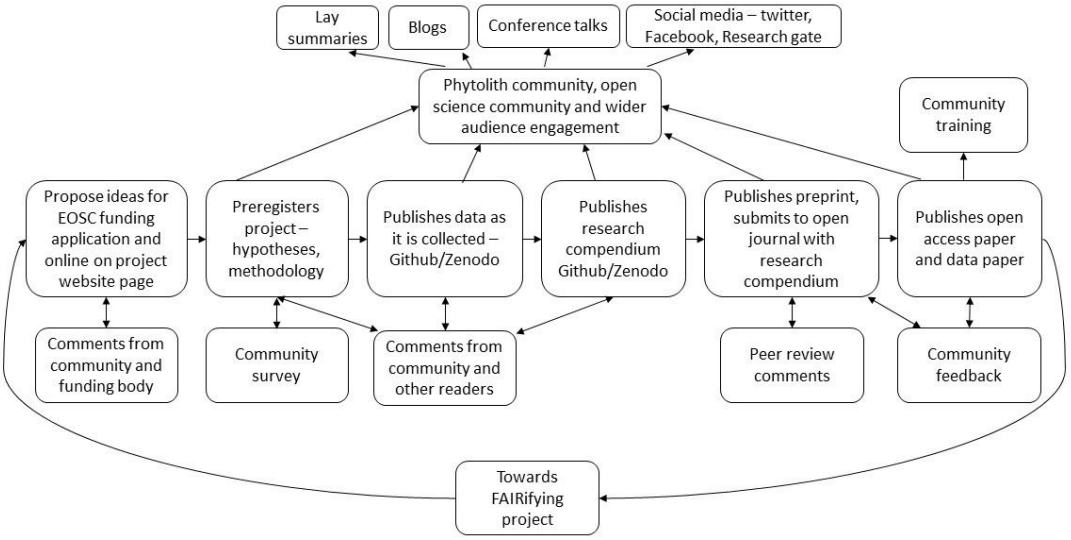
Additional files for Improving the FAIRness of Phytolith Data

Documents include:

1. Workflow diagram for project.
2. Table 1: Moving towards FAIR phytolith data.
3. Data Management plan.
4. Letter of support from Open Life Science.
5. Letter of support from the International Phytolith Society.
6. Letter of support from Historic England.

1. Workflow diagram for project



2. **Table 1: Moving towards FAIR phytolith data**

Actions	Findable	Accessible	Interoperable	Reusable
Easy wins				
Datasets from published articles in an open repository with DOI	X	X		X
Dataset deposited as csv file			X	X
Research articles - gold or green open access	X	X		
Data availability statement in research article	X	X		
Add a CC0 license to dataset				X
Stating what nomenclature used in published article				X
Stating clear method in research article/referencing one methodological paper				X
Medium effort				
Datasets of raw data in an open repository with DOI	X	X		X
Writing a data paper for the dataset	X	X		
Well described metadata in repository				X
Providing pictures/photos of all morphotypes in repository				X
Full protocol in repository			X	X
Use standard geographical terminology			X	X
Harder to implement				
Using standardised processing method			X	X
Using standardised			X	X

nomenclature				
Using standardised dating terminology			X	X
Research compendia with research articles including R stats code for data analysis			X	X

Easy wins = simple adjustments to usual practices by individual researchers.

Medium effort = adoption of new practices by individual researchers.

Harder to implement = need community consensus for successful adoption.

3. Data Management Plan

Administrative data	
Project name	Increasing the FAIRness of phytolith data
Project description	This project aims to take the first step towards changing the culture of data sharing in phytolith research. It will initiate movement towards more standardised, FAIRer and available data by conducting a FAIR assessment of existing datasets, developing clear recommendations to change the current data sharing practices for existing and future datasets and start to introduce these recommendations to the phytolith community by modelling data sharing best practice.
Funder	EOSC-life digital life sciences open call
PI/ Researcher	Dr Emma Karoune
PI ORCID	0000-0002-6576-6053
Project data contact	Dr Emma Karoune - ekaroune@googlemail.com
Date of first version	22/12/2020
Date of last update	N/A
Related policies	Historic England uses the ADAPt toolkit for managing and preserving data. This project will follow these principles to embed the high standards of successful data creation, management, and long-term data preservation that Historic England promotes. The DMP below sets out the specific requirements of this project.
Data collection	
What data will you collect or create?	Existing phytolith datasets will be collected from published journal articles from studies concerning two geographic regions - South America and Europe. Initially, this will be from the last 5 years and then an assessment of the size of the dataset will be conducted to establish whether it is representative. If the dataset needs to be increased, another 5 years of published articles will be collected. The existing data will be quantitative. These are small datasets, usually collected in excel spreadsheets. From the existing datasets, a FAIR assessment will be conducted and this will create a new dataset in terms of the type of data and metadata included in the articles and other aspects related to the FAIR principles. New survey data will be created.
How will the data be	Existing datasets from published articles will be found through a

