TOWARDS A NATIONAL COLLECTION

UKRI Arts and Humanities Research Council

SECOND REPORT
DISCOVERY PROJECTS

Unpath'd Waters: Marine and Maritime Collections in the UK

MAY 2023

Table of Contents

Authors

Barney Sloane (Historic England), Julian Richards (Archaeology Data Service, University of York), Fraser Sturt (Centre for Maritime Archaeology, University of Southampton), Ann Coats (University of Portsmouth), Michael Roberts (Centre for Applied Marine Sciences, Bangor University), Vince Gaffney (School of Archaeological and Forensic Sciences, University of Bradford), Stuart Jeffrey (School of Simulation and Visualisation, Glasgow School of Art), Andrew Choong Han Lin (National Maritime Museums), Sara Perry (Museum of London Archaeology)

Executive Summary

The UK has a rich maritime heritage, stretching back over 23,000 years. It is impossible to tell the story of our islands without talking about our relationship with the sea. This maritime past is becoming increasingly important. People are more aware of our exploitation of the sea and topics such as colonialism, slavery and immigration.

Unpath'd Waters therefore aims to increase interaction with the UK's maritime heritage by making it easier to research and easier for the public to discover and share stories in new ways. Despite its importance, it is not always easy to study our maritime heritage. Records, maps and objects are scattered across hundreds of different archives, museums, libraries and galleries. A large part of our work is to develop new ways of making information across these different collections easy to search and find.

This will help everyone – from researchers asking new questions to members of the public having direct access to records. We hope this will encourage more experts from all disciplines to use maritime collections in their own work. To make sure this project has a lasting impact, we will publish all our methods, code and research so anyone can use it in their work and help the future of UK maritime heritage.

Unpath'd Waters is tackling the challenge through eight key activities (we have called them Work Packages). The first is aggregating and assessing the character of more than 90 different maritime collections, matching their core attributes and creating a metadata framework which will allow them to be connected. The second is testing artificial intelligence and machine learning opportunities to help search these connected collections in new ways. Three connected research activities will then test this framework and these new tools to try to resolve three real-world challenges. *People and the Sea* is looking at how people value wrecks in museums (the *Mary Rose* and the *Holland 1*) compared to some wrecks still on the seabed of the English Channel which have been surveyed and even excavated in part. *Science and the Sea* is using the connected collections to find the identity of wrecks in the Irish Sea and establish their stories and their management needs. *Lands Beneath the Sea* is using collections to build a simulation of now-lost landscapes under the North Sea where prehistoric peoples once dwelled before sea level rise.

Using the outcomes of this research and the collections framework and exploration tools, the project will then reach out to some key sample audiences: cross-disciplinary researchers, visually impaired people, and those who live far from the sea and have little connection with it. These audiences will help us co-design new ways of accessing this connected information so that we can create new portals for the public which are designed around what people themselves want rather than what we think they want. On top of this, we expect to be able to enhance the national inventories of all four home nations – England, Scotland, Wales and Northern Ireland.

Finally, we will prepare a report with the lessons we have learned and recommendations on what works best to help the UK move towards a genuinely national collection.

Unpath'd Waters is a big project and has a large number of people working on it. It is led by Barney Sloane of Historic England aided by a programme manager and administrative support. A Management Group of experts in digital humanities, marine archaeology and history, ocean science and audience engagement and

evaluation leads a consortium of eight universities, three Independent Research Organisations and five collaborating organisations including archaeological trusts, museums and charities.

This consortium is working well and the research programme is proceeding as planned. Real impacts are beginning to arise from the research, in addition to those reported previously (relating to ontologies and metadata standards). Our work on national inventories has exposed and begun to offer solutions for the historical inconsistencies across the UK, and to show how these data could be enriched. Work with student cohorts has shown the potential for co-designing visualisation solutions for connected collections. Integration of marine environmental data with heritage and commercial vessel data is starting to provide insights into the interdisciplinary value of connecting collections. Our work on provision of access for visually impaired people has generated new relationships. A number of interim recommendations are now emerging which we hope will help frame the business case and strategic approach to developing a National Collection. New partners have joined us as a result of the gradually improving visibility of the project, such as the British Geological Survey and Lloyd's of London (who join the Lloyd's Register Foundation). Both have provided key datasets relating to their collections. Manx National Heritage, another partner, has used Unpath'd Waters as a catalyst for significant investment in their own national marine heritage inventory.

We have provided access to the integrated national inventories via the UNPATH portal alpha version made public for user testing; and we are showcasing UK maritime data via the international ARIADNE RI portal.

We have also developed our project Values Framework – a crucial ethical consideration in establishing genuine representation within the project consortium and with our public audiences whose support will be critical as we move forward with the project.

We have also published a project website www.unpathdwaters.org.uk where we will provide updates on the work of the project.

The products of Unpath'd Waters will be held securely and sustainably beyond the lifetime of the project and are being developed to accord with the FAIR principles, so as well as informing the strategy for Towards a National Collection, the products of the project will outlast it and be available for others to use.

Interim recommendations arising from the project are as follows:

- Discovering collections: scoping of a future infrastructure should be clear about the degree to which it involves metadata or rich datasets. Both are needed to unleash the potential of a true national collection and a clear criterion-based strategy will be needed.
- Analogue to digital: standards, guidelines and principles for digitising current analogue collections will be required to ensure FAIR outputs.
- Authority and control: the extent to which authority files and controlled vocabularies should be imposed is a complex ethical and technical issue. A strategic approach will be needed to ensure a balance for different kinds of collections.
- Artificial Intelligence: Given the rapidly expanding capabilities of AI, the need for tracking of how outputs and labels have been created will be essential, especially for collections used for legislative or strategic management purposes.
- Dynamic data: some collections are designed to grow and change (eg national inventories). Others are static (eg a historical collection of objects). Efforts to enhance the former will require workflows that allow enhanced data to feed back into original repositories and be recognised there.

- Connectivity: formatting of new, future collections will be crucial in ensuring that the full range of abilities of users (for example the visually impaired) is built-in at the outset. This has significant implications.
- Ethics: related to connectivity, we recommend that human and more-than-human centred design is made a core component of any future national collection, at an infrastructure as well as collection level.
- Inherent issues in collections: collections created in the past use language and approaches rooted in that past. Linking such collections may surface difficult and inappropriate language and consistent approaches to this risk are required.

Abstract

The UK Marine Area covers 867,400 km2, 3.5 times the terrestrial extent. Our marine heritage is extraordinary. Shipwrecks from the Bronze Age to the World Wars bear testimony to Britain as an island nation, a destination for trade and conquest and, in the past, the heart of a global empire. Coastal communities have been shaped by their maritime heritage with stories of loss and heroism. Deeper in time, what is now the North Sea was dry land, peopled by prehistoric communities. Our current land would have been distant uplands above hills and plains and rivers now lost and forgotten.

Numerous collections represent this heritage, covering 23,000 years, including charts, documents, images, film, oral histories, sonar surveys, seismic data, bathymetry, archaeological investigations, artefacts, artworks and palaeo-environmental cores. These are unconnected and inaccessible. This matters because the story of our seas is of huge interest to the UK public, with millions visiting maritime museums annually, and marine exploitation increasing dramatically for energy, minerals, trade, food and leisure.

To unlock new stories and effect sustainable management, we must join up our marine collections. Unpath'd Waters brings together universities, agencies, museums, trusts and experts to confront this challenge. AI is being applied to innovate searching across collections, simulations to visualise landscapes, and science to identify wrecks and research their artefacts. Unpath'd Waters will deliver management tools to protect our most significant heritage and invite the public to co-design new ways of interacting with the collections. The methods, code and resources created will be published openly so they can be used to shape the future of UK marine heritage.

Aims & Objectives

Unpath'd Waters is split into Work Packages (WPs). These are designed to run concurrently and reflexively, informing and being informed by each other, forming a coherent research arc aiming to characterise and organise the collections, resolve obstacles to cross-searching, apply these tools iteratively to real-world marine challenges and develop mechanisms to engage the public in co-designing new access routes.

Throughout this report we refer to the project as UNPATH.

WP1 and 2 are addressing the more technical challenges of connection, access and searching, WP3.1 – 3.3 represent 'test lenses' to work through real-world research demonstrators, and WP4 and 5 are focused on visualisation, audience engagement and evaluation. WP6 is the underpinning programme management work. The five research WPs have the following research objectives.

WP1 Aggregation and Characterisation: *How can we integrate the UK's marine heritage collections?* WP1's objectives are to assess, enhance and aggregate the digital components of the UK's marine heritage collections; review the demands and capabilities of existing digital infrastructures; and enhance these to deliver the UK marine heritage collections to researchers and the public via human and machine-readable interfaces. It will develop appropriate data standards and cataloguing practices to achieve interoperability and deliver sustainable access routes and a primary data archive for UNPATH data outputs. This will lead directly to major improvements in national marine heritage inventories, including Coflein (Wales), Mariner (England), Canmore (Scotland) and, in due course inventories for Northern Ireland and, beyond UK waters, the Isle of Man.

WP2 Discovery: *How can we transform our ability to search those collections?* WP2's objectives are: to harness recent developments in artificial Intelligence and machine learning and build on work done in the Towards a National Collection Heritage Connector Foundation Project, to enrich sparse data, increase interoperability and enable a diversity of search options. It will do this through adopting a low-shot/zero-shot approach to machine learning leveraging the assets created in WP1 via named entity recognition to open linked data. In addition, computer vision (object detection) and increased spatial/temporal awareness/visualisation will be explored within the cases studies detailed in WP3. This will enable distinct modes of discovery and recovery (text, image, 3D shape, sound, spatial, temporal).

WP3.1 People and the Sea: *How can we enhance the significance of submerged and displayed wrecks?* WP3.1 will use a sample of submerged Protected Wrecks (and nearby unprotected wrecks) in the English Channel (The Needles, Mary Rose and Holland 5) and displayed wrecks (Mary Rose and Holland 1) by employing new stakeholder processes for elevating the significance of underwater wrecks in relation to displayed wrecks. Approaches will test the capacity of connected collections, including wreck site surveys, recovered artefacts, documentary records and scientific samples, to engage audiences in co-creating new narratives and innovative ways of engaging with the wrecks. WP3.1 aims to evaluate these novel approaches in enhancing public availability of wreck databases, serving as proof of concept for accessing other wreck sites and developing strategies for Protected and unprotected wrecks.

