Digital epigraphy in 2022: state of the art

Petra Hermankova[[1]](#footnote-20)

Marietta Horster[[2]](#footnote-23)

Jonathan Prag[[3]](#footnote-26)

2022-03-14

Table of Contents

# Introduction

The field of digital epigraphy has seen significant development in recent years: not only the traditional epigraphic corpora are continuously being digitised and made accessible via their websites for anyone to browse and search, but several resources were already born digital without their printed edition, e.g., *Inscriptions of Greek Cyrenaica* ([Roueche *et al.*, 2020](#ref-roueche2020)), *Inscriptions of Roman Tripolitania* ([Roueche, 2022](#ref-roueche2022)) or *Inscriptions of Aphrodisias* ([Bodard *et al.*, 2007](#ref-bodard_inscriptions_2007)). Most inscriptions contain references to places, people, events or contain spatio-temporal data related to the place and time of their creation and provide an ideal resource to study the past communities as a whole. However, in order to be able to harness their full potential and for example access *all* inscriptions from a place of interest or of a given type, we need to link the existing datasets together. The concept of *Linked Open Data* (LOD) provides means of connecting various digital datasets while enriching the text with broader spatio-temporal context as well as prosopographic data, leading to the creation of new connections between individual inscriptions as well as archaeological sites or potential re-evaluation of historical narratives ([Tupman, 2021](#ref-tupman2021)). Although many epigraphic datasets have been using LOD, especially to record the spatial component by using the Pleiades or Trismegistos, there is still a considerable gap in the LOD implementation across the discipline and thus the accessibility of the data.

The **FAIR Epigraphy Project** (<https://www.csad.ox.ac.uk/fair-epigraphy>) has the ambition to fill in the gap between digitisation of inscriptions and being able to use their full potential as a digital resource. The FAIR Epigraphy project has been established as a collaboration between Johannes Gutenberg University in Mainz (Prof. Marietta Horster) and Oxford University (Prof. Jonathan Prag), funded by the Arts and Humanities Research Council (AHRC) and Deutsche Forschungemeinschaft (DFG) and will run for 36 months from 2022 to 2025. The **FAIR Epigraphy** aims to create an interactive platform for all epigraphic projects, aligning their digital needs with the principles of FAIR science. The overall desirability for **FAIR** (*Findable*, *Accessible*, *Interoperable*, *Reusable*) data is fundamental to advance research into the epigraphic, linguistic, and material culture of the ancient world.

*“The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.”* (FAIR Principles website, <https://www.go-fair.org/fair-principles/>)

With the increase in Linked Open Data and novel interface technologies and standards, the FAIR Epigraphy project will be able to create the tools and the community needed to transform epigraphic research in the digital age. However, the FAIR Epigraphy project does not wish to replicate any current efforts, but align existing initiatives and bring them together to create a hub of high-quality tools and FAIR compliant standards and resources for the modern epigraphic discipline. Our internationally collaborative approach will enable and support innovative research across epigraphic data, and the wider linked web of data (especially archaeological data), such that all epigraphic data is increasingly FAIR for both the research community and the wider public. To that end, we aim to:

1. consolidate community-wide standards (vocabularies and ontology);
2. develop the tools for community implementation of those standards (vocabulary and ontology hosting and publication);
3. host and make fully accessible the resulting linked open data published by individual projects (RDF/XML data publication).

In order to map the existing field of digital epigraphy, current practices and standards, as well as find out the (digital) needs of the discipline, we have circulated the two scoping surveys in February 2022 (*FAIR Epigraphy: Scoping survey for partners and collaborators* and *Digital epigraphy in 2022: scoping survey* for all digital epigraphy projects). The results of the survey, presented in the current report, will be used to plan the activities and efficiently allocate the resources of the FAIR Epigraphy Project in the next three years. The survey answers are anonymised so the individual projects cannot be identified on the basis of their contents and the data is stored as TSV (tab-separated value) file within the same GitHub repository (<https://github.com/FAIR-epigraphy/scoping_survey_report>) as a supplement to the text of the report and can be accessed under the CC-BY-SA 4.0 International License.

# FAIR Epigraphy partner projects

This section summarises the results of the online survey *FAIR Epigraphy: Scoping survey for partners and collaborators* aimed at established digital projects that became official partners and collaborators of the FAIR Epigraphy Project. We have sent the survey to 14 partner projects. We have received 9 responses to the survey, with a response rate of 64.29%.

The partner projects represent relatively established projects with the average duration of a project shortest being 6 years. The shortest participating project reported their duration as 3 years and the longest 30 years.

