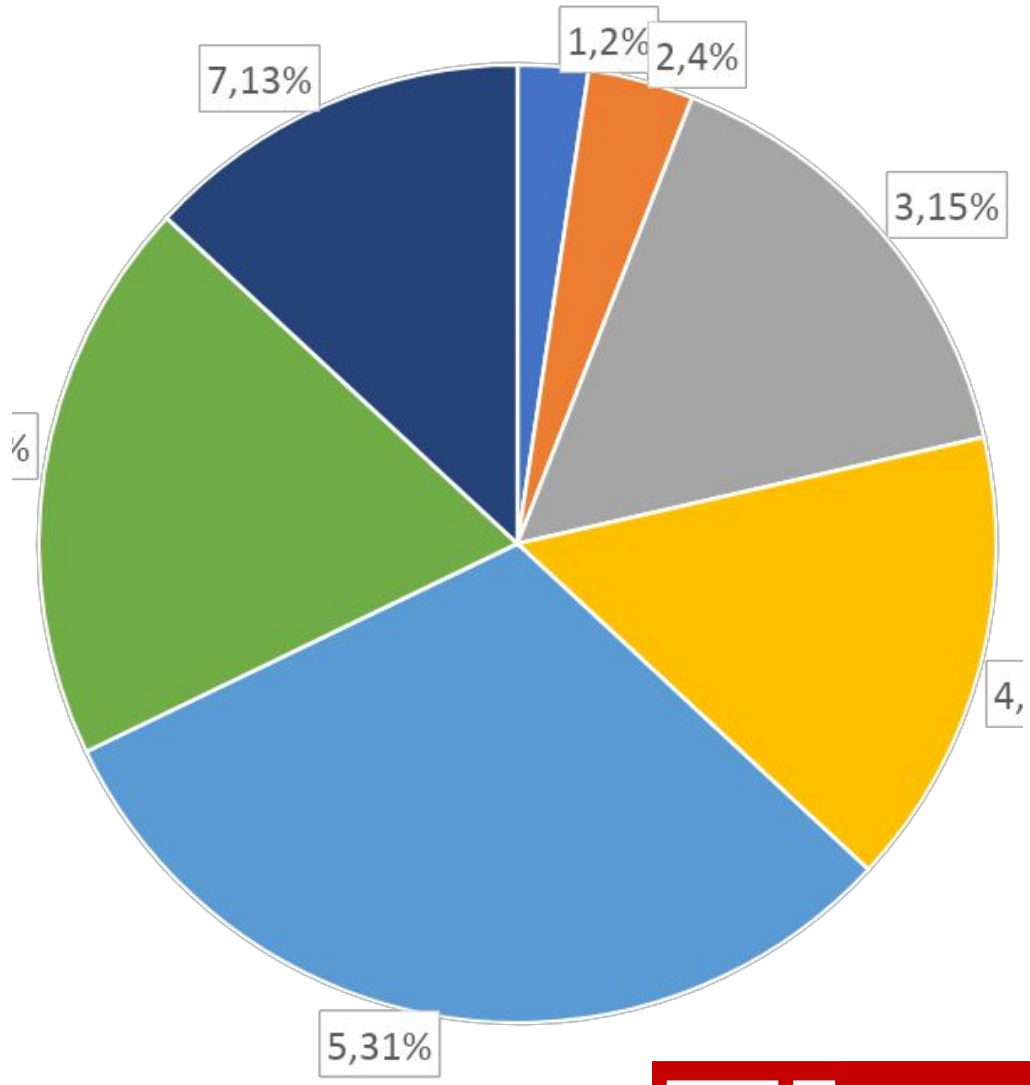


UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE

Are the results reliable?