WP3.2 Science and the Sea: *How can we use collections to identify located wrecks and improve their management?* Using an Irish Sea pilot, WP3.2 aims to test connectivity between disparate marine scientific collections to identify wrecks and enable more effective heritage management of shipwrecks through better understanding of their condition and integrity. WP3.2 aims to link a collection of high-resolution multibeam sonar surveys from >100 incorrectly identified or unknown wreck sites in the Irish Sea to other collections (shipping and

naval archives, vessel plans, insurance records, personal correspondence and photographs etc), to transform the value of heritage assets whose research access is limited typically to vessel names and location of loss data. WP3.2 will also test integration of site-related geophysical and hydrodynamical/sediment transport datasets and marine ecological data to enable site-specific assessments on localised site conditions, integrity and sustainability of the national collection of shipwrecks.

WP3.3 Lands beneath the Sea: *How can we open access to complex collections representing submerged ancient landscapes?* WP3.3 will pioneer a method of accessing complex digital scientific collections relating to the underlying processes that created the archives representing the inundated prehistoric North Sea landscape. These processes link users, data and unifying concepts between different datasets, including evidence for climate change and rising sea levels. As process-led access models are largely untried, a bespoke computer simulation will be created to allow users to experience the dramatically changing world which created the data collections, providing a more comprehensible, and interactive method of engaging with them. WP3.2 aims to enhance researcher and public access to an otherwise entirely occluded prehistoric landscape and to support far more accurate predictive models of likely archaeological potential in an area set to see massive exploitation for renewables.

WP4. Designing, Connecting, Immersing: *How can we co-design and co-create research and engagement access for new audiences?* WP4 will address issues of diversifying audiences, improving cross-collection and cross-disciplinary research by co-designing novel interfaces with integrated maritime datasets through the use of immersive systems, through three user-group case studies (see WP5). Objectives are to test: (1) What new affordances do integrated maritime datasets offer cross-disciplinary researchers? (2) Can modes of navigation be co-designed that counter the complexity of multiple datasets, from multiple sources and in multiple formats (2D/3D) which work for both research and public audiences? (3) Can we reach audiences that do not traditionally engage with these data but could directly participate in the design of access modes that encourage their future engagement? (4) Can the technical challenges posed by User Generated Pathways through the UNPATH datasets be meaningfully addressed via multi-modal immersive interfaces (i.e. delivered as full immersive systems, but also via e.g. desktop and mobile devices)?

WP5. Engaging and Evaluating Audiences: *How can we ensure that the audiences are truly representational, empowered and their inputs formally evaluated?* WP5 will underpin WPs 3 and 4 by aiming to identify, engage with, and evaluate impacts on users. WP5 aims to: (1) discover shared values for the project (and to guide audience-focused work) in line with the [Digital Culture Charter](#), and develop metrics to assess project accountability to the values; (2) conduct audience mapping, to specifically identify and articulate a programme to ensure representation of three key target user-groups: non-coastal (inland) communities who do not identify with the sea; visually-impaired publics; and cross-disciplinary natural science and cultural researchers; (3) facilitate co-design with these audiences (qv WP4) and assess the efficacy and inclusiveness of the co-design approach; (4) facilitate testing and evaluating of the pilot designs that emerge through the interaction of WPs3 & 4 with our target audiences; and (5) coordinate and evaluate an exhibition programme of the final designs, including enrolment of other museums partners and industrial and third sector collaborating organisations into the programme to connect with their existing initiatives.

Partnership Structure

The partnership structure for Unpath'd Waters is set out in Fig 1 schematically.

Historic England (Barney Sloane) is the Principal Investigator (PI), supported directly by a Programme Manager (TaNC-funded) and an Administrator (HE funded). An external Advisory Board provides the PI with an independent sounding board.

The PI chairs a Management Group made up of the lead Co-Investigators from each Work Package (WP), the Programme Manager, Administrator and the Communications lead (HE-funded). This Group monitors progress, addresses cross-project issues and provides reports down to each WP.
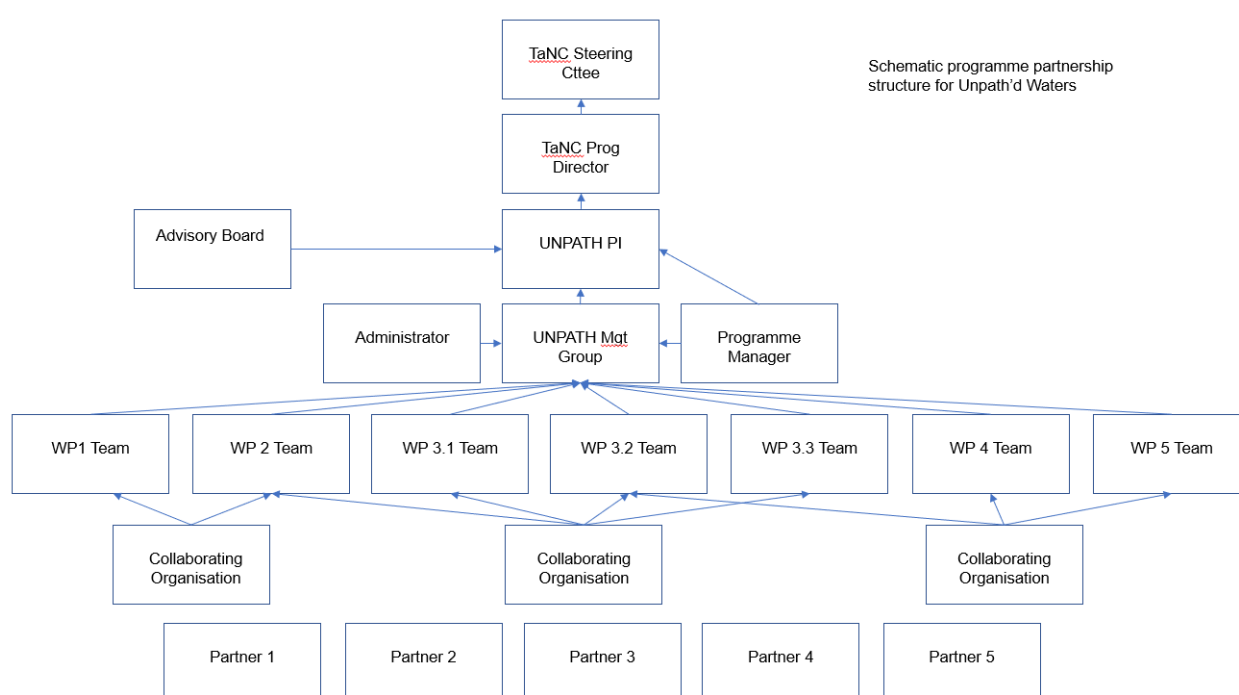

Fig 1: Schematic programme partnership structure for Unpath'd Waters

Work Package participants are:

- WP1: Lead: University of York (Archaeology Data Service). Co-Is: MEDIN, Historic Environment Scotland, Historic England; Collaborating Orgs (CO): Wessex Archaeology
- WP2: Lead: University of Southampton (UoS). Co-Is: Historic England. COs: RCAHMW
- WP3.1: Lead: University of Portsmouth (UoP). Co-Is: MOLA. COs: Mary Rose Trust, Nautical Archaeology Society, Maritime Archaeological Trust.
- WP3.2: Lead: Bangor University (BU). Co-Is: University of Ulster (UoU); COs: RCAHMW
- WP3.3: Lead: University of Bradford (UoBr). Co-Is: Historic England
- WP4: Lead: Glasgow School of Art (GSA). Co-Is: MOLA, University of St Andrews (UStA); Cos: Wessex Archaeology and Maritime Archaeology Trust

- WP5: Lead: MOLA. Co-Is: National Maritime Museum, Glasgow School of Art
- WP6: Lead: Historic England (HE).

Partner organisations are:

- Cadw (Welsh Government)
- Historic Environment Division, Department of Communities Northern Ireland
- Manx Heritage, Isle of Man's heritage department
- Marine Management Organisation
- Lloyd's Register Foundation
- Lloyd's of London
- Protected Wrecks Association
- The British Geological Survey

Staffing Structure

WP1: ADS Co-I Julian Richards leads WP1. MEDIN Co-I Clare Postlethwaite leader of the MEDIN data-providing organisations, will facilitate data access, and ensuring compatibility with the MEDIN. HES Co-I Peter McKeague will support Scottish (Canmore) inventory input. HE Co-I Paul Jeffery supports English marine inventory input and oversees RA Frankie Lau on UNPATH/Mariner integration. ADS RA Holly Wright supports RA Jamie Geddes on implementing the UNPATH ontology and metadata standards; and provides day-to-day management of WP1. ADS RAs Tim Evans and Paul Young support the upload of data to the triplestore.

WP2 UoS marine and archaeological computing expert Co-I Fraser Sturt leads WP2, manages RA Jack Pink and links to CO input, and leads on spatiotemporal data discovery tools. UoS data science Co-I Adrienne Chapman leads on application and integration of AI methods for search and discovery.

WP3.1 UoP social/cultural maritime historian Co-I Ann Coats leads, coordinating wreck collections, managing the core UoP RAs, coordinating CO input. MOLA education/outreach Co-I Caroline Barrie-Smith oversees RA Sherman in the development and evaluation of co-creative public engagement and ensuring links to WPs 4 and 5. UoP RAs Garry Scarlett, Robert Inkpen, Tarek Teba, Karen McBride are creating visualisations of wreck sites and investigating access to collections. UoP RA Claire Bailey-Ross, a digital heritage engagement impact specialist, supports a programme of immersive/viewer resources. MOLA RA Andy Sherman, community archaeologist, supports development and evaluation of co-creative public engagement approaches.

WP3.2 BU Marine geoscientist Co-I Michael Roberts leads and manages WP3.2, coordinating wreck and environment input from UoU marine geoscientists Co-Is Wes Forsythe and Rory Quinn; supervising RA (current recruit) and Kathy Blakey in linking collections; and coordinating CO input.

