

# Reimagining Archaeological Data Management Workflows through the Lens of Reuse

**DR HOLLY WRIGHT**

*ARCHAEOLOGY DATA SERVICE, UNIVERSITY OF YORK*

CAA AUCKLAND, 9 APRIL 2024



# **PRESENTATION CONTENTS**



**Goal of the TETRARCHs Project**



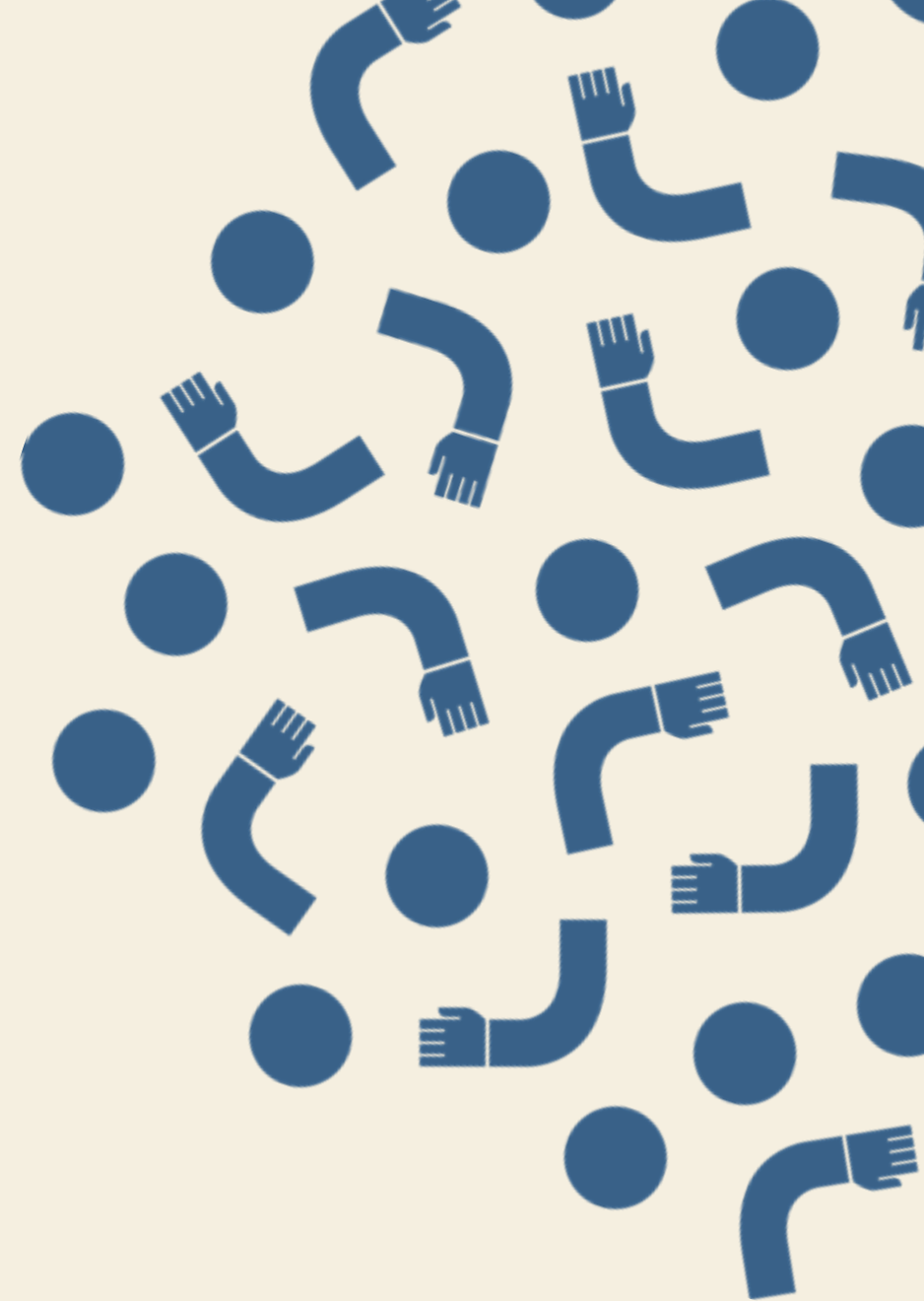
**TETRARCHs Data Mapping Strategy**



**TETRARCHs Workshop on Data  
Evaluation and Recommendations**



**Discussion and Results**



# Goal of the TETRARCHs Project



**A central tension exists** between the need to preserve cultural resources, and the dynamic potential for their use and reuse in democratic, just and compelling ways.