Mean	Modal	Partially met	Fully met
4.79	5	3.5	5.25



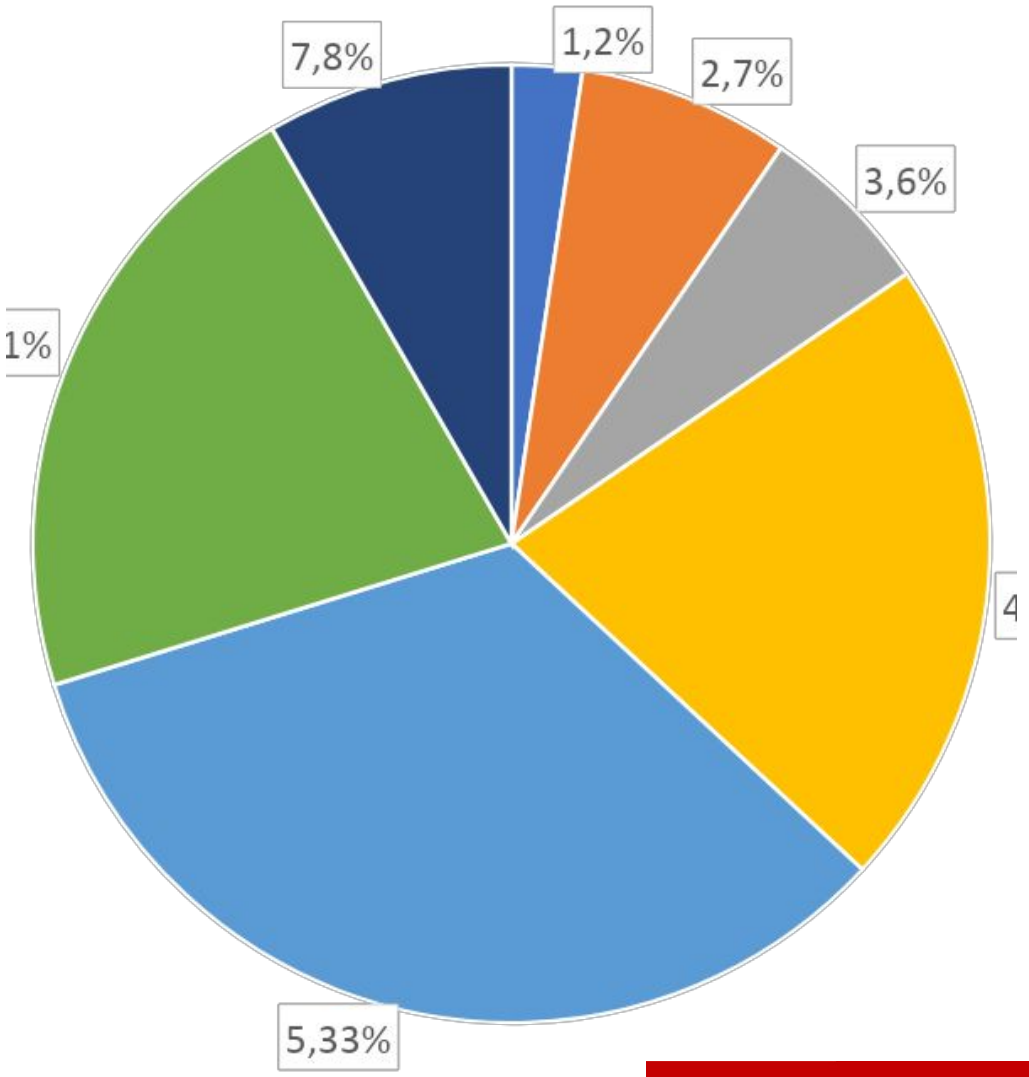
(Authors own)

Historic England

ads ARCHAEOLOGY DATA SERVICE

Is it time saving?