WP3.3 UoBr landscape archaeologist Co-I Vince Gaffney leads, overseeing UoBr RAs Phil Murgatroyd and (current recruit) in integrating diverse data sets and related software development programmes required for modelling; contributes to the visualisation development in WP4; and ensures academic outputs. RA Murgatroyd is providing agent modelling/simulation expertise to design and develop software and supporting material to provide user access to the landscapes. RA recruitment for marine landscape modelling expert, to continue integrating required archaeological, geophysical, topographic and core collections/data sets.

WP4 GSA heritage data visualisation expert Co-I Stuart Jeffrey leads WP4, managing development and delivery of the UNPATH Navigator. He is also coordinating RAs (Scott Carballo and Alan Matthews) and COs and liaising with WPs 3.1 and 5. UStA community archaeologist Co-I Tom Dawson is supporting the co-design and evaluation documentation process and will also develop Navigator content including integration of maritime heritage AV content.

WP5 MOLA audience engagement and evaluation expert Co-I Sara Perry leads, providing a values framework and ensuring digital equity and access across UNPATH, and supporting Co-I Barrie-Smith (WP3.1) and Co-I Jeffrey (WP4); NMM museum/collections expert Co-I Andrew Choong will facilitate access to museum partners and their target audiences; coordinate final dissemination/ exhibition activities while supervising an RA (formerly Lawrence Northall, now recruiting for his replacement) and support Co-I Perry in delivering other outputs of WP5.

WP6 HE heritage management specialist PI Barney Sloane will provide overall leadership (WP6) of the consortium, finances, programme and delivery; manage input of HE Co-Is Paul Jeffery (Partner coordination) and Antony Firth (marine strategy lead); and oversee RAs Lau (UNPATH/Mariner interface); Carla Venterop-Martin (UNPATH programme manager); and in year 3, RA (tbc), an evaluation specialist, who will help evaluate UNPATH using HE's Public Value Framework.

Collaborating organisation staff are: Wessex Archaeology (Graham Scott): Senior Maritime Technical Specialist providing WA datasets (WP1) and input on wrecks (WP3.1 and 5); Maritime Archaeology Trust (Julie Satchell and Brandon Mason): working on Needles wrecks (WP3.1) and audience engagement (WP5); Mary Rose Trust (Alex Hildred, Alastair Miles and Hannah Matthews) providing Mary Rose collections and audience access (WP3.1); Nautical Archaeology Society (Mark Beattie-Edwards) providing data and mapping for Holland 1 and 5 wrecks (WP3.1); Royal Commission on the Ancient and Historical Monuments of Wales (Julian Whitewright) providing access to Coflein datasets (WP1, 3.2).

Overall Programme

The following table provides a summary of the project's programme

	Start	Finish
UNPATH WORK PACKAGE 1: Aggregation & characterisation		
Ontology Development	Nov 21	Nov 22
Data Auditing	Nov 21	Jun 23
Portal Development	Nov 21	Aug 24
Data Uploads	May 22	July 24
Standards and Good Practice Guide	Mar 23	Aug 24
Data Archiving	Sep 23	Aug 24
UNPATH WORK PACKAGE 2: Discovery		
User Needs / Audience Mapping	Nov 21	Apr 22
Development	May 22	Oct 22
Explore use of NLP, zero shot and object detection on national datasets and WP case studies	Oct 22	Sep 23
Embed extended search option and publish	Sep 23	Jul 24
UNPATH WORK PACKAGE 3.1: People & the sea		
Data collection, review and analysis of sites and study areas	Nov 21	Nov 22
Prototype development	Jul 22	Apr 23
Working with audiences	Dec 22	Oct 23
Public & schools engagement	Mar 23	Jul 24
Analysis and evaluation	May 23	Jul 24
UNPATH WORK PACKAGE 3.2: Science & the sea		
Identification of wreck sites, relevant scientific data sets, historical archives	Dec 21	Aug 22
Creation of metadata catalogues, UNPATH science data toolkit & site ID methodology	May 22	Oct 22
Assessment of collection connectivity and methodology testing link data/archives/collections approaches development	Nov 22	Jul 24
Recommendations for candidate Protected Wrecks	Dec 23	May 24
Analysis and evaluation	Dec 23	Nov 24
UNPATH WORK PACKAGE 3.3: Lands beneath the sea		
Prototype terrain model	Nov 21	Nov 22
Climate-sensitive temporal model	Nov 22	Sep 23
Prototype integrated landscape model, tech testing and user engagement	Oct 23	Dec 23
Beta version of final product, tech testing and public engagement	Nov 23	Jun 24
Version 1.0 of software package	May 24	Sep 24

Version 1.1 of final product and design document/report	Oct 24	Nov 24
UNPATH WORK PACKAGE 4: Designing, Connecting, Immersing		
Define co-design methodology & recruitment	Oct 22	May 23
Audience engagement and co-design activities	May 23	Mar 24
Create Unpath Explorer technical workplan & technical framework	Nov 22	Sept 23
Curated Pathways co-design	Sep 23	Nov 23
UNPATH Navigator re-casting for multiple dissemination modes	Nov 23	Mar 24
Multi-modal output testing and evaluation	Mar 24	Sep 24
UNPATH WORK PACKAGE 5: Engaging and evaluating audiences		
Values Framework	Nov 21	Jun 22
Audience mapping	Apr 22	Dec 22
Specify and test audience evaluation methodology	May 22	Dec 23
Co-design research & methodology development	Oct 22	Dec 22
Co-design sessions	Dec 22	Nov 23
Agree outline exhibition plans	Oct 22	Mar 23
Exhibition development	Mar 23	May 24
Report on user analyses	Nov 23	Feb 24
Roll out audience evaluations & report on outcomes	Dec 23	Oct 24

Events & Consultations

Subject	Date	Number of attendees/responses
WP5 Values framework development		
Initial brainstorming workshop	8 Mar 2022	13
Consultation on draft values	Completed w/c 4 Apr 2022	13
Consultation with the Consortium on refined values	Completed 30 Apr 2022	76
WP5 Audience Mapping		
Consultation and data gathering with Unpath'd Waters Consortium	Apr - Jul 2022	76
Working with audiences		
WP3.1 Pilot student hackathon – task was to visualise biological research data, creating a visual immersive/interactive using genetic data from the Mary Rose	Jul 2022	15
Events		
Unpath'd Waters Consortium event, National Museums Liverpool	10 Nov 2022	35
UKRI leadership visit to Menai Bridge ocean research facility and vessel with presentation on Unpath'd Waters	1 Dec 2022	10
Talks & Conferences		
WP3.2 Science & the Sea (University of Bangor): public lecture by MJR at a Bangor U3A group meeting	25 Mar 2022	40
WP5 Audiences & Evaluation (MOLA): Transforming Collections Project Values Exploration Presentation, University of the Arts London	12 Apr 2022	50
WP3.1 People & the Sea (University of Portsmouth): Naval Dockyards Society Conference, National Museum of the Royal Navy Portsmouth: presentation by Antony Firth: Placing Warships: Reconnecting vessels and dockyards	9–11 Jun 2022	59 in-person; 11 online delegates
WP3.2 Science & the Sea (University of Bangor): presentations by MJR and JW at the Research Framework for the Archaeology of Wales mini-conference	27 Jun 2022	40
WP3.2 Cylch Glasinfryn public lecture	12 Oct 2022	30
WP1 Unpath'd Waters portal previewed in CHNT keynote conference presentation, Vienna	11 Nov 2022	c.120 attended in person or online

WP4 & 5: SJ and SP Invited panellists at ICOMOS-UK Digital Technology National Committee's Digital Technologies for Visitor Engagement with Cultural Heritage. Online.	29 Nov 2022	100
WP1 Aggregation & Characterisation (ADS, University of York): Hackathon, Linked Pasts Conference, York	29 Nov-1 Dec 2022	Triple store and portal tested by c.25 users
WG3.3 Invited lecture on Doggerland South Oxfordshire Archaeological Society	26 Jan 2023	70
WG3.3 Invited lecture to Intellectual Property Office. "Not drowning but waving". British Science Week lecture on theme of contacts (VG)	10 Mar 2023	105
WP3.3 Kiel Conference 2023: Scales of social, environmental and cultural change in past societies On course to drowned landscapes in the North Sea – cultural aspects, preservation potential and geoscientific challenges.". (VG)	14 Mar 2023	30
WP2 Discovery (University of Southampton), NLP workshop, Online	Jun 2022	35
WP2 Discovery (University of Southampton) TANC workshop (Greenwich) by F. Sturt	10 Dec 2022	40
"Why do we need real spaces for research using extended reality?": Introducing ARC-XR as a space for arts and humanities research with immersive technologies (Advanced Research Centre, University of Glasgow).	1 Dec 2022	20
Economic Growth & Cultural Tourism: XR Experiences, Produce, Sustainability (Advanced Research Centre, University of Glasgow	23 Jan 2023	40

Research Approach

The research design of the project is linear overall: An Incubator Lab in the first third of the programme has discovered, aggregated and characterised the collections, developing search and access tools, and setting the research basis for the three marine themes. The project has moved into an Innovation Lab, where we are applying the tools in real-world scenarios to test feasibility, resolve barriers. We will then employ these outputs to explore genuine innovation in co-design and co-creation capability to permit consumption, enhancement and evolution of the connected collections through visualisation, immersion and exhibition engagement. Impact Delivery, the final phase, will complete the work's evaluation and development of the formal documentation and final reporting outputs.

The seven Work Packages are progressing through these three phases, with iterative connections to ensure that lessons learned within one WP can be transmitted quickly to other WPs, providing robust course-correction flexibility. This design is ensuring that the project Research Associates, and in particular the Early Career Researchers, are benefitting from a full spectrum viewpoint and see their roles within the context of the entire programme. Further, this design integrates insights from industrial, third sector and heritage management organisations at each step of the project to ensure real-world applicability.