## Language coverage

**Question:** *What is the predominant language of epigraphic data in your project (for mixed collections or collections where other languages are predominant provide details in Other)*

## language n ratio  
## 1 Latin 6 30  
## 2 Greek 5 25  
## 3 Other 2 10  
## 4 Ancient Celtic 1 5  
## 5 Etruscan 1 5  
## 6 Gaulish 1 5  
## 7 Hebrew 1 5  
## 8 other epichoric languages from the west provinces (ex. Africa) 1 5  
## 9 Punic 1 5  
## 10 Raetic 1 5

**Commentary:** The language coverage of the participating projects consisted predominantly of Latin and Greek projects (representing 55% of the answers). The Other category encompassed a substantial part of the surveyed projects, documenting the need to expand beyond the traditional Latin and Greek focus of the epigraphic discipline. The languages listed as Other consisted of Ancient Celtic, Etruscan, Gaulish, Hebrew, other epichoric languages from the west provinces (ex. Africa), Punic, Raetic.

## IT infrastructure

**Question**: *Does the project have a website?*

## # A tibble: 1 x 2  
## Website n  
## <chr> <int>  
## 1 Yes 9

**Commentary**: All of the participating projects maintain an online presence (as of February 2022).

**Question**: *Does you project have an IT specialist(s)?*

## # A tibble: 4 x 3  
## IT\_spec n ratio  
## <chr> <int> <dbl>  
## 1 Yes, equivalent of part-time (<1.0 FTE) position 6 66.7  
## 2 No 1 11.1  
## 3 Yes, equivalent of full-time (1.0 FTE) position 1 11.1  
## 4 Yes, equivalent of more than full-time (>1.0 FTE) position 1 11.1

**Commentary**: Vast majority of digital projects have an IT specialist, yet only 2 projects have an equivalent of 1.0 FTE or more at their disposal. The 66.67 % of projects have access to part-time IT support for their projects, which in some instances may be only a few hours per week per project.

**Question**: *Does your project store epigraphic data in the following formats…?*

## # A tibble: 6 x 3  
## format n no\_format  
## <chr> <int> <dbl>  
## 1 Epidoc XML 4 1  
## 2 Epidoc XML, JSON, CSV 1 3  
## 3 Epidoc XML, JSON, RDF 1 3  
## 4 Epidoc XML, SQL or similar 1 2  
## 5 RDF, SQL or similar 1 2  
## 6 SQL or similar 1 1

**Commentary**: The majority of projects use Epidoc XML as their main output data format (representing 77.78% of participating projects). 22.22% of projects use only one type of data format, while 44.44% two or more data format types (such as Epidoc XML, JSON, CSV, RDF, SQL or similar).

## Data sharing

**Question**: Do you share your data outside of your project?

## # A tibble: 4 x 3  
## share n ratio  
## <chr> <int> <dbl>  
## 1 Yes, under a Creative Commons license 6 66.7  
## 2 Not currently, but we are thinking about it 1 11.1  
## 3 Yes, on demand 1 11.1  
## 4 Yes, without any license 1 11.1

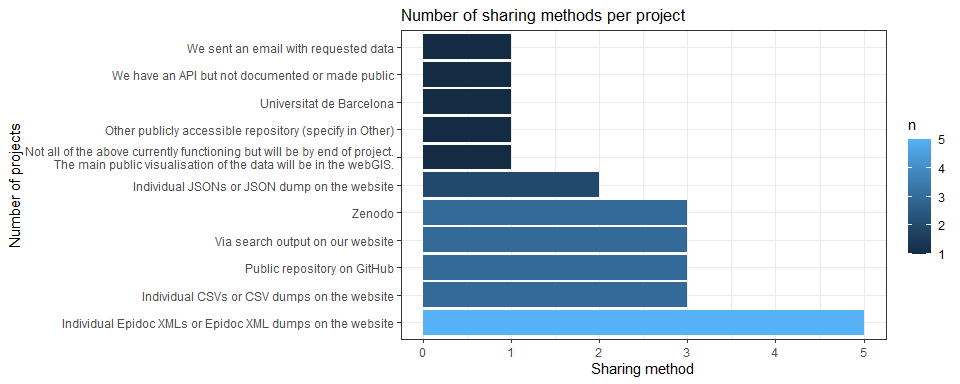
**Commentary**: The majority of partner projects share their data: 66.67% shares the data under a Creative Common license, which is the preferred mode according to the FAIR data principles. All partner projects reported their willingness to share the data, even if they are not currently doing it, or they provide the data only on demand.