Mean	Modal	Partially met	Fully met
4.74	5	3.5	5.25



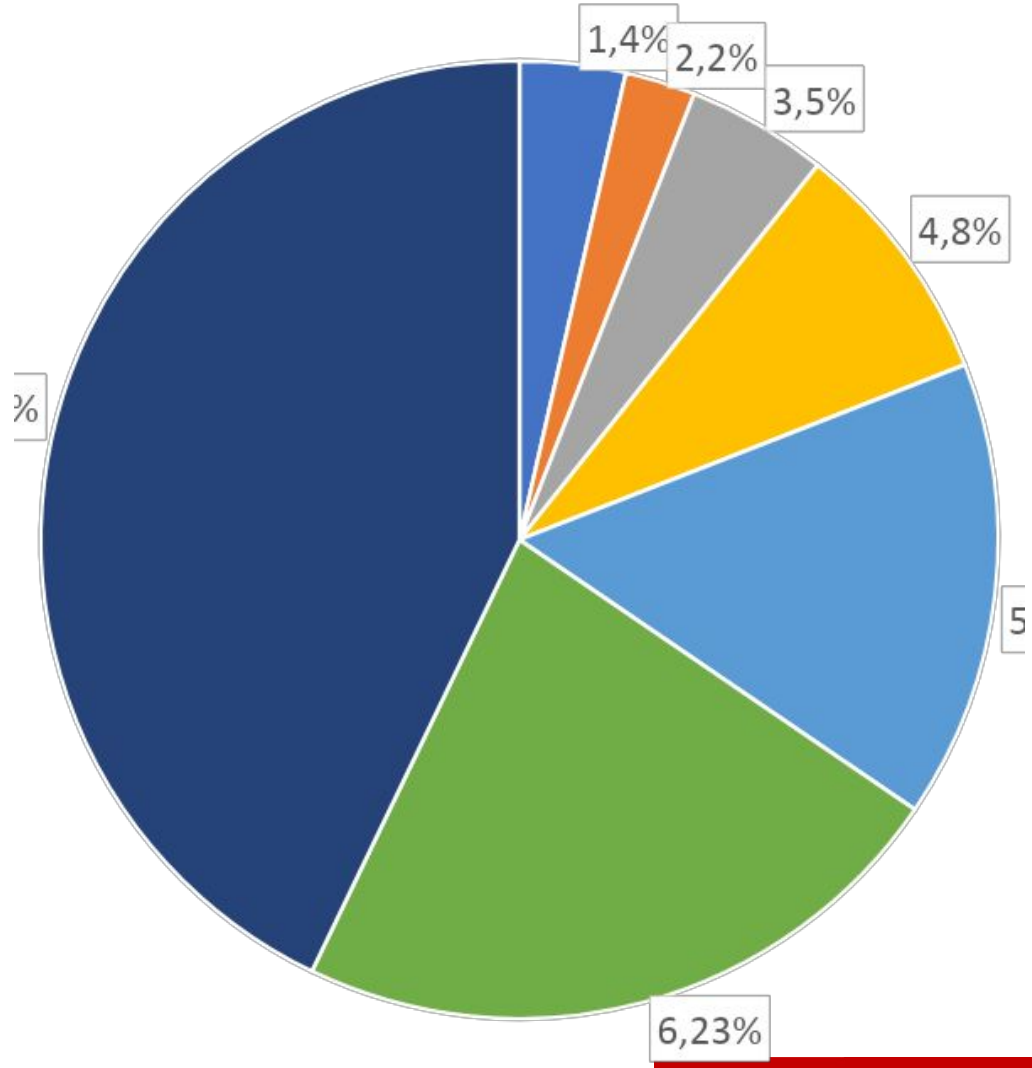
(Authors own)

Historic England

ARCHAEOLOGY DATA SERVICE

How accessible was it?

Mean	Modal	Partially met	Fully met
5.69	7	3.5	5.25



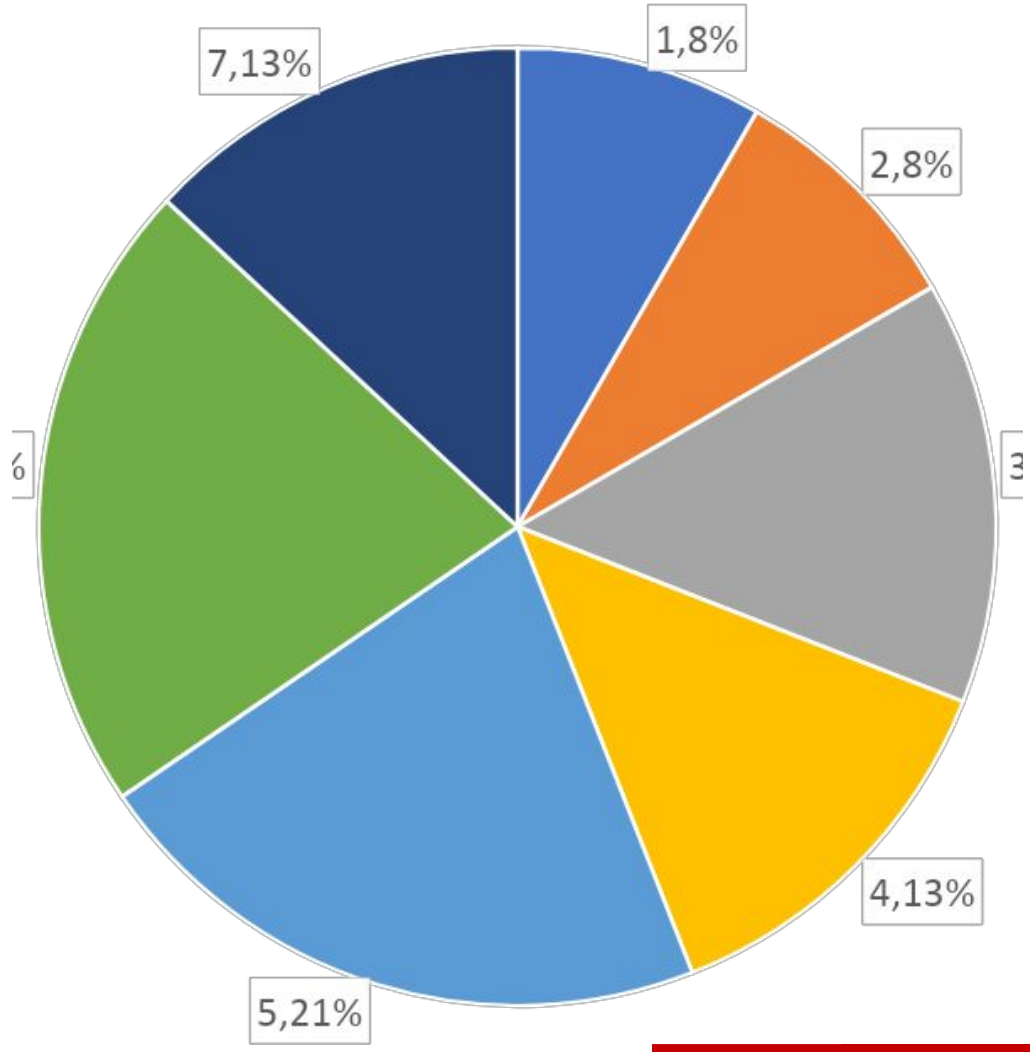
(Authors own)