WP1 has undertaken a partner data structure survey and survey of existing data infrastructures, feeding development of the UNPATH core metadata model and core data standards. The range of collections included in this survey (and the resulting ontology, now confirmed) is considerable, covering 94 different collections which range from records of artefacts from shipwrecks, through to seismic datasets from the oil and gas industry which reveal submerged landforms. WP1 has enhanced existing portals with aggregation of UK marine data (national bodies – Coflein, Canmore, Heritage Gateway; ADS ArchSearch; MEDIN, ARIADNE); building upon work done by ADS in the EU-funded VENUS project, UNPATH is working to provide up-to-date data standards guidance in the ADS Open Access world-renowned Guides to Good Practice. FAIR and CARE principles sit at the heart of the UNPATH approach (Co-I Richards and Co-I Perry were PI and Co-I respectively on the TaNC *Making it FAIR* project).

WP2 is trialling Natural Language Processing (NPL) approaches with national agency and CO datasets. This focuses on extracting critical hard to search information contained within these data, but hidden within complex text, that will open routes to discovery. It is also helping to make the diverse datasets of our national collection more comparable and interoperable. This is achieved through ensuring that each dataset can map as well and informatively as possible onto the ontology created in WP1. Computer vision-based approaches are also being explored. This sees artificial intelligence approaches being used for identification of sites underwater in sonar data, as well as within images taken of objects in museum collections. We are working with WPs 3, 4 & 5 to undertake audience/stakeholder behaviour/engagement analysis (defining search and discovery ontologies) to assist with developing UNPATH search tools (text, image, sound, space, time) for application to these enhanced collections. The aim is to move towards 'free-association' search, where users are less constrained than they might be by a rigid ontology, and by enabling links beyond core datasets. Finally, WP2 has created 3D datasets for WP3.1 through laser scanning the Holland 1 submarine.

WP3.1 will deliver four strands. (i) Dissolve boundaries between museum-based Mary Rose and Holland 1 and the still-submerged Holland 5, Mary Rose and The Needles Protected Wrecks and adjacent non-protected wrecks, to explore how conflict or trade vessels connect individuals and coastal communities; relating to

construction, provisions, ordnance, crews, health, rescue and display; (ii) develop digital connectors linking disparate wreck artefact and archival collections; (iii) explore potential applications of Holland 1/5 real-time immersive digital experiences through XR; (iv) map the cultural and social attributes of the wrecks with contemporary communities, values and debates. These are being used to develop innovative co-designed, co-created wreck audience engagement and learning tools and packages and prototype connectors, enabling a range of consumers (Protected Wrecks Association, museum partners, wreck divers, volunteers, researchers, school/university students and the public) to interact with and potentially enhance connected collections. Co-design approaches have now been tested with post-graduate students at the University of Portsmouth acting as a test group, applying a hackathon approach where teams of students were invited to consider collections from the Mary Rose and create new approaches to access and public engagement.

WP3.2 has defined a sample of unidentified wreck sites and will identify maritime collections and data sets required to support wreck identification. A documentary and data tagging initiative is being developed by creating a search criteria classification system designed to link key relevant historical collections and data. In addition, a site characterisation and classification approach will also be constructed, linking marine collections and data sets and principal geological and hydrodynamical processes and associated biological conditions to create a mechanism to evaluate likely conditions and survival trajectories. The two protocols will be assessed in the context of the creation of appropriate adaptive management strategies, and on the applicability and relevance of this approach at a wider UK and international scale.

WP3.3 is developing a computer simulation to combine a variety of data sources to facilitate public access to Doggerland. Sources are disparate in spatial scale, storage size and computational demand (e.g. sonar bathymetric maps, seismic profiles, publicly available archaeological collections such as CITiZAN, local Historic Environment Records, ADS online library, and palaeo-environmental seabed cores). The European Research Council Europe's Lost Frontier's output (the modelled landscape at the beginning of the Holocene) forms the base. New data collections are being integrated to test and advance the simulation. The simulation's front end will ultimately allow the user to 'fast forward' from the last Ice Age, melting glaciers and changing sea level to any selected time. Through a small band of Doggerland hunter-gatherers inhabiting the landscape at the time specified, the user will be able to either observe or take significant decisions for the simulated community, thus accessing the data patterns of the submerged landscape via the processes which formed them. The integrated collections will not tell the story of Doggerland but allow users to find their own stories within it.

WP4 is setting about engaging with existing audiences, finding new audiences and communities and co-designing the interface underpinning new ways of exploring and generating novel and unexpected narratives from data which may use multiple channels, e.g. embodied and sound-based. It will reveal user requirements and preferences for engaging with interoperable collections. It will deploy a Rapid Application Development methodology to transform requirements into new and sustainable interfaces for three case-studies, addressing specific audiences, audience requirements and three key research themes. (i) Co-design and User Generated Pathways data layer for non-coastal communities. This case study will develop new audiences, engage via new interaction modes and is intended to surface unexpected, community specified, connections between the interoperable data. (ii) The visually impaired public, using spatialised audio content (e.g. derived from WP3.2 sonar datasets) to both represent and allow interrogation of interoperable datasets, and address barriers to access presented by traditional complex database interfaces. Using creative response this case study will be extended to generate novel means of engagement with wider audiences whose interest goes beyond the historic and scientific to the aesthetic, social or spiritual (qv Burra Charter); and (iii) Cross-disciplinary research, using natural and cultural heritage collections derived from WP3.3 simulation, allowing researchers from both

natural and cultural disciplines to explore and discover connectivity between these domains. The interoperable UNPATH datasets will be made amenable to multiple modes of access through an UNPATH 'Navigator' which will be demonstrated and promoted as a web-based resource, a mobile application, and an immersive VR interface (e.g. HMD) and installations evaluated and deployed via WP5.

WP5 is leading the co-design of evaluation methodologies for audiences of the products and applying those methods to assessment of the Incubator, Innovator and Impact components of UNPATH. Following the Digital Culture Charter and informed by the digital equity methodology of D'Ignazio & Klein (2020), WP5 began by establishing a values framework and associated metrics for the project and will audit progress against the framework on an annual basis. Via audience mapping, WP5 identified target audiences for recruitment into UNPATH in Year 1 and Year 2, and has worked between them and across WPs1-4 to ensure that project outputs accommodate user needs, that the project's exhibition programme is curated for accessibility, and that inclusive evaluation methods are developed and applied across the project to accommodate neurodiverse and visually-impaired participants. WP5 is now about to start coordinating user evaluation activities in the Innovator and Impact phases of the project, utilising (accessible) survey, interview and automated data acquisition tools. This includes an assessment of the efficacy and inclusiveness of the co-design approach itself at the close of the Innovator Lab, and analysis and publication (with WP3 and 4 leads) of learning, social and emotional outcomes of audience workshops, exhibitions and other dissemination activities. WP5 will ensure delivery of a final exhibition programme, with the National Maritime Museum coordinating museum partners and industrial and third sector collaborating organisations to extend the reach of the project further. The nature of the exhibition programme will be established in consultation with our key audience groups (including virtual and pop-up events), integrating (where possible) with existing programmes such as anniversary and thematic events. WP5 will also track the application of the research tools enhanced in the Incubator Lab across the life of the project, and map the novel interpretations, questions and research findings that emerge.

In summary, the research approach adopted by UNPATH is addressing the key Towards a National Collection Impact Areas in the following ways.

(a) **The exploration of how multiple collections can be explored in the future**. UNPATH does this by forming core metadata model and standards, linking disparate collections into a single schema, by testing those links through three real-world research queries, and by then freeing the results to be accessed by carefully defined invited research and public audiences using new tools co-designed by them themselves.

(b) **The dissolution of barriers and opening of public access**. The collections we are looking at range from born-digital to physical objects and from summary inventory data to point-cloud datasets. There are significant challenges (both technical and social) that we expect to come up against and either to resolve or to identify as currently beyond our ability to resolve. Some examples of these challenges include:

The analogue data and metadata of written sources, charts and records. Considerable numbers of archive, museum and library holdings remain analogue and while digital datasets connect to them, there is no capability to link. We are exploring a Digital/Analogue connector (which combines input from WP2, WP3.1 and WP3.2 particularly) which will at least begin to chart these connections (see Research Results, and Initial Recommendations sections below).

The complexities of some of the collections/datasets. For example: national inventories are not internally or comparably consistent; a 'wreck' can often include multiple sites (through drift of material); sonar survey technology over time has changed leading to differences within apparently coherent collections; some collections have a temporal dimension (seabed surveys, wreck surveys) complicating access.

By working with audiences to co-design pathways into these collections, we will develop a bottom-up approach to solutions which should resolve technical issues as well as create desirable access. The solutions are being designed to be transferrable.

(c) **The diversification of audiences and innovation of access**. The character and method of establishment of our three key target audiences (WP4 and 5) will ensure that UNPATH reaches a diverse group of potential users and can accommodate ranges of (dis)ability and neurodiversity. The 'Unpath Navigator' (WP4) will extend access (via web, immersives, mobile app and museum installations) to the interoperable datasets which we anticipate will provide a model for future access. We have selected the three audiences to test each of the three kinds of motivation identified by Bailey-Ross (2021) (i.e. Casual Use, Personal Interest, Scholarly & Professional Research). In the case of the researchers, we anticipate cross-disciplinary interest in the interplay between natural and cultural heritage collections (especially as a result of the work in WP3.2 on the integration of ocean and seabed sciences and historical and archaeological sciences).

(d) **Setting a global standard for others building their own collections**. Work on integrating and enhancing marine inventory collections (national datasets) interoperability will allow international aggregation (such as the completed Perio.Do definitions for all of Ireland). The protocols established in WP3.2 likewise will have an international value. UNPATH has already led to a Getty Art and Architecture Thesaurus mapping for maritime craft which represents an early impact on international standards and will deliver an ADS Open Access world-renowned Guide to Good Practice.

(e) **Enhancing collaboration between UK and national collections worldwide**. As the UK national data sets are aggregated in WP1 they become searchable in the UNPATH portal but are also simultaneously available via the international heritage portal managed by the ARIADNE Research Infrastructure, where they are interoperable with archaeology and heritage datasets drawn from over 30 other collections drawn from international heritage organisations, covering the archaeology of four continents and over 40 countries. International collaboration and impact will increase through the outputs above (d), and through international events where UNPATH will be promoted (European Archaeological Association, Computer Applications in Archaeology conferences). We are pleased to have colleagues from Ireland and the Netherlands on our Advisory Board.