collected?	search engine such as google scholar or web of science. The datasets will be in a number of formats - tables in the text, pdf, word and excel files in supplementary files and excel and csv files in repositories. All existing data will be made into csv files and uploaded to a Github repository. The FAIR assessment data collection will be standardised through the formation of a google form for data entry that creates a google spreadsheet. The final spreadsheet will then be added to a Github repository as a csv file and archived on Zenodo. Survey data will be collected in a google form that creates a google spreadsheet. The final spreadsheet will then be added to a Github repository as a csv file and archived on Zenodo.
Documentation and Metadata	
What documentation and metadata will accompany the data?	All aspects of the project will be fully documented and all metadata will be compiled as a research compendium during and at the end of the project. This will be published with the research articles and data paper.
Ethics and legal compliance	
How will you manage any ethical issues?	All survey data will be anonymised. It will be kept in line with GDPR regulations by not collecting any personal data.
How will you manage copyright and Intellectual Property rights (IPR) issues?	It is the aim of the project to make all new datasets publicly accessible under a CC0 license. All documentation created will have CC-BY 4.0 licenses. However, this may not be possible for all existing datasets due to the licenses and copyright the data is published under. Authors and journals will be contacted to seek permission to publish all existing datasets on the Github repository and Zenodo as csv files under CC0 licenses. Zenodo is being used to give all research outputs persistent identifiers (DOI's) so that all parts of the project are citable.
Storage and Backup	
How will the data be stored and backed up during the research?	Locally created documentation will be stored in Historic England's network, which is backed up daily. As this is a distributed team, cloud services will be used to aid collaborative working. All new datasets will be stored during the data collection phase on a project Google drive but will then be uploaded to the Github repository as csv files. This will be done for existing datasets when permission is given by authors and journal publishers.

	The Github repository will then be archived on Zenodo.
How will you manage access and security?	The aim of the project is to be as openly accessible as possible therefore we are using free to use software and tools such as Google drive and Github. The project will be open for contributors to edit and make comments on during the project. Github and Google have version control to enable fair contributions and the ability to step back to earlier versions of documents if necessary.
Selection and Preservation	
Which data should be retained, shared and/or preserved?	All new data and documentation in this project, as well as the existing datasets that give permission, will be deposited in the Github repository and archived on Zenodo for long-term storage. This will also provide Digital object identifiers for all of the research outputs. It is the objective of this project to be fully transparent to aid reproducibility.
What is the long-term preservation plan for the dataset?	Zenodo is a long-term data repository built and operated by CERN and OpenAIRE. This means that all research outputs are sustainably archived.
Data Sharing	
How will you share the data?	This is an open project so all research outputs will be in an open Github repository that will be publicly accessible. Findability and long term accessibility will be increased by archiving on Zenodo.
Are any restrictions on data sharing required?	There will be no restrictions on any new outputs created in this project. Existing datasets will only be shared under their existing licenses and copyright agreements, unless permission is given by authors and journal publishers to make them publicly open under CC0 licenses.
Responsibilities and resources	
Who will be responsible for data management?	Dr Emma Karoune
What resources will you require to deliver your plan?	All software and tools are free and open to use.

This DMP is based on DCC. (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

4. Letter of support from Open Life Science.

18 December 2020

Dear EOSC-Life Committee members,

I am writing this letter on behalf of Open Life Science (OLS) and The Alan Turing Institute's project *The Turing Way* in support of the grant proposal by Dr Emma Karoune and her team to fund their proposal for Open Science in Phytolith Research designed under the Digital life sciences open call by EOSC-Life.

OLS is a training and mentoring program for researchers at different career- and skill-levels to learn about and apply open, reproducible and FAIR research principles in their work. We are grateful to receive the next round of funding from EOSC-Life to involve and train researchers in EOSC Research Infrastructure. *The Turing Way* is an open source community-led project in data science that involves and supports a diverse community of researchers, funders, educators, and learners to create, learn, and apply best practices in reproducible research. The programme's goals are to build trustworthy systems, embed transparent reporting practices, promote inclusive interoperable design, maintain ethical integrity, and encourage respectful co-creation.

I have had the opportunity to work with Emma in both these projects. I have seen Emma's persistence and work ethics closely while she developed *The Turing Way* chapters and resources for people new to data science. Emma has recently graduated from the second round of OLS as a leader of the project "Creating community awareness of open science practices in phytolith research". I have witnessed her passion and dedication to creating resources that highlight the importance of open science in the phytolith community. She has been spearheading the Working Group for Phytolith Open Science to explore existing frameworks and build pathways to ensure the sustainability of research components in her field. They aim to actively involve and empower individuals by training them on open and FAIR practices in an accessible and meaningful manner.

Field scientists and experimentalists from various disciplines in life science rely on computational approaches to build research workflows and publish their work. They are extensively trained in skills in their respective field of research, however, they are often not taught to effectively apply computational methods with the formal understanding of open, FAIR and reproducible practices. Online resources on FAIRification of research components, when combined with hands-on training and workshops tailored for data types used in their field, can facilitate researchers to apply best practices in their professional and academic work. The phytolith community is at the very start of the open science journey and this project will therefore make a large impact to move researchers towards producing FAIRer data.