Historic England

ads ARCHAEOLOGY DATA SERVICE

Would you use again?

Mean	Modal	Partially met	Fully met
4.48	5&6	3.5	5.25



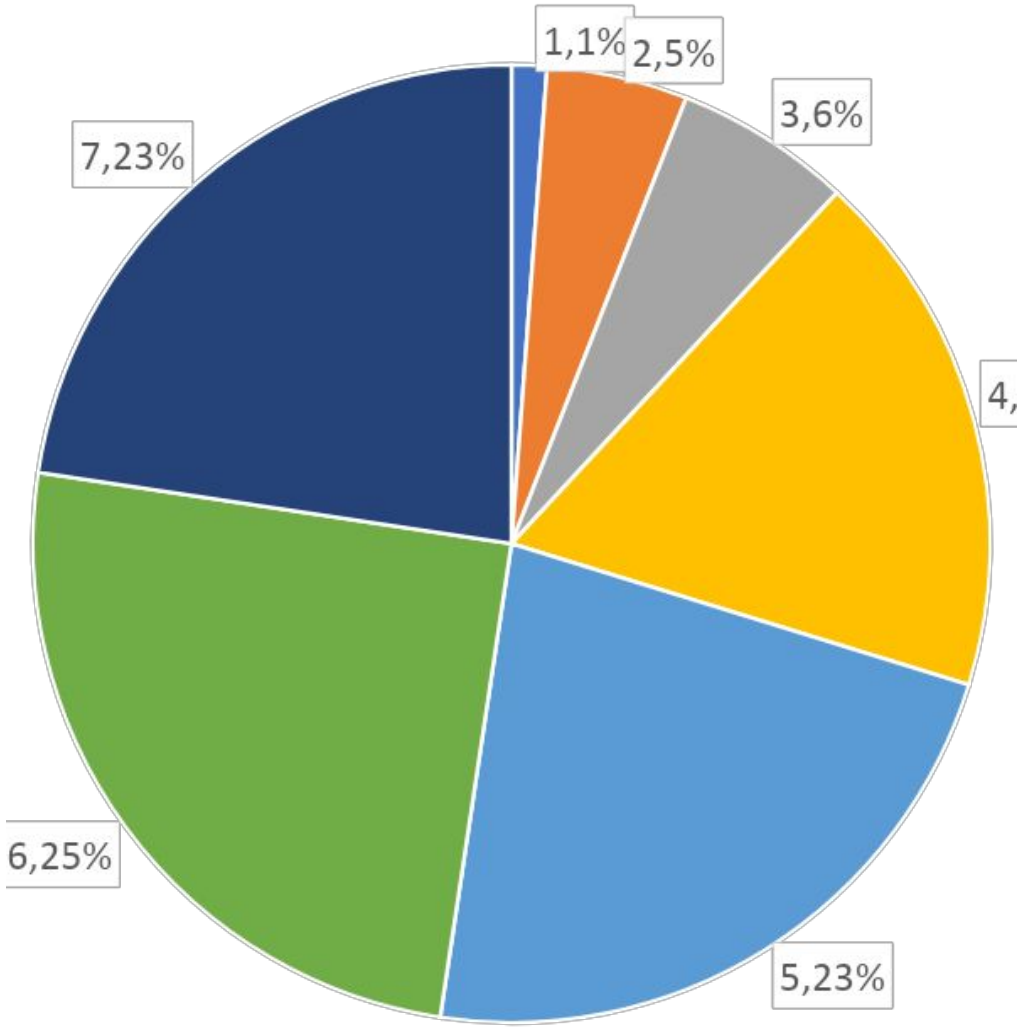
(Authors own)