At the same time, the introduction of the **tetrarchy** of FAIR Guiding Principles (Findable, Accessible, Interoperable, **Reusable**) for data management and stewardship has set an important challenge: *that each of the four principles is of equal importance and must therefore be engaged with equally.*



# The central tension...

Infrastructures allow **static resources to be updated and cross-searched**, but the metadata for these assets **must be mapped in a centralised and controlled way**.

This reflects the types of terminology and relationships **defined by the data creators, and those charged with archiving and disseminating the data** (like ADS) not those who might use the data in **new and innovative ways**.



# Goal of TETRARCHs

TETRARCHs is working to provide those who capture, curate and apply cultural heritage data with **critically aware methodologies to prepare their data for enhanced reuse**, then experiment with such reuse **through storytelling scenarios** involving cross-European audiences.

As both an early adopter and user of a wide range of digital methods, archaeology is an ideal lens through which to develop and test these methodologies and scenarios.



# TETRARCHs Partners

Three-year project funded by the CHANSE ERA-NET  
Co-fund programme

Six partners representing five countries

**Project Leader - Sara Perry**

Anna Simandiraki-Grimshaw

**MOLA** (Museum of London Archaeology)

**University of York:** ADS (Holly Wright) + Department of  
Archaeology (James Stuart Taylor and Colleen Morgan)

United Kingdom



# TETRARCHs Partners

Rimvydas Laužikas  
Ingrida Kelpšienė and  
Indrė Jovaišaitė-Blaževičienė

**Vilnius University**

Lithuania

Edisa Lozić  
Benjamin Štular

**Znanstvenoraziskovalni center  
Slovenske akademije -  
Department Inštitut za  
arheologijo (ZRC SAZU)**

Slovenia

Nicoló Dell'Unto  
Paola Derudas

**Lund University**

Sweden

Hélène Verreyke  
Piraye Hacıgüzeller  
*Aida Fadioui*

**University of Antwerp**

Christophe Verbruggen  
*Lise Foket*

**Ghent University**

Belgium





# TETRARCHs Workpackages

- **Co-Design and User-Centred Development and Evaluation** (Lead: MOLA + all PIs)
- **Data Mapping Strategy** (Lead: Antwerp + Ghent, MOLA)
- **Data Capture Strategies** (Lead: ZRC SAZU + York, Lund)
- **Data Experimentation** (Lead: Lund + York, ZRC SAZU)
- **Repository Experimentation** (Lead: York + Vilnius)
- **Quality in Use Analysis for Archaeologists** (Lead: Vilnius + York, MOLA)
- **Storytelling and Creative reuse** (Lead: MOLA + all PIs)

# TETRARCHs

## Data Mapping Strategy



# Metadata Methodology

Under development by the partners  
Piraye Hacıgüzeller and *Aida Fadioui*  
University of Antwerp