I would like to express my support for the proposed plan to make phytolith data FAIR at source and I look forward to seeing their project succeed. I believe that by awarding this grant, EOSC can boost the project started by this working group and make a significant difference in the phytolith community.

Yours Sincerely,

Malvika Sharan

Community manager, *The Turing Way,* The Alan Turing Institute
Co-founder of Open Life Science
Email: malvikasharan@openlifesci.org

5. Letter of support from the International Phytolith Society.

International Phytolith Society

LETTER OF SUPPORT

To: Emma Karoune, ekaroune@googlemail.com
Subject: IPS Letter of Support for the Project *Increasing the FAIRness of Phytolith Data*

The IPS wishes to declare its enthusiasm and support for the project *Increasing the FAIRness of Phytolith Data*, proposed by Emma Karoune, Juan José García-Granero, Marco Madella and Carla Lancelotti.

While phytolith analysis has a long history of application in paleoecology and archaeology, there has recently been an explosion of interest in phytoliths in many scientific disciplines as researchers have found phytoliths useful to address scientific questions in fields such as ecology, phylogeny, paleoclimatology, and biogeochemistry. Unfortunately, most phytolith data published to date have only rarely been presented in a form that can be easily shared with and reused by other researchers. The IPS believes the phytolith research community would benefit greatly from having increased access to phytolith data in an easily shared form.

Over the past twenty years or so, the community of phytolith researchers represented by the IPS has developed several initiatives that have markedly improved phytolith identification and classification. The standardization of phytolith data resulting from the project proposed by Karoune *et al.*, if funded, will enable the phytolith community to move on to an exciting new phase.

IPS fully supports this project and is committed to forming an ad-hoc committee (the International Committee on Open Phytolith Science - ICOPS), chaired by Dr. Emma Karoune, to help guide the phytolith research community toward transforming data generated by phytolith research into a more useful and shareable form.

On behalf of the IPS Board,

Ákos PETŐ
President

17th December 2020.

Contact :
https://phytoliths.org/
phytolithsociety@gmail.com

6. Letter of support from Historic England.

Historic England

Our ref: Gill Campbell

Telephone 02392 856780
Fax 02392 856701

21 December 2020

Dear Sir/ Madame

Increasing the FAIRness of phytolith data

Historic England (HE) is delighted to write in support of this bid to the European Open Science Cloud (EOSC) Digital Life Science Open Call by Dr Emma Karoune.

Historic England is a Non-Departmental Public Body established by statute under the National Heritage Act 1983. We are the statutory advisers to Government on the management, protection and advancement of understanding of our historic environment whether built, buried or on the seabed. We are a Public Sector Research Establishment, and an Independent Research Organisation recognised by UKRI.

As laid out in our Corporate Plan[1], our strategic activities include: 1.) identifying and delivering strong research collaborations with Higher Education Institutions and Independent Research Organisations (IROS) that will deliver real impact benefitting all partners; 2.) building and developing sector capacity and capability to make the most of the historic environment. This collaborative project between Historic England, the Spanish National Research Council and Pompeu Fabra University, Barcelona meets these ambitions. The project will also contribute to delivery of the Heritage Information Access Strategy (HIAS)[2] which seeks to simplify and improve public access to heritage data. Improving access to palaeoecological data, including that obtained though phytolith analysis, is a key challenge for the sector as is the ability to archive datasets effectively in ways that continuously advance archaeological knowledge. This project will tackle these challenges in a highly original and integrated way.

The project will take forward the #Inform theme of Historic England Research Agenda[3] by enabling the sharing, linking and interoperability of phytolith data and research resources. It will also take forward the Human Environment topic of the Research Agenda by delivering better ways of understanding past environments through the study of these important palaeoenvironmental proxies.

Historic England will host the PI at the HE Fort Cumberland Laboratories in Portsmouth, providing her with direct access to our heritage science infrastructure and sitting her within the HE National Specialist Services Department (NSSD) with more than 100 staff spread across six locations nationally. HE will also administer the grant, distributing funds to Pompeu Fabra University for the employment of the Postdoctoral Research Assistant, Javier Ruiz-Perez.

Historic England

I trust that this letter demonstrates our interest in and commitment to Dr Karoune's application.

Yours sincerely,

Gill Campbell
Head of Environmental Studies
gill.campbell@HistoricEngland.org.uk

Cc. Jen Heathcote, Head of Investigative Science,
John Cattell, National Head of Research

[1] https://historicengland.org.uk/images-books/publications/he-corp-plan-2020-23/historic-england-corp-plan-2020-23/
[2] https://historicengland.org.uk/research/support-and-collaboration/heritage-information-access-strategy/
[3] https://historicengland.org.uk/images-books/publications/he-research-agenda/research-agenda/