Historic England

ads ARCHAEOLOGY DATA SERVICE

Is it useful for archaeologists?

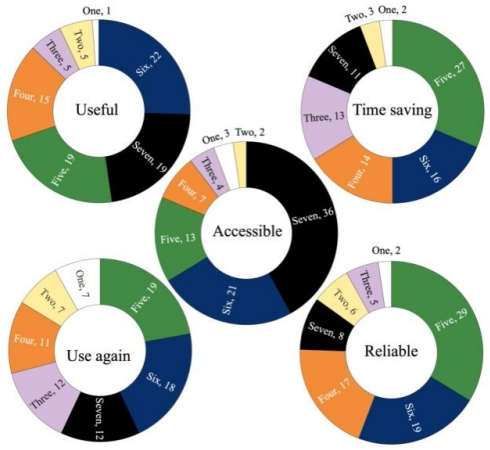
Mean	Modal	Partially met	Fully met
5.21	6	3.5	5.25



(Authors own)

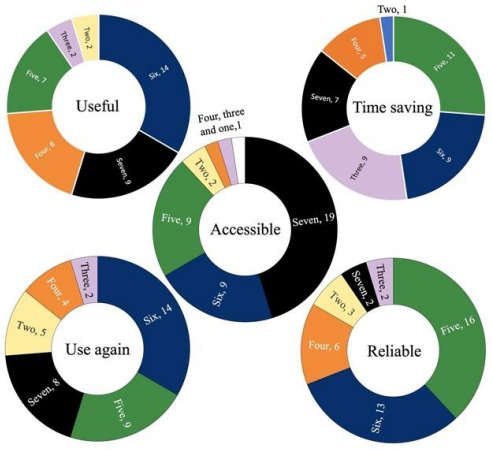
Historic England

ads ARCHAEOLOGY DATA SERVICE



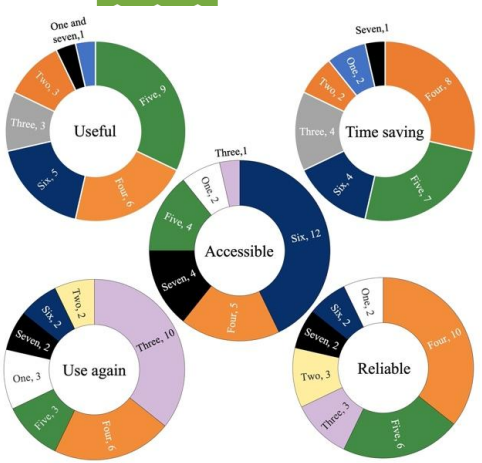
All

Number of participants: 86	Time saving		Reliable		Accessible		Use again		Useful	
1	2	2%	2	2%	3	3%	7	8%	1	1%
2	3	3%	6	7%	2	2%	7	8%	5	6%
3	13	15%	5	6%	4	5%	12	14%	5	6%
4	14	16%	17	20%	7	8%	11	13%	15	17%
5	27	31%	29	34%	13	15%	19	22%	19	22%
6	16	19%	19	22%	21	24%	18	21%	22	26%
7	11	13%	8	9%	36	42%	12	14%	19	22%
Minimum	1		1		1		1		1	
Maximum	7		7		7		7		7	
Mode	5		5		7		5		6	
Mean	4.78		4.79		5.70		4.51		5.19	
Range	6		6		6		6		6	



Students

Number of participants: 42	Time saving		Reliable		Accessible		Use again		Useful	
1	0	0%	0	0%	1	2%	0	0%	0	0%
2	1	2%	3	7%	2	5%	5	12%	2	5%
3	9	21%	2	5%	1	2%	2	5%	2	5%
4	5	12%	6	14%	1	2%	4	10%	8	19%
5	11	26%	16	38%	9	21%	9	21%	7	17%
6	9	21%	13	31%	9	21%	14	33%	14	33%
7	7	17%	2	5%	19	45%	8	19%	9	21%
Minimum	2		2		1		2		2	
Maximum	7		7		7		7		7	
Mode	5		5		7		6		6	
Mean	4.93		4.95		5.81		5.17		5.33	
Range	5		5		6		5		5	