Our involvement with our Partner Manx National Heritage (MNH) has been particularly valuable. In the words of their Collections Information Manager Jude Dicken, it has enabled them to capitalise on UNPATH benefits in developing their own inventory of marine heritage and in early 2023 they have been able to include their own national marine heritage data inventory within the UNPATH Portal. Their involvement acted as the catalyst for them to reach out formally to shipwreck research specialist Adrian Corkill devising a programme for him to work with them to create 1,919 underwater site and 1,074 vessel resources on the public website https://isleofmanher.im/. In February 2023, this data was shared and uploaded to the public portals https://unpathd.ads.ac.uk/ and https://portal.ariadne-infrastructure.eu/search?q=&publisher=Manx%20National%20Heritage. This comprehensive and authoritative dataset of wreck sites, compiled by Adrian over three decades but previously accessible only to him, is data now secured in the Isle of Man Historic Environment Record and discoverable by researchers and stakeholders. UNPATH has given buoyancy to a long-held aspiration by MNH to document and reveal more about the maritime heritage of the Isle of Man through data and the stories told through its collections and the Irish Sea wreck sites. Throughout 2023 MNH will continue to add maritime data to https://isleofmanher.im/ (including archive & bibliography and crew resources) and collaborate with the UNPATH data specialists to mine stories from the data. UNPATH will make an important contribution to the data and digital legacy at MNH in its telling and monitoring of maritime history.

Research Results

UNPATH continues to test boundaries of linking collections and providing access to them. We have made some highly significant advances.

WP1: The UNPATH WP1 programme is progressing on schedule. We have completed the primary dataset collation (c 94 different datasets) and established the UNPATH ontology (UO-CAT). This is a sub-set of the CIDOC-CRM ISO data standard, ensuring interoperability beyond the TaNC programme. We have also confirmed that this has a clear mapping to the MEDIN metadata standard, which will allow an easy export of core data in year 3. The project triplestore is operational, and in February 2023 we released a proof of concept public instance of the European ARIADNE portal badged and hard-wired for UK maritime data for UNPATH (https://unpathd.ads.ac.uk/).

The ingest of national level inventory datasets is almost complete, with 86,509 resources for maritime research made available via the triplestore and searchable in the UNPATH portal. This includes 50,580 records from the National Marine Heritage Record (NHMR) from Historic England; 33,379 records from the Historic Environment Scotland CANMORE database; 1,916 records from Manx National Heritage; and 244 records from the Department for Communities, Northern Ireland. This allows researchers to interrogate the aggregated datasets via a combination of "What" "When" and "Where" search parameters (Subject matter, Period dating, and Location), including both map and timeline functionality. Links are provided to the primary data resources, hosted by the national records. This allows users who have found the resources they are interested in, to drill down to the primary evidence, whether in digital format, or held as references to analogue paper resources. It is the first time the Isle of Man data has been made available online and this is a direct result of UNPATH. These datasets will be enhanced as the project progresses and work on the Welsh national record, Coflein, and the national Intertidal zone data set derived from the public participatory project CITIZAN (managed by MOLA) is ongoing. In addition, the portal provides access to 207 unpublished maritime fieldwork reports collected via the OASIS archaeological event recording system, and 82 rich fieldwork archives curated by the Archaeology Data Service, including those provided by collaborating organisation Wessex Archaeology.

We have established Perio.Do (ie chronological) definitions for Scottish and Irish data (including the whole of Ireland), and we have agreed Getty AAT thesaurus mapping for maritime craft. The AAT has proved to be adequate for our needs although it lacks granularity at the level of specific vessel sub-types and suffers from its one-dimensional hierarchy (e.g. should vessels be classified by type, or method of propulsion?) Nonetheless the portal subject search takes advantage of the hierarchical structure of the AAT thesaurus, resulting in powerful search capabilities across multiple data providers. A search for "cargo vessels", for example, not only returns the 18,198 vessels directly classified as Cargo vessels, but also 261 Lighters, 123 Tankers, 185 Colliers etc. We have worked with the Mary Rose Trust to assist them in mapping their subject vocabulary to the Getty AAT thesaurus.

The UNPATH data is also simultaneously available via the international portal managed by the ARIADNE Research Infrastructure AISBL (https://portal.ariadne-infrastructure.eu) as it draws on the same Linked Open Data triplestore.

We have undertaken a light touch FAIR audit of the data being drawn upon and collated by all the WP3 research lenses to assess the level of FAIR compliance, identifying opportunities for improvement during the project, or simply reporting on a microcosm of the challenges faced by a future National Collection. The FAIR audit is designed as a living document and will continue to be updated through to the end of the project.

WP2: Considerable progress has been made within this work package focused on discovery utilising artificial intelligence (AI) and machine learning (ML) methods. We have received copies of the national heritage agency datasets, mapped to the UNPATH ontology from WP1. Analysis of these datasets revealed that their structure and content led to varying levels of insight being gained through queries based on this ontology: in some cases datasets allowed for detailed information on ship type to be searched (e.g. Schooner), while in others all lost vessels were referred to as 'shipwrecks' or just 'wreck'. The broad nature of these mapped terms hid more specific details contained within rich text in more generic 'notes' or 'description' fields.

To access this information we have trialled zero shot (ZSL) and few shot (FSL) named entity recognition (NER). In zero-shot learning the computer is asked to analyse a dataset and extract what it sees as key terms (using language models to identify these entities). In few-shot learning the computer is given a limited dataset, labelled by a domain expert, to help it understand extract entities from a supplied authority file. In each case the output is a modified dataset, with new fields included, reflecting what the computer has determined to be the appropriate output.

Trials have been completed with two datasets where ontology mapping only gave broad terms, but for which associated data was rich (Coflein and the Isle of Man). This was achieved through use of a series of different machine learning base-models; SPaCY, Google BERT, Facebook BART and compared against SVM and Naïve Bayes models. Although each of these models is publicised as a highly flexible NER system, significant differences have been found in the accuracy results for each. The best accuracy results have been extracted from Google's BERT algorithm. There are several reasons for this variance. Each of the models possesses their own variance as they have been coded and trained differently before their deployment as part of UNPATH. This means they also react differently to domain-specific training data provided for classification tasks. The nature of the data available for UNPATH means that for some classes there may only be one or two entities in an entire National Monuments Record that relate to each term or type. This highly varied degree of presence for each craft type means certain approaches are better suited than others. Of the ZSL and FSL approaches utilised, the highest degree of accuracy has been observed in the latter. Due to the presence of entities with only a handful of identifiable examples in the dataset (e.g., 'Passenger Liner' with only two entries in Coflein) the FSL deployment is often a 'One-Shot' model.

Specific explorations of the datasets themselves have demonstrated the potential to streamline the approach being taken. A significant proportion of the records within the two datasets could be termed 'simple' NER tasks and focus on word-matching an entity already identified in description text. For example, a record might describe "The ABBEY was a schooner registered at Beaumaris" in this case the task is to extract the craft-type "Schooner". These tasks do not require complex semantic or contextual solutions. We have also seen that a handful of craft types make up a majority of the records. This is a feature of the dataset that would influence the decision of whether a particular technique is needed or not. Finally, this process of enrichment and expanding tags is only required once for each of the National Datasets. Once the enrichment is complete and a format established for the specific fields in the dataset any new additions will contain this information by default.

The enrichment and exploration of the dataset has also highlighted the value of the rich "Description" fields within the National Monuments' Records. These are not only an essential source of information for the NLP and NER models but are also invaluable as sources of narrative and story related to each entity. As part of the work done with the National Heritage Agencies, a concerted effort is being made to ensure similar data is recorded across each of the individual national records. Furthermore, through the collaboration with RCAHMW a new approach to the format of description fields has been developed to ensure that at a minimum this field holds the essential data for each entity.

In order to better understand the impact of the work we are doing, and to help identify future actions for enabling a national collection to be developed, we have been charting the steps that researchers take to link information to create new knowledge. For example, we have followed the process for shipwreck identification, from seabed anomaly, through diagnostic features and then links to different archives. Key within this is helping to identify when researchers need to shift from Findable, Accessible, Interoperable and Reusable (FAIR) digital data, to online (but not FAIR) data and/or physical archives. In essence, this is picking out the 'air gap', or point at which online search and discovery capability hits the analogue wall. Understanding this barrier is providing knowledge key to the future of Towards a National Collection. For example, the identification of resources which could easily become FAIR, but currently are not, and the consideration of the physical archives for which digitisation would yield the greatest return. Work with Lloyds Register on this theme has begun, exploring how records within the UNPATH triplestore could be linked.

Related to the work on linking UNPATH and the English inventory (Mariner), we are testing concepts of connectivity between collections and marine inventories, including a 'Spatial Slider' capability: linking historical maps to modern charts (qv the British Library Georeferencer project) and a 'Temporal Slider' capability to test access to time series (e.g. journeys of vessels) in map form, to see whether these capabilities can be linked to national inventories.

WP3.1. To refine our AI searches and portal development and FAIR evaluation of the project, the *People and the Sea* research focus has identified target sample wreck data provided by Maritime Archaeology Trust for exploring the relationship (and challenges) of the barrier between physical (non-digital) and digital collections. We are calling this the Analogue-Digital Connector. Located off the Needles, Isle of Wight, the sample wrecks have a fairly wide date range and intriguing histories, and underline the current inaccuracy of digital databases. They include Dutch East India vessel *Campen*, lost 1627; *HMS Looe*, 5th rate, 1705; *HMS Incendiary*, fire ship, 1780; *Queen Charlotte*, troop ship, 1815; *Anglo Saxon*, Channel Island brig, 1879; *Serrana*, steamship, 1918; *SS Varvassi* cargo ship, 1947. To better understand the barriers between digital and analogue, and devise potential solutions, we will survey and subsequently interview cross-disciplinary researchers who manage selected maritime collections (Portsmouth Museum Service collections, University of Newcastle Marine Technology Special Collection, Chris Donnithorne's Navy List, Volunteers of Portsmouth Royal Dockyard Historical Trust, University of Portsmouth Map Collection and the Gosport Diving Museum).