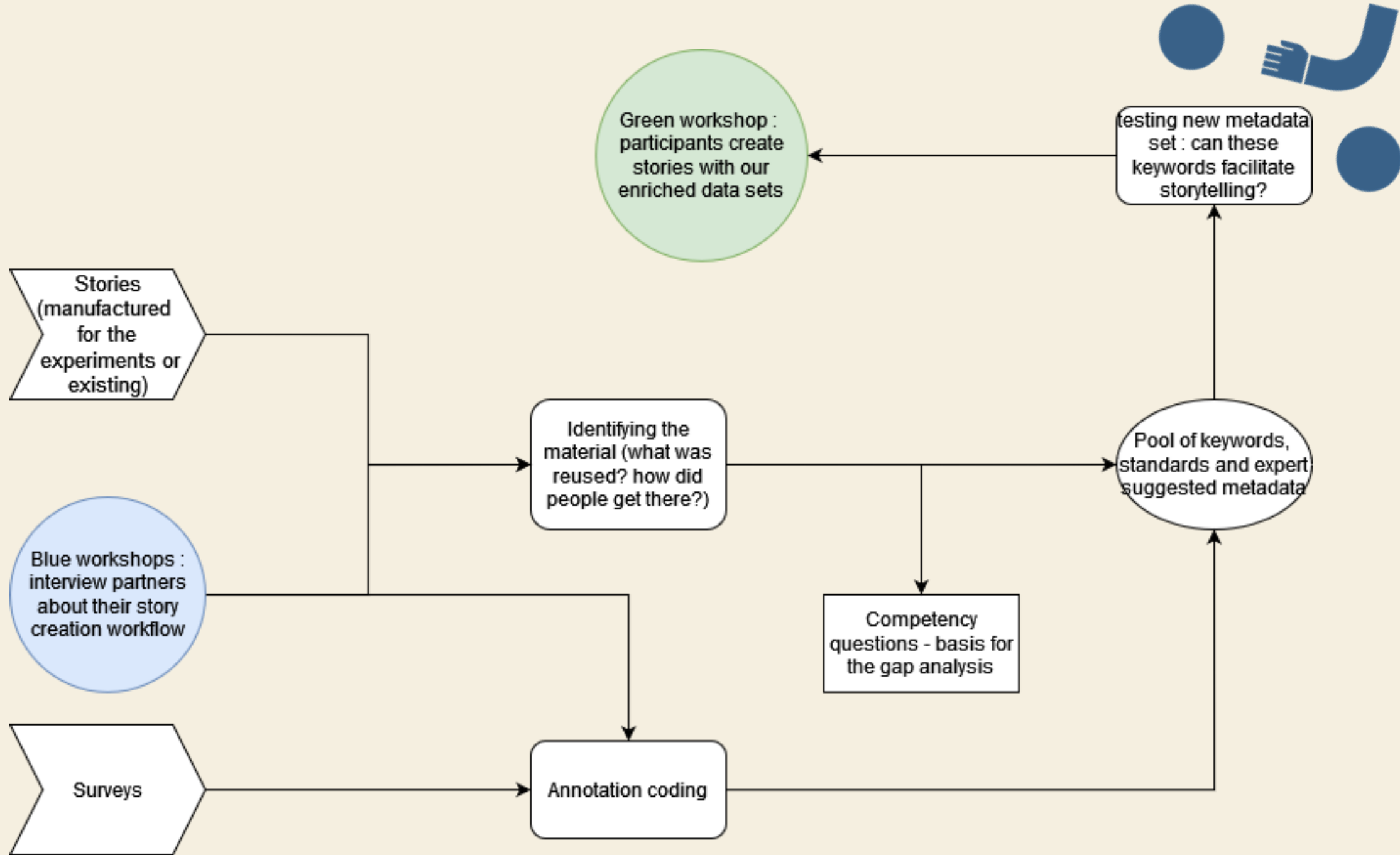
Using the *Co-Design and User-Centred Development and Evaluation* methodologies, a range of data gathering exercises have been undertaken by Antwerp, MOLA and Lund, involving narrative-building responses to text, photos and 3D models.

Exercises took the form of annotation and reuse experiments, surveys, focus groups, workshops and existing story analysis, and the data produced by these workshops is forming the basis for the data model.

*For example, users were asked to annotate images with words or phrases to build a narrative by the participants using the types of words or phrases that were most useful and meaningful to them. These annotations were then put through a process of concept coding, which was then categorised and synthesised to form classes.*

Currently these classes are part of a “keyword pool”: some of which can be mapped to existing ontologies and vocabularies, while those that can’t are coded to create new metadata classes.





## Codes

<input checked="" type="checkbox"/> Name	Files	References
<input type="checkbox"/> find or feature type	8	176
<input type="checkbox"/> condition	8	64
<input checked="" type="checkbox"/> material - size - quality	7	60
<input type="checkbox"/> function	6	38
<input type="checkbox"/> texture - materiality	6	12
<input type="checkbox"/> Feelings	6	119
<input type="checkbox"/> curiosity	6	42
<input type="checkbox"/> confusion	5	10
<input type="checkbox"/> surprise	3	7
<input type="checkbox"/> Sensations - stimuli	6	98
<input type="checkbox"/> Actions	8	96
<input type="checkbox"/> Physical features - description environment	7	73
<input type="checkbox"/> modern infrastructure	6	21
<input type="checkbox"/> natural features - landscape	4	11
<input type="checkbox"/> F9 Place	8	58
<input type="checkbox"/> excavation site	4	20