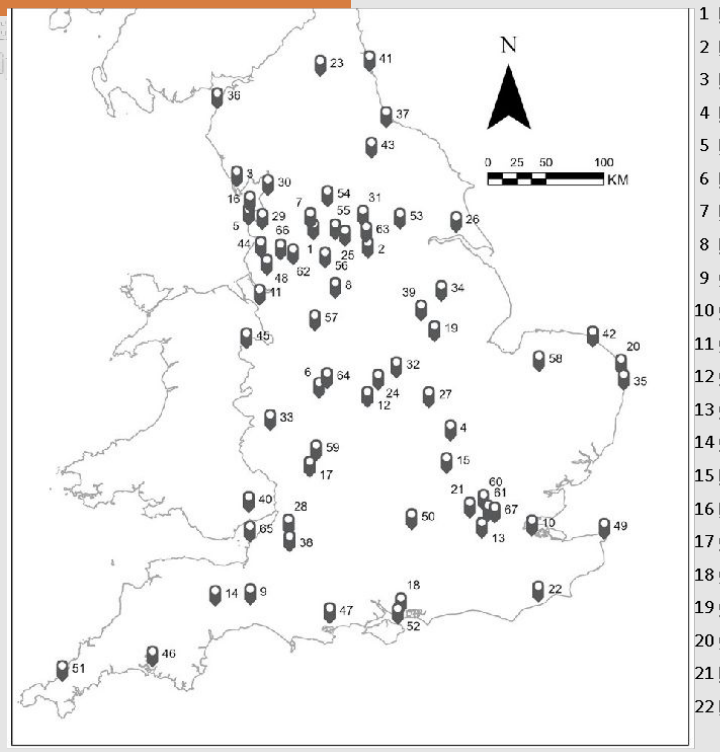


Experts

Number of participants: 28	Time saving		Reliable		Accessible		Use again		Useful	
1	2	7%	2	7%	2	7%	3	11%	1	4%
2	2	7%	3	11%	0	0%	2	7%	3	11%
3	4	14%	3	11%	1	4%	10	36%	3	11%
4	8	29%	10	36%	5	18%	6	21%	6	21%
5	7	25%	6	21%	4	14%	3	11%	9	32%
6	4	14%	2	7%	12	43%	2	7%	5	18%
7	1	4%	2	7%	4	14%	2	7%	1	4%
Minimum	1		1		1		1		1	
Maximum	7		7		7		7		7	
Mode	4		4		6		3		5	
Mean	4.14		4.04		5.18		3.64		4.36	
Range	6		6		6		6		6	

(Authors own)

Osteoarchaeological and palaeopathological entity search



1 Bacup, Rossendale	23 Hexham	46 Plymouth
2 Barnsley	24 Hinckley	47 Poole
3 Barrow in Furness	25 Huddersfield	48 Prescot
4 Bedford	26 Hull	49 Ramsgate
5 Blackpool	27 Kettering	50 Reading
6 Brierley Hill	28 Keynsham	51 Redruth
7 Burnley	29 Kirkham	52 Ryde
8 Buxton	30 Lancaster	53 Selby
9 Chard	31 Leeds	54 Skipton
10 Chatham	32 Leicester	55 Sowerby Bridge
11 Chester	33 Leominster	56 Stalybridge
12 Coventry	34 Lincoln	57 Stoke on Trent
13 Croydon	35 Lowestoft	58 Swaffham
14 Cullompton	36 Maryport, Cumbria	59 Tewkesbury
15 Dunstable	37 Middlesbrough	60 Tottenham
16 Fleetwood	38 Midsomer Norton	61 Tower Hamlets
17 Gloucester	39 Newark-on-Trent	62 Tyldesley, Greater Manchester
18 Gosport	40 Newport	63 Wakefield
19 Grantham	41 North Shields	64 Wednesbury
20 Great Yarmouth	42 North Walsham	65 Weston-super-Mare
21 Harlesden	43 Northallerton	66 Wigan
22 Hastings	44 Ormskirk	67 Woolwich
	45 Oswestry	

(Authors own)