To support development of immersives (see WP4 below), a student-led hackathon at Portsmouth piloted an immersive competition in 2022. Two teams of five postgraduate students explored visualisation of genetic data from the Mary Rose, including from the Archer Royal skeleton, to create a visual immersive/interactive. They also looked at other datasets (one team looked at ancestry.com's DNA service display data). The brief (which was also student-led) was to design and build an interactive innovative interface that connects scientific and heritage datasets to wider audiences. The challenge was to take mtDNA sequence data that is scientifically abstruse and present it engagingly to a non-specialist audience. The two outputs showed that

negotiating cross-disciplinary boundaries can inspire new ideas, insights and conversations for new audiences. We have developed digital datasets of the *Holland 5* submarine wreck and will complete digital scans at National Museum of the Royal Navy (NMRN) Submarine Museum of the sister ship *Holland 1* to enable students at University of Portsmouth for a second planned co-creation competition.

The CITiZAN (Coastal and Intertidal Zone Archaeological Network) App has uploaded records to ADS for 3,878 features around the UK worked on by its volunteers, with 5,760 images. We have tied 37 CITiZAN records to UNPATH sites: two for the Hollands, 27 for Needles wrecks and eight for the *Mary Rose*. The study area is currently a 1,122 square mile rectangle encompassing the Isle of Wight, the River Test as far as the Redbridge Causeway and the 'Solent' between Hengistbury Head and Selsey. There are also three outlier sites, one off the Lizard, one inside the port of Dover and one at Barrow-in-Furness (where the Hollands were built).

Work on access and visualisation has resulted in the Needles immersive viewer (developed by the Maritime Archaeology Trust), which has developed multiple routes to access 75 shipwrecks, crashed aircraft and a number of maritime focused terrestrial sites which lie within a 2.5km radius of the Needles. This has included inventory information, photographs, video, audio, 3D models and documents. Some additional work on the database has been undertaken to extend the types of data held, such as the ability to add and search on people associated with wrecks, to add additional options for research as part of UNPATH. An important aspect for the project was the ability for the database to link out to other digital archive sources related to each site. A complementary project being undertaken by MAT - Fathoming the Future – involving digitisation of MAT archive records from the last 30 years, has enhanced the available collections. The online viewer is a web interface in HTML using JavaScript, CSS and PHP. The web interface presents a range of ways for visitors to access the data with options to follow wreck stories, people stories, artefact stories or choose the type of resources one wishes to view. It is also possible to filter the sites and data to start building specific research questions, which will enable new stories to be developed. The Needles viewer will be available online and in Shipwreck Centre from its Easter 2023 opening and will be taken around the Solent area on the MAT Discovery bus during the summer and autumn. Reactions to it will be tested through interactive audience surveys.

A further pilot 'Archaeology on the Seabed - diversifying our engagement with historic shipwrecks', is a regional pilot study based upon a geographical area encompassing the Solent and Isle of Wight and possibly an area around the *Holland 5* historic submarine wreck, that seeks to develop new people-focused heritage narratives about the historic wrecks lying within it. Themes will include slavery and bonded labour, colonial links, the global origins of non-British seamen and migration. Keyword searches backed by further research are identifying wrecks and losses associated with these themes. These searches have identified the predictable, for example the association between the wreck of *HMS Assurance* and a significant figure in the colonial history of Jamaica. They have also identified the less predictable, for example the Bermudan connection between *HMS Mistletoe* and enslaved labour. In addition, they have pointed to gaps, for example the lack of results when keywords referring to the world's great religions are searched. AI searches on UNPATH linked datasets will be used to see if additional sites or thematic links are generated. The 'human' harvested data will also be used to train the AI.

WP3.2: Our *Science and the Sea* (WP3.2) lens has completed sample selection of unidentified wrecks in an area west of the Isle of Man, and correctly formatted related multibeam datasets for analysis. We are now identifying relevant collections and archive sources likely to aid in their identification. For the environmental work, development of a scientific metadata catalogue is underway, as is evaluation of scientific data topics

and outputs (e.g. hydrodynamical model outputs and relevance/value of specific parameters such as spring/neap variance, flow depths and tidal direction).

Assessment of relevant available multibeam data (extent and quality), following the generation of 'cleaned' point clouds and the production of digital terrain models (DTMs) for each of the 80 sites surveyed off the west coast of the Isle of Man has been completed and these data have been cross-referenced with the UK Hydrographic Office database of 'Wrecks and Obstructions' in Arc GIS to determine adequacy of data coverage in the principal case-study area of interest. To identity wrecks to a high level of confidence, 100% coverage of all known wreck sites (with multibeam) in an area of interest is essential. Given the data set was collected opportunistically with no specific research framework/objective in place, this is not the case. This initial exercise highlighted numerous data gaps within our preferred initial study area (a complex polygonal area extending over many 1,000s km2) with over 32% of charted wreck sites having not been surveyed or having minimal data coverage. We therefore adjusted the primary study area to take the form of a circular region extending 10Nm in diameter from a point located halfway between the Isle of Man and Northern Ireland. This principal study area, surrounded by two successive 5Nm 'buffer zones', has resulted in 91% data coverage of all wreck sites within the main area of interest and approximately similar levels of coverage in the surrounding 'buffer zones'. Ongoing principal activities (using the historic record, existing UNPATH resources and via collaboration across the UNPATH work packages) are now focused on identifying all potential losses associated with the revised study area. This presents a challenge to the team but also provides a significant opportunity, where project resources and ambitions will be developed through close collaboration with WP2 (particularly with respect to AI/machine learning).

Research on 'preservation modelling' is also underway through assessing the vulnerability and preservation potential of wreck sites based on a range of scientific data and commercial activity (such as bathymetry, hydrodynamics, sediments, temperature, salinity, fishing intensity, cabling and dredging activities). Significant contributions have been made by our Ulster partners and the intention is to produce 'risk-level' values for sites within the study area and the wider Irish Sea region.

WP3.3: For *Lands Beneath the Sea* (WP3.3) we have integrated the necessary datasets and collections to permit the development of the primary simulation infrastructure. The first version of an integrated terrain model for the southern North Sea has been completed. This will be developed as further data sources are analysed and combined with our existing data. This model has been used as the base for the initial infrastructure of the accessible simulation, and contains the framework required to accommodate all the functionality of the final model. It currently contains the terrain, along with user-friendly controls and a combination of external Glacio-Isostatic Adjustment models of sea level change and shorter-term sea-level processes. Simulated plant communities are present in template form, into which real world plant data can be input. The technical infrastructure has been developed using the Unity cross-platform game engine to enable compatibility with computers, smartphones and tablets, and allows user interaction with a changing, interactive environment. Integration with the outputs of the visualisation work (WP4 below) have been planned to be flexible, depending on the future development of both work packages. Co-design activities scheduled for late 2023 and 2024 have been discussed. The unique nature of the work so far has attracted collaborative research interest from the Netherlands (Groningen) and a tentative approach from a major international cruise group for further development.

WP4: There has been significant progress in developing the co-design process and researching key technical issues in the UNPATH Navigator – our approach to visualising the connected collections. The initial Methodology Report has been finalised. This remains a living document and dependent on feedback from

consultations with VIP organisations, who are advising on best practices for engaging with visually impaired users. Initial co-design sessions with visually impaired organisations have now been finalised. Glasgow Disability Alliance, VICTAR, VocalEyes have all confirmed attendance and provided staff members and additional consultants. These initial sessions will co-design the procedures and approaches required for VIP participants in the 'Navigator'. The logistics of these sessions are in advanced planning, and best practice is being developed for recording group sessions online and face-to-face sessions, to ensure consistent approaches throughout the project. Recruitment for other test groups (non-coastal communities and cross-disciplinary maritime researchers) is now underway. Sessions with these test groups (alongside VIP audiences) will begin in June 2023. The Audience Evaluation documentation is in development, and we are working on updating our 'digital confidence' survey. The Navigator Technical Workplan document for UNPATH Navigator has been completed in draft. This document includes a provisional outline of the expected behaviour and technical requirement of the Navigator. It delineates the major design challenges already identified and outlines the development timetable which aligns with the overall project timetable. Ingest of sample UNPATH metadata, per the Technical Workplan, has begun, drawn from Archaeology Data Service/ARIADNE triplestore format.

Research for the UNPATH Navigator has identified a number of key design questions, to be addressed through the co-design process, as well as the development of a novel approach for documenting the research/development process using videography to create 'explainers' (this may also include use/usability 'explainers' as well as research development) as part of the core Navigator dataset. Navigator scoping research has also highlighted several core technical challenges, with suggested mitigation, as well challenges around dataset content and structure that are not directly addressable via the Navigator, but which the Navigator can signpost for future development. The key design challenges are the geo-location of aggregated data points (ADPs): many of them do not have precise location data but must be presented in a visual interface along with those that do. Furthermore, all ADPs with coordinate data need translation to a local virtual space grid but how depth data will be handled is yet to be confirmed.

A key consideration of the richness of the data that is available to be drawn into the immersive (beyond the ADPs, which are essentially enhanced metadata) is that it is still not clear that beyond 'landing page' URLs, what content is currently addressable via DOI/URI and whether this can be programmatically associated with an ADP (addressing the knowledge gap has been flagged as a requisite in the technical workplan). However, this can be mitigated by the manual assignment of links – especially if there is a focus on enhanced content for 'Curated Pathways'. There are also system design issues around rich external content and formatting, especially for 3D data (e.g. OBJ vs FBX which is native to UNITY), however, we believe these can be mitigated either at the data provider end (preferred) or through the coding of a 'loaders' for these data. Whilst not yet confirmed, it is likely that the VIP user group will present the greatest challenge here with regards the formatting of external datasets – the Navigator is designed to allow users to navigate to data, however, if this is not in a format that is usable for VIP users then this would be problematic. Current mitigation for this approach involves investigating web usability tools in the context of UNITY 'webview' functionality.