## Codes

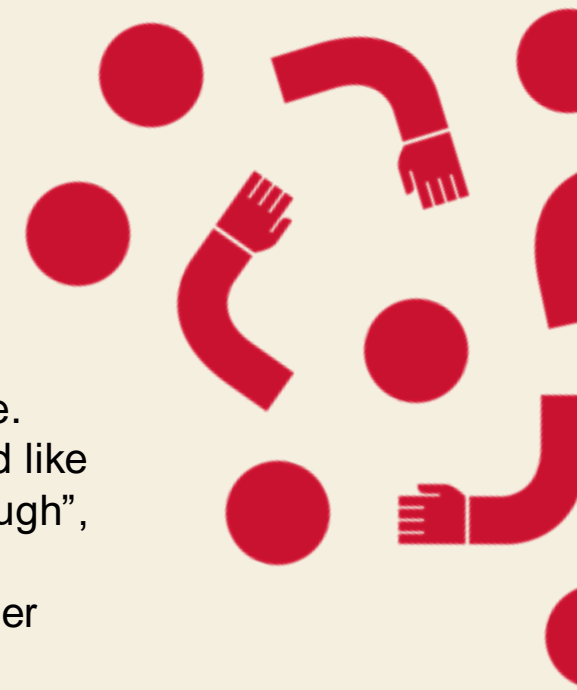
<input checked="" type="radio"/> Name	↔	Files	References
<input type="radio"/> Time		8	54
<input type="radio"/> comparaison - interpretation - process		5	49
<input type="radio"/> Linked discipline		7	48
<input type="radio"/> Archaeological process and technical information		6	44
<input checked="" type="radio"/> Media characteristics		7	40
<input type="radio"/> contrast		5	22
<input type="radio"/> POV		3	3
<input type="radio"/> atmosphere		6	31
<input checked="" type="radio"/> Media type		3	29
<input type="radio"/> game		4	13
<input type="radio"/> podcast		4	11
<input type="radio"/> Outreach~community		4	22
<input checked="" type="radio"/> religious - sacred - death		8	21
<input type="radio"/> burial type		2	14
<input type="radio"/> biblical concept - greed		1	1
<input type="radio"/> creativity and craft - human process		6	20

## Codes

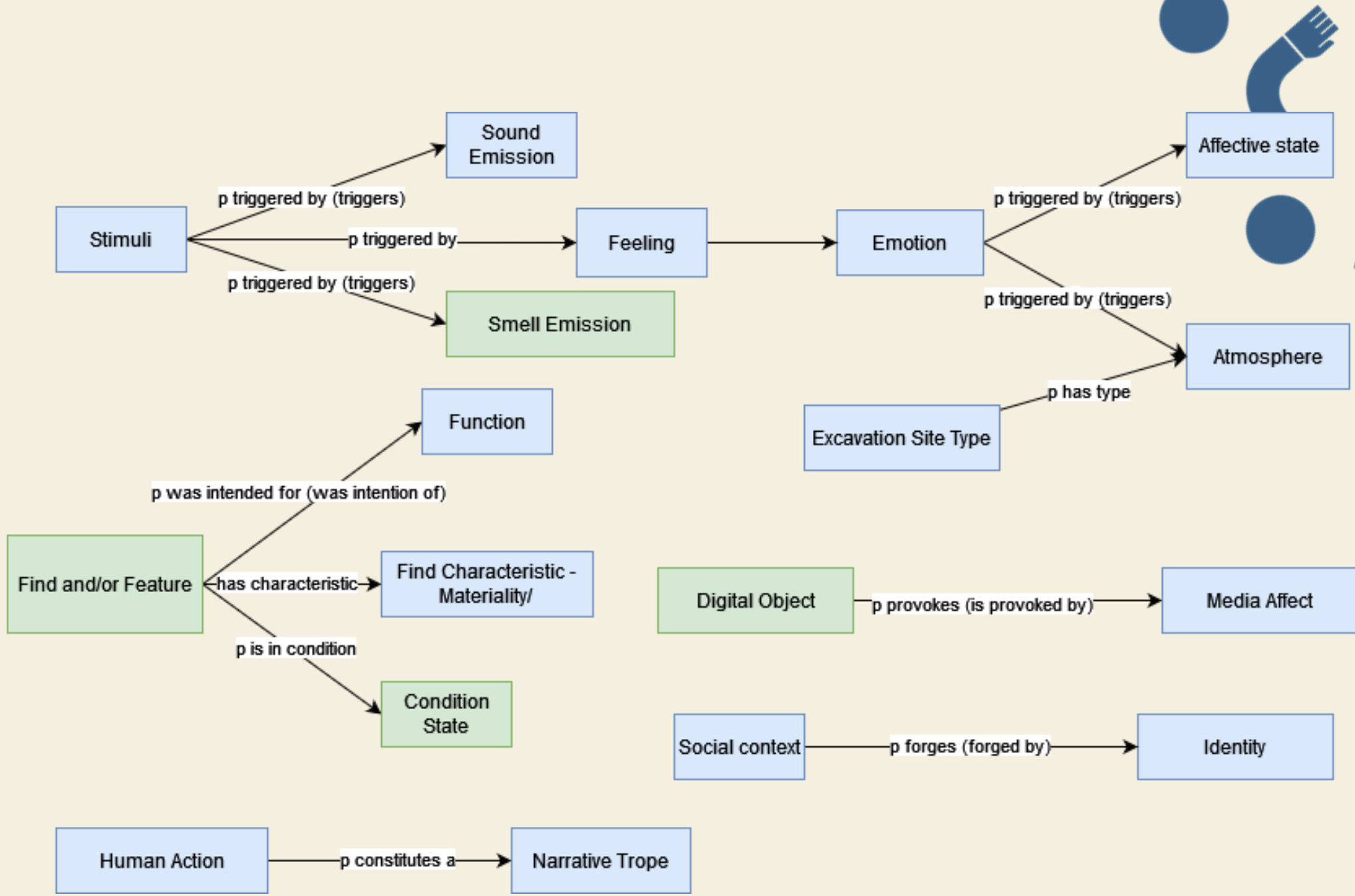
<input checked="" type="radio"/> Name	↔	Files	References
<input type="radio"/> E21 Person		4	20
<input type="radio"/> E5 Event		3	19
<input type="radio"/> Sound		5	18
<input type="radio"/> Stratigraphic layer		4	17
<input type="radio"/> embodied experiences and expressions		5	17
<input type="radio"/> Characteristics and actions relative to the people		3	15
<input type="radio"/> E2 Temporal Entity		3	14
<input checked="" type="radio"/> Mood		5	14
<input type="radio"/> reflection		6	23
<input type="radio"/> Empathy		3	13
<input type="radio"/> Nostalgia		4	13
<input checked="" type="radio"/> context - lense		2	14
<input checked="" type="radio"/> lived experience		5	8
<input type="radio"/> gender		2	4
<input type="radio"/> Indigenous		1	4
<input type="radio"/> class		2	3
<input type="radio"/> disability		1	1
<input type="radio"/> environnement issues		1	1