Historic High Street

- Ensuring the accessibility and reuse of data created from the High Street
- HAZ – stakeholders for past
- HSHAZ - economic, social and cultural recovery
- Many datatypes

Previous studies

Historic England

- Historic Town Atlas – interoperability of datasets between cities
- EUS and HLC – how characterisation assists FAIR
- Mapping Medieval Chester – the relationships between datasets
- City Witness – how interoperability helps with lack of contemporary
- Know Your Place – inclusion of community datasets
- Layers of London – how to access community datasets with iteration
- CHARTEX – how to access textual documents using NLP

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council

Collaborative Doctoral Partnership

Historic England

Methodology

Needs Analysis

Ensure the long-term preservation and reusability of data to researchers and public

Iterate strategies of FAIR data

4 case studies

University of York

ads ARCHAEOLOGY DATA SERVICE

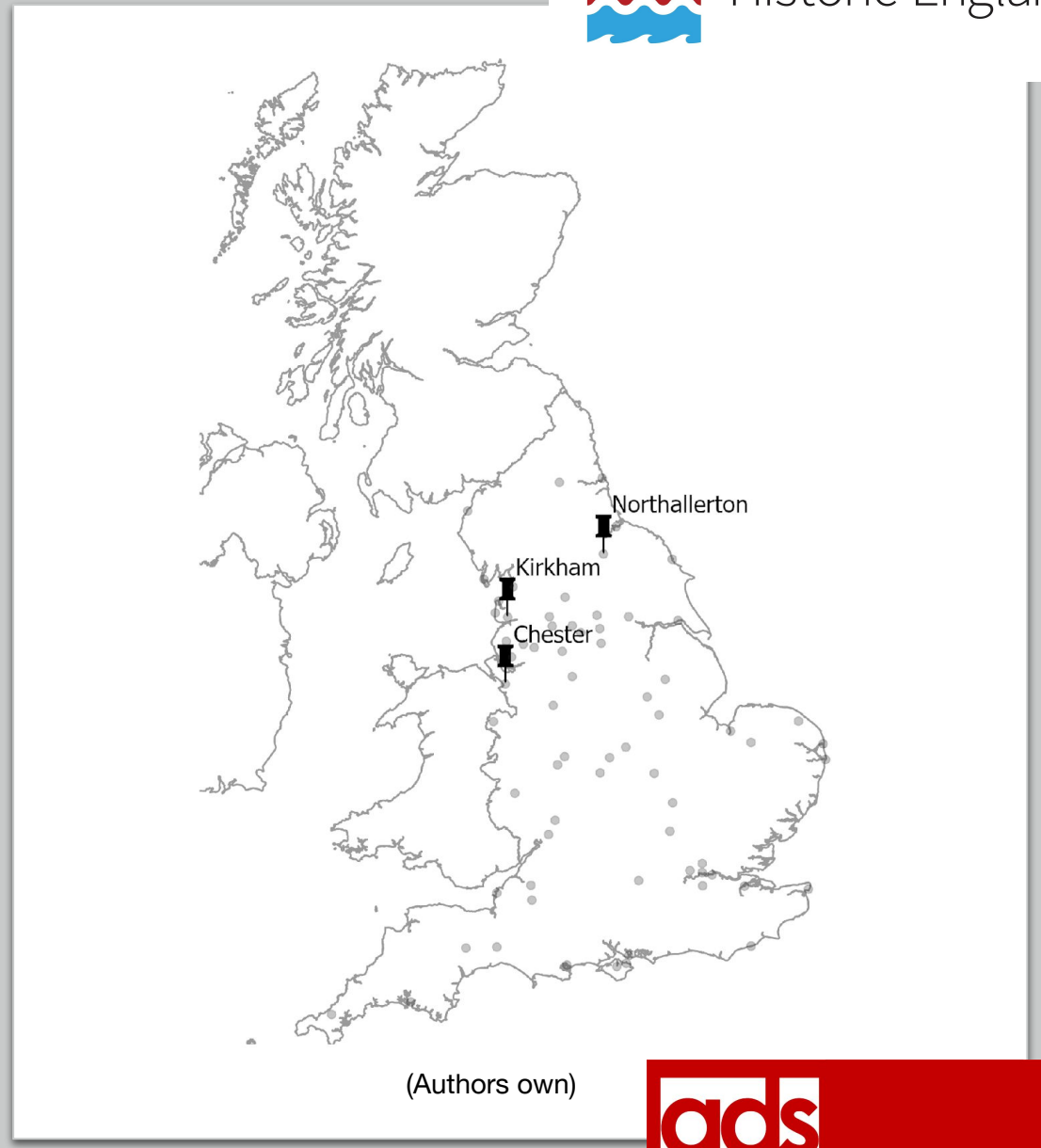
Arts and Humanities Research Council

Collaborative Doctoral Partnership

Historic England

Case studies

1. Chester – “complete” dataset, for data capture and management practices
2. Northallerton – what data is being reused
3. Kirkham – beginning of HSHAZ work
4. Fourth?