WP5: Work in WP5 has centred around five major activities, including extensive research, workshopping and survey to produce a [Values Framework](#) for the project. An internal report, as well as an associated plain-language description of the methodology, have been developed, the latter published on the UNPATH website. The framework consists of six values monitored according to 14 measures, with at least one indicator of success per measure. While values and associated accountability measures are a known tools in the design and Human-Computer Interactions fields, robust and transparent processes of developing and

evaluating them are rare. Even in the cultural and museums sectors, where values are commonplace within institutions, open, accessible and rigorous reflections on their usefulness are uncommon, and for the most part are uninformed by current scholarly research into values for design. Our approach in UNPATH is informed by the ethics-oriented efforts of the Transforming Collections TaNC project, which complements but does not duplicate UNPATH. Indeed, in UNPATH, we have approached our values with prefiguration in mind (i.e., shaping people's behaviours before research begins) and hence have assumed a more top-down process, whereas Transforming Collections has assumed a more organic, bottom-up process driven especially by early career researchers. Our complementary but different approaches will bring us together in 2023 to co-host a cross-project TaNC ethics workshop, exploring requirements for ethical infrastructural practice.

Our values and their measures are the basis for UNPATH's overarching Evaluation Strategy, which has also been the subject of research and development over the past six months. Each measure is linked to one or more of a total of nine evaluation methods, centred primarily around ethnography, survey and interviews, co-design, self-reflection, and internal auditing. UNPATH's massive consortium and varied public engagement activities, which span dozens of institutions and audiences around the UK, have necessitated a tight approach which enables some flexibility but which ultimately collects comparable data from more than a dozen different audience engagement activities and outputs. An internal audit of existing evaluation tools has been combined with a review of immersive (and other) exhibition evaluation methods to produce a common core set of evaluation questions. These core questions form the first part of a multi-part Values/Evaluation Toolkit, a living resource which will be published internally in late Spring 2023. We believe we are unique in designing our evaluation strategy around our values, and UNPATH will eventually be able to testify empirically to the utility of values in realising (or not) wider research project goals.

To ensure that our work caters to specific audiences, on their own terms, in ways that resonate directly with them, we needed to understand these audiences better. We conducted an extensive Audience Mapping exercise collating 60 different audience-related datasets (including interviews, ethnography, web analytics, observations, and other forms of reporting), from 21 organisations, and assessing these in relation to our interests in reaching non-coastal communities, cross-disciplinary researchers and visually-impaired people (VIP). Our audience mapping report offers 22 insights and six key recommendations for reaching UNPATH's audiences and addressing key gaps in practice, including the current near-total lack of engagement with visually impaired audiences in maritime heritage. This report led us to establish contacts with key gateway organisations and consultants in the VIP community, and through these contacts we are now working with WP4 to coordinate, host and evaluate co-design activities.

Project Outputs (to date)

The following project outputs have been created as of end March 2023.

WP1

- UNPATH ontology (UO-CAT) v1.1 finalised
- Metadata licence drafted
- Provisional Perio.do definitions for Scotland and Ireland published
- Getty Art and Architecture Thesaurus for Marine Craft established
- Triple-store set up and Graph DB interface available for public usage
- UNPATH portal alpha version made public for user testing: https://unpathd.ads.ac.uk/
- UK maritime data showcased in international ARIADNE RI portal: https://portal.ariadne-infrastructure.eu/search?q=&ariadneSubject=Maritime

WP2

- Parsing of .json file from Coflein for initial processing
- Integration of dictionary into Natural Language tool kits
- Scripts (Jupyter notebooks) for FSL and ZSL approaches
- Clean datasets for working with NLP and FS/ZS approaches
- Enhanced versions of Coflein and Isle of Mann datasets
- GIS of data received to date from national services
- Conference Paper for Computer Applications in Archaeology 2023 conference
- Journal Article draft: Modest Doubt: Enabling Discovery Across Maritime Heritage Records
- 3D model of Holland 1 Submarine
  MSc Thesis: *Natural Language Processing in Maritime Archaeology* by Shrikriti Singh

WP3.1

- Public Engagement & Co-creation Ethics Application agreed
- Cultural mapping observations & co-creation workshops Ethics Application agreed
- Holland 1 submarine scans incorporated into project
- Photogrammetry surveys of Holland 5 (2017 & 2022) incorporated into project
- Two visualisation products from the student hackathons

WP3.2

- Assessment of data coverage and revision of study area to ensure >90% multibeam coverage for sites of interest.
- Initial development of models for integration of scientific data with marine activity information to determine preservation potential and risk levels for wreck sites.

WP3.3

- Extracted 2D seismic layer for integration into simulation.
- Base GIS created.
- Prototype surface created.
- Corpus of reports relating to submerged landscapes created for AI investigation by other WPs
- Simulation infrastructure developed for portable, functional terrain viewer

WP4

- Co-Design Methodology (final stages)
- UNPATH Glossary (continually updated)
- Technical Workplan for UNPATH Navigator development

WP5

- 1st release of Co-designed UNPATH Living Values
- Online update on methodology behind the creation of Living Values
- Co-designed UNPATH Living Values (1st release)
- [Online update](#) on methodology behind the creation of Living Values (published 27 June 2022)
- [Online video](#) on Living Values scholarship and intent (published 17 March 2022)
- Audience Mapping methodology
- Values Measures methodology
- Audience Mapping report published for consortium
- 1st and 2nd Drafts of Values Measures, presented to Management Committee

WP6

- Publication of UNPATH project [website.](#)
- Test Storymaps from national inventory and archive data
- Test Mapsliders from national inventory and archive data
- Test 3D modelling from national inventory and archive data

Cross Project Collaboration

In the set-up stages of the project, we have had little opportunity to begin to work with the four other TaNC projects in any great depth outside of the strong inter-project collaborations initiated by the Programme Director and her team. These TaNC sponsored events (all-projects meetings, workshops on technical aspects, project management approaches, decolonisation of research, and media interaction training) have provided fertile grounds for identifying some key cross-project collaboration strands. One example of strong cross-project collaboration which has emerged from UNPATH is the work undertaken as part of WP5 on project values, as mentioned above. At a cross-TaNC workshop, enthusiasm was shown from all the other projects on understanding what we were trying to accomplish with work on Values and how this might tangibly affect TaNC project engagement with different audiences. Our WP5 team is now coordinating, with researchers from Transforming Collections and The Congruence Engine, a half-day, cross-project workshop on ethics to be hosted at Tate Britain in Autumn 2023.

Another of UNPATH's key strengths – the strong spatial relationships embedded in most of the collections with which we are working – provided a similar opportunity for cross-collaboration through discussions at the Towards a National Collection programme workshop at Royal Museums Greenwich in December 2022.

Sustainability and Infrastructure

Sustainability of the UNPATH project products and outputs will be ensured by adhering to FAIR principles and using international data standards.

In the short term (i.e. during the duration of the project), university partners use networked computers which are regularly backed up to secure storage, using a mix of tape, remote hard drive and AWS or Google cloud provision. Project management information and shared documents will be held on shared team drives, or cloud services, securely managed and backed up by the respective IT services and to which only authorised staff have access. Non-HEI partners all follow industry-standard backup and data management procedures, and use networked drives and computers, and either designated data centres, or third-party servers, operated by Microsoft and Oracle and located in Western and Northern Europe, backed up daily. For new primary data transferred to ADS, core data in the form of AIPs and original (the slightly altered SIP) files are regularly synchronised from a local copy in the University of York to AWS.

Long-term preservation of new digital data generated or enhanced within UNPATH will be the responsibility of the archivists within the national heritage bodies, supported by ADS for data that would otherwise be orphaned, and specialist data types. All follow the OAIS model and ADS and RCAHMW are able to assign Datacite DOIs to all data sets and reports but will also observe recommendations of the TaNC Persistent Identifiers Foundation Project. In ADS spatial data is published using Geoserver and incorporated into web interfaces using Openlayers (v5). The Highslide (javascript) application is currently used to provide zoomable access to high resolution JPG images in the archive, but ADS & RCAHMW are investigating the deployment of IIIF and will build upon the recommendations of the TaNC IIIF Foundation Project.

ADS archives are preserved into perpetuity; additional data held by university partners and not transferred to ADS will be held for a minimum of 10 years by the institutional repositories. Beyond the project lifetime, data which forms the UK marine record will continue to be updated by the national bodies for their areas of responsibility as part of their national roles. For England, the National Marine Heritage Record is currently held within Warden, a bespoke database developed on the Getty Arches platform; in Scotland, Canmore is Oracle-based; in Wales Arches provides the front-end and Preservica is used to ensure long-term preservation and access; in Northern Ireland data is held on a SQL Server with public access provided by an ESRI web application.

ADS and the national bodies will maintain access to the distributed marine record and associated digital archives using their respective current and enhanced search interfaces: ArchSearch, Heritage Gateway, Mariner, Canmore, Coflein etc. For external portals we are reliant on their sustainability plans, but MEDIN's marine portal is a national service, and there is a commitment by the Swedish National Data Service and CNR in Italy to maintain the ARIADNE until 2025, with plans being developed for longer term support as part of the ARIADNE Research Infrastructure, a not-for-profit company registered under Belgian law. Similarly, the ADS will maintain the UNPATH iteration of the portal for at least 2 years beyond the project end. All data held by ADS is available online free of charge, and under a CC-BY or other open licence, and metadata is CC-0.

The TANC Github site will become a repository for all source code generated in WP2-WP4 and made available CC-BY-NC. WP3.1 Holland submarine visualisations will be maintained on web platforms by the University of Portsmouth, and/or Nautical Archaeology Society for five years after the project. For WP3.3 the software simulation source code will also be freely available via the CoMSES simulation repository. The software will be available free as an installer for desktop PCs via Itch.io. While the installer is vulnerable to future operating

systems changes and may not function reliably in the long term, availability of the source code will allow the software to be altered and recompiled to ensure compatibility with any such changes. There will be no dependencies on commercial libraries to facilitate the recompilation. For WP4, the Navigator project files will be maintained on Github and the Glasgow School of ARt but will be vulnerable to changes in dissemination technology. Versions created during the project for particular events e.g. standalone, web, mobile etc, will be archived for re-use by SimVis at GSA. Historic England will maintain a dedicated project website, with links to all outputs, for 10 years after the project end date.