# Preliminary Results

- Highest number of annotations refer to **type of find or feature** and are primarily descriptive. Many refer to their **condition, but also their material, size, aesthetic quality** (words used like “rare”, “special”, “delicate”, “unusual”, “beautiful”), their function, texture and materiality (“rough”, “smooth”, “heavy”).
- **Feelings/emotions** represent the second largest number of annotations. Those coded under “curiosity” include questions about the find or feature themselves (material, function), their interpretation and the excavation process (already seeing where potential linkages are forming).
- **Actions** represent the third largest group, followed closely by **sensations** that reference one of the five senses.
- **Time** can refer to seasons, time of day, historical periods
- **Media characteristics** refer to things like the composition of the photos, the quality of the 3D models or text
- **Context or Lense** includes elements such as “feminist perspective”, social context at large, indigenous vs. non-indigenous, etc.
- Begun to **see the common threads between the codes** and are working to refine them and understand how they’re interrelated.
- Classes were determined based on how the codes were represented, but also based on which elements were **deemed useful to supporting storytelling** based on user requirements.







# TETRARCHs

## Workshop on Data Evaluation and Recommendations



# Data and Workflow Evaluation

## Repository Experimentation Workpackage

Focuses on resolving the point of central tension between the need to preserve cultural resources, and the dynamic potential for their use and reuse.

With ADS as a test case, the WP aims to determine if/how data optimised for co-designed and user-centred reuse can be incorporated into digital preservation workflows.

Purpose was to evaluate the data mapping strategy regarding its appropriateness for accessioning, preservation and dissemination in an accredited repository with a focus group comprised of representatives who have worked extensively with data modelling and mapping, staff from ADS and from other accredited repositories that hold archaeological data (DANS, SND, HES).





# Focus Group Questions

Sara Perry gave a short presentation on the background and aims of the TETRARCHs project, followed by a presentation by Aida Fadioui on the work being undertaken at Antwerp, followed by three 20 minute discussion sessions were undertaken for each of the sets sets of questions:

## Initial Impressions from the Group

- As data management and data mapping professionals, what is your initial impression?
- Does it follow what we currently think of as best practice? If not, in what way does it differ? Does it matter?
- Are there additional things we should be considering?



# Focus Group Questions

## Questions for Data Mappers

- Do you see issues around the way the data model is structured?
- Should we try to incorporate elements of CIDOC CRM, or other existing vocabs/thesauri or ontologies or map to them later?
- Do you see issues with interoperability that should be considered?