(Authors own)

UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council
Collaborative Doctoral Partnership

Historic England

Conclusion

WHAT IS THE FAIR DATA PRINCIPLES

WHY USE THEM

HOW FAIR IS BIOARCHAEOLOGY

HOW TO ACCESS DATA INSIDE PDFS

HOW THEY ASSIST WITH THE HISTORIC HIGH STREET

UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE

Arts and Humanities Research Council

Collaborative Doctoral Partnership

Historic England

References

- Wright, H. and Richards, J. (2020). D.5.3 Data Curation Policy. E-RIHS. [Online]. Available at: https://doi.org/10.1080/00934690.2018.1511960 (accessed 13 May 2020).
- Oakley, K. (2005). Forensic archaeology and anthropology: An Australian perspective. *Forensic science, medicine, and pathology*, 1(3), 169–172.
- Pálsdóttir, A. H., Bläuer, A., Rannamäe, E., Boessenkool, S. and Hallsson, J. H. (2019). Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. *Royal society open science*, 6(10), 191059.
- Green, E. D., Rubin, E. M., and Olson, M. V. (2017). The future of DNA sequencing. *Nature News*, 550(7675), 179.
- Evans, T. N. and Moore, R. H. (2014). The use of PDF/A in digital archives: a case study from archaeology. *International journal of digital curation*, 9(2), 123-138.
- Kansa, S. W., Atici, L., Kansa, E. C. and Meadow, R. H. (2020). Archaeological analysis in the information age: guidelines for maximizing the reach, comprehensiveness, and longevity of data. *Advances in archaeological practice*, 8(1), 40–52.
- Sobotkova, A. (2018). Sociotechnical obstacles to archaeological data reuse. *Advances in archaeological practice*, 6(2), 117-124.
- Brandsen, A, Verberne, S, Wansleeben, M and Lambers, K 2020 Creating a Dataset for Named Entity Recognition in the Archaeology Domain. In: Calzolari, N, Bechet, F, Blache, P, Choukri, K, Cieri, C, Declerck, T, Goggi, S, Isahara, H, Maegaard, B, Mariani, J and Mazo, H (eds.). *Proceedings of the 12th Language Resources and Evaluation Conference Marseille, 11 to 16 May 2020.* Marseille: The European Language Resources Association, pp. 4573-4577.
- Richards, J., Jeffrey, S., Waller, S., Ciravegna, F., Chapman, S. and Zhang, Z. (2011). The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing. In: E. Kansa, S. Whitcher Kansa and E. Watrall (eds). *Archaeology 2.0 new approaches to communication & collaboration*. California: Cotsen Digital Archaeology, pp. 31-56.

UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE

- Tudhope, D., Binding, C., Jeffrey, S., May, K. and Vlachidis, A. (2011). A STELLAR role for knowledge organization systems in digital archaeology. *Bulletin of the Association for Information Science and Technology*, 37(4), 15-18.
- May, K., Binding, C. and Tudhope, D. (2015). Barriers and opportunities for linked open data use in archaeology and cultural heritage. *Archäologische informationen*, 38(1), 173-184.
- Binding, C. and Tudhope, D. (2016). Improving interoperability using vocabulary linked data. *International journal on digital libraries*, 17(1), 5-21.
- Talboom, L. (2017). *Improving the discoverability of zooarchaeological data with the help of Natural Language Processing*. Unpublished: University of York. MSc Archaeological Information Systems.
- Richards, J. D. (1997). Preservation and re-use of digital data: the role of the Archaeology Data Service. *Antiquity*, 71(274), 1057-1059.
- Holly Wright, Archaeology Data service, pers. comm. February 2019
- Keily, J. (2017). Tunnel: The archaeology of Crossrail. London: Museum of London Docklands.

Arts and Humanities Research Council

Collaborative Doctoral Partnership

Historic England

Any questions?

Alfie Talks: AT1186@york.ac.uk

UNIVERSITY of York

ads ARCHAEOLOGY DATA SERVICE