Interim Recommendations

Our interim recommendations are based around the challenges we are encountering as the project proceeds.

DISCOVERING COLLECTIONS – At an early stage in the project it became clear that our scoping needed to be clearer about how far we would explore the integration and connectivity of metadata and of data themselves. We are deliberately targeting both. Metadata, obviously common to all collections, are perhaps best exemplified by the summary level inventory data contained within the five (including Manx Heritage) national records of the historic environment. Data related to these inventories exist in numerous forms, including born-digital (such as sonar surveys of wrecks), digitised (such as specific data relating to charts, objects, and investigation reports) and purely analogue/physical (collections which do not yet have any meaningful digital footprint). Our recommendation is that a strategic decision is made early on about how to approach this complexity. Such a strategy should be based on an evidence-based characterisation of likely Return on Investment (ROI). An incremental long-term investment strategy of this nature should in this way maximise engagement with the collections. Establishing the criteria for ROI will be essential, as these will be more complex than number of 'units' and may well include consideration of accessibility to wider and more diverse audiences, response to current challenges and issues and, potentially addressing current in-built biases.

ANALOGUE TO DIGITAL – Connecting inventories to deeper datasets of collections rapidly brought us to the Analogue-Digital 'air-gap'. It became clear that the analogue collections which are currently beyond machine access are often the very resources which will unleash the full potential of the national collection. At the half-way point of the project we have witnessed this in relation to construction and loss records for vessels now existing as wrecks, for crew and cargo manifests, and for artefacts, images and maps and charts. The barriers to digitisation are numerous and complex, including legal, ethical, physical, technological, and economic issues. Furthermore, those resources where we have been able to 'test' newly digitised collections show up a variability in consistency and thus approach to FAIRness. We recommend that the way in which digitisation is approached is supported by framework guidance and principles to maximise FAIRness.

AUTHORITY AND CONTROL ON VOCABULARIES – The creation of our Uo.Cat ontology and the expansion of terms for wrecks to help enrich the Getty Art and Archaeology Thesaurus has already indicated to us that a national collections strategy will need to consider the degree to which controlled vocabularies are a feature. Our cross TaNC discussions have shown that authority and control need to be managed very carefully. Some collections must have a level of authority which can be depended upon for (e.g.) planning or environmental requirements, but that is by no means universal. The extent to which such controls should be international or global schemas is also not clear, and panels such as those managing biological classifications might be needed.

ARTIFICIAL INTELLIGENCE – As AI tools rapidly develop, and increasingly have the potential not only for surfacing but creating information (generative AI), the need for tracking of how outputs and labels have been created has become clear. This is a particularly acute issue within this project, where these datasets (and the collections they represent) are used as the basis for national strategy, cultural heritage management as well as planning and development decisions. It is also clear that much of the work being conducted as part of this

project is about enabling linking of data which for historical reasons is not always well formatted for contemporary needs. Once datasets are enhanced, interoperable and discoverable, small changes in data collection and management for future records/enhancement will be needed. This is about making data fit for 21st century needs.

DYNAMIC DATA – Many collections are static, representing a particular point in time assemblage created by single individuals or entities, for example the collection of artefacts from a since-destroyed wreck, the works of art by a specific historical artist, the dive logs of a particular individual. However, many more are deliberately envisaged to be dynamic, adding, modifying and, potentially, deleting components of the collection. Examples of these are national inventories of marine heritage which include 'monument' entities, 'event' entities and 'reference' entities. Systems and flowlines must be developed which ensure that links between collections are 'live' and able to connect with dynamic collections as those collections grow. Only in this way will we be able to counter the risk of 'snapshot research'.

CONNECTIVITY – Two gaps have been identified which are likely issues for the UNPATH Navigator, and thus possibly for other solutions. First, rich content is not always directly addressable from the metadata – i.e. there are URL links to HTML landing pages, but the content on these pages is not (often) discoverable programmatically. An example might be a text, video or image associated with a wreck held by a data provider and visible via their wreck record web page – but not directly connected to the metadata record itself, or necessarily logically/consistently addressed from the web page. Second, a question of what a data navigation tool can and cannot address – the Navigator can explore interesting and meaningful ways for the visually impaired to find a metadata record, but if the richer data that this record points too is not formatted in an accessible way then an issue still exists as far as the user is concerned.

ETHICS – We will need to foreground human and more-than-human centred design from the outset in developing and refining both our collections and the infrastructures that house them. The majority of these (if not all of them) have evolved over the past 100+ years without such attention to humane and environmentally-sensitive practice. As a result, all aspects of our work today (from vocabulary development, to AI, to discoverability and engagement) are challenging, having originally been developed with niche specialist expertise and competencies at their core. Future activities must prioritise pre-design, as we're doing with VIP audiences, of all outputs and of methods themselves with more expansive audiences and needs in mind.

INHERENT ISSUES IN THE COLLECTIONS – We are also aware of issues that may emerge with regards to presumption around audiences and bias both within datasets and search terms. Within any historical record there is a risk that enhanced search will help surface outdated information and outmoded ways of thinking. The data examined so far has not revealed negative outcomes in terms of perspective, but has revealed missing terms and labels that people might want to search for. For example, initial exploration around terms such as 'slavery' reveal few associated records, but this does not equate to little of the archaeological record having links to the slave trade in the past.

Contacts

Barney Sloane: barney.sloane@historicengland.org.uk for general overview of project research

Julian Richards: julian.richards@york.ac.uk for technical data information

Fraser Sturt: F.Sturt@soton.ac.uk for AI information

Stuart Jeffery: S.Jeffrey@gsa.ac.uk for simulation and visualisation information

Sara Perry: sperry@mola.org.uk for audience research information

Annex 1

Data Aggregation workflow in UNPATH

The flowline of national inventory data from agencies to the project triple store, aggregation and then distribution to portals. Data is collected in XML format (or harvested via OAI-PMH) and the 3M mapping tool is then used to convert the data to RDF according to the project's ontology (UO-Cat, which is a subset of the CIDOC-CRM). In a second pipeline agency-specific subject vocabularies are mapped to the Getty Art and Architecture Thesaurus, while in the third input pipeline the PeriodO web application is deployed to provide absolute date ranges for the equivalent agency-specific period terms. The three workflows merge to form RDF in the GraphDB LoD triple store, with additional metadata enrichment from AI where required. The triple store is then used to generate Open Search indices which provide the basis of search and retrieve in existing portals, including ARIADNE, as well as the UK maritime instance of the ARIADNE portal created specifically for Unpath'd Waters.
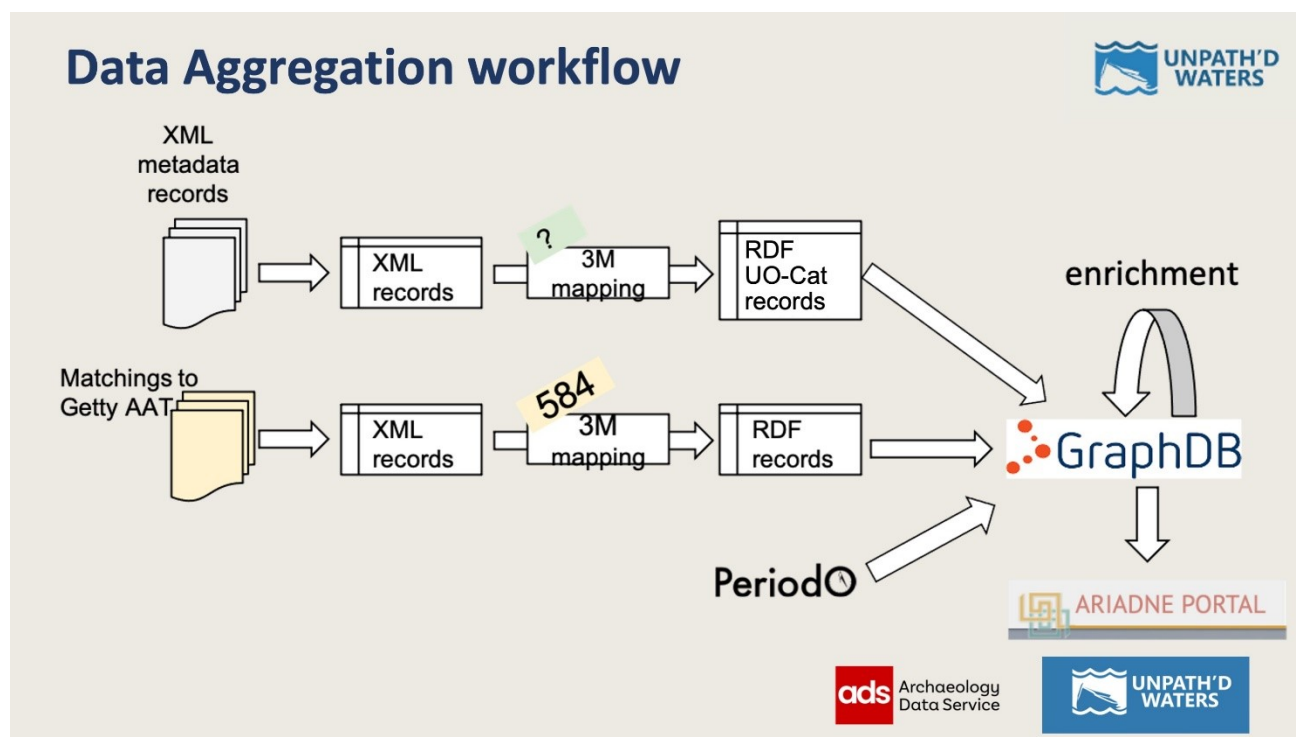
Zero Shot Learning in UNPATH

ZSL is a variant of transfer learning, enabling recognition of patterns without specified training examples. It learns from past knowledge. It is an important tool since marine archaeology as a discipline is currently in a

digitization phase, as a result of which established training datasets are not readily available and new categories of object/data will keep coming up. ZSL has been tested on the Welsh Coflein (national inventory) data set and has showed approx. 80% accuracy with expert-based rules. The approach is benchmarked against powerful supervised methods like Support Vector machine and Naïve Bayes.