## Questions for Data Managers

- What aspects of this do you think will be problematic for your existing data management workflow? What aspects will be easy to implement?
- What do you see as the reuse implications of incorporating a data model like this? What are the challenges? What are the opportunities?
- Will incorporation of this type of model change the way you think about how you manage your data and make it available for reuse?

# Discussion and Results



# Discussion

## Initial Impressions from the Group

**As data management and data mapping professionals, what is your initial impression?**

- Generally very positive, with multiple comments about appreciating being able to accommodate subjectivity by using a storytelling approach.
- Concern that the result of the work is still going to be imposing a structure, and perhaps will not that useful for storytelling in the end. Should also be looking at how media outlets are handling this type of content.
- Response from Sara Perry that even going through these exercises is giving us **really useful data about processes by which data are collected** in the first place, the prompts that we're using, the way our digital recording systems are operating. It's also helping us to **explore the broader storytelling landscape inside archaeology** so it can have an impact on data managers and data mapping professionals going forward.

# Discussion

## Initial Impressions from the Group

Does it follow what we currently think of as best practice? If not, in what way does it differ? Does it matter?

- **Best practice is whatever promotes reusability and widens the impact of what we hold, so it fits perfectly.**
- It's about **changing our mindset about what best practice is**, and everything flows from that, which is a very different way of thinking about it than we have thought about as digital preservation specialists.
- Concern about the incremental ways we build our datasets, some of it happens in the field, some during post-ex, but always being processed in some way and skewed towards what we need to put into a publication. **Hard to envision what different types of narrative outputs might look like and how we might cater to them**, when we are so geared toward this specific type of output.





# Discussion

## Initial Impressions from the Group

### **Are there additional things we should be considering?**

- Need to bring funders into the conversation, as they tend to be focussed only on traditional outputs
- At the same time, this project was funded without the traditional need for some sort of technology output (new app, new search interface, new visualisation tool) which feels like a complete departure compared to how we have needed to structure our funding applications in the past.
- We shouldn't shy away from this type of work, because it will likely be seen more positively as we go on.

# Discussion

## Questions for Data Mappers

- Again, **the response was very positive** but there was a good discussion about whether we should be modelling the narrative elements separately and then linking them to existing vocabularies/ontologies, or if we should be using existing vocabularies/ontologies as the basis for the structure and then adding to them.
- **Consensus from the data modellers/mappers was surprisingly unconcerned.** For the sake of our sanity, they seemed happy that we should model the data as we see it emerging and worry about making links and/or mappings to existing structures later. If we were mindful of what was out there already and prepared to make those links/mappings, they were happy.
- **They deemed it much more important that we were figuring out ways to accommodate more voices than adhering to any structural purity.**

# Discussion

## Questions for Data Managers

- **Discussion centred on power** and who has it within an organisation or project. If the person responsible for managing the data is receptive, then change can happen, if not, it won't.
- Consensus was that **change will happen incrementally at different rates in different places**. That doesn't mean it isn't happening, but it is up to projects like TETRARCHs to demonstrate how easy or difficult this type of change is, and what the returned value can be.
- Lots of emphasis on the fact that what we are trying to accomplish in TETRARCHs is quite humble, but a **recognition that we are just at the beginning in this research area**. The point is to not to just say we should be doing these things, but to try; to make decisions that result in action even if those decisions aren't perfect. What we are trying to do can never be perfect and inclusive of everyone and everything, but we don't know how to move closer to that ideal if we don't create practical implementations that can be critiqued and adjusted.

# Results

## **The results were a bit...well...shocking!**

There was excellent and nuanced discussion about detailed aspects of the questions, but overall the consensus was that there were virtually no issues that couldn't be addressed or accommodated into best practice for data mappers and data managers. In fact, they were supportive and even enthusiastic about the “chaos” we were proposing.

This is a sea change in attitude, but still met with pragmatism about the challenges. So, full steam ahead from the workshop attendees.

## **What does CAA think?**



# Get in touch

holly.wright@york.ac.uk

WWW.TETRARCHS.ORG



TETRARCHS is supported by the Arts and Humanities Research Council (AHRC) in the UK, the Research Council of Lithuania, the Ministry of Education, Science and Sport in Slovenia, the FORTE Swedish Research Council for Health, Working life and Welfare, and the Research Foundation – Flanders (FWO) in Belgium under the CHNSE ERA-NET Co-fund programme, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement no 101004509.