**Question**: *How do share your data with users outside your project?*

## # A tibble: 9 x 3  
## share\_all n share\_method  
## <chr> <int> <dbl>  
## 1 Individual Epidoc XMLs or Epidoc XML dumps on the website 1 1  
## 2 Individual Epidoc XMLs or Epidoc XML dumps on the website;~ 1 2  
## 3 Other publicly accessible repository (specify in Other); U~ 1 2  
## 4 Public repository on GitHub 1 1  
## 5 Public repository on GitHub; Individual Epidoc XMLs or Epi~ 1 3  
## 6 Public repository on GitHub; Zenodo; Individual JSONs or J~ 1 7  
## 7 Via search output on our website; We sent an email with re~ 1 2  
## 8 Zenodo; Individual CSVs or CSV dumps on the website 1 2  
## 9 Zenodo; Individual JSONs or JSON dump on the website; Indi~ 1 4

**Commentary**: All partner projects provide at least one way of sharing the data (whether it may be currently accessible to the public or not, or it is intended to be accessible in the future). The average (median) number of sharing methods per project is 2.

**Question**: *What is the number of sharing methods across projects?*



**Commentary**: Epidoc XML is by far the most popular format for data sharing, however other Open Science services are starting to make their way to established digital epigraphy projects, such as sharing via public repository on GitHub or Zenodo, as well as providing raw data in the CSV (comma-separated value) format, or JSON (JavaScript Object Notation) files. Only a relative minority of participating partner projects shares the data on an on-demand basis or have a non-public API access point to their data.

## Institutional policies

**Question:** *Does your institution of funding body require your project to comply with any data policies (e.g., FAIR principles, data storage, data sharing, Open Science)?*

## # A tibble: 3 x 3  
## policies n ratio  
## <chr> <int> <dbl>  
## 1 Yes 5 55.6  
## 2 No 3 33.3  
## 3 The ERC open data policies don't apply to this project, but we ar~ 1 11.1

**Commentary**: The majority of projects (represented by 55.56%) are required to comply with data-related policy introduced either by their institution or a funding body. Only a minority of partner projects (33.33%) are not required to follow any data policy, nor follow it on a voluntary basis.

**Question**: *If you have answered YES in the previous question, please specify what are the policies, or provide a link.*

## [1] "The funding body (the ERC) expects that research results are available in open access"   
## [2] "Open access publication."   
## [3] "Actualització de la Política d’Accés Obert a la Universitat de Barcelona (http://hdl.handle.net/2445/142065)"  
## [4] "Data Sharing, Open Science"   
## [5] "usual ERC requirements"

**Commentary**: Several of the partner projects follow the ERC data and open access policies. More information on ERC Open Research Data and Data Management Plans can be found at <https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf> or at ERC Open Science policies page <https://erc.europa.eu/managing-your-project/open-science>.

## Open Science practice

**Question**: *Standardized terminologies: The project uses the following systems:*

## standard\_terminologies  
## 1 Own version of EAGLE vocabularies (edited for our project)  
## 2 Internal authority lists  
## 3 EAGLE vocabularies as provided at https://www.eagle-network.eu/resources/vocabularies/  
## n ratio  
## 1 6 66.67  
## 2 5 55.56  
## 3 4 44.44

**Commentary**: The lists of vocabularies for the epigraphic discipline created by the EAGLE project (<https://www.eagle-network.eu/resources/vocabularies/>) are used by most of the projects: either in the original form (representing 44.44% of participating projects) or in the form modified for the needs of the project (representing 66.67% of participating projects). The need for modifications suggests that the EAGLE vocabularies do not in fact form a community-wide standard and need to be improved before becoming one. The process has been already started by the Epigraphy.info Vocabularies working group of which Hermankova, Horster, and Prag are all members. For more details see <https://epigraphy.info/vocabularies_wg/>.

**Question**: *Standardized terminologies: data on combination of vocabularies systems*

## # A tibble: 3 x 3  
## standard\_method\_no n ratio  
## <dbl> <int> <dbl>  
## 1 1 5 55.6  
## 2 2 2 22.2  
## 3 3 2 22.2

**Commentary**: The majority of projects (55.56%) uses only one method to record their standard terminologies, while 44.44% of projects use a combination of two or three methods. The internal authority lists are used in combination with the EAGLE vocabularies both in their original and modified form. Sharing or publication of internal authority lists would be thus highly beneficial for improving the existing EAGLE vocabularies.

**Question**: *Linked Open Datasets: The project uses the following systems:*

## linked\_data  
## 1 Pleiades  
## 2 Trismegistos  
## 3 EAGLE vocabularies  
## 4 EDH People  
## 5 PIR  
## 6 Adriatlas  
## 7 Cartapulia  
## 8 LGPN  
## 9 OxREP mines database  
## 10 Period.O  
## 11 We provide TM references in our bibliography but inconsistently and without cross linking  
## n ratio ratio\_all\_proj  
## 1 8 25.81 88.89  
## 2 8 25.81 88.89  
## 3 5 16.13 55.56  
## 4 2 6.45 22.22  
## 5 2 6.45 22.22  
## 6 1 3.23 11.11  
## 7 1 3.23 11.11  
## 8 1 3.23 11.11  
## 9 1 3.23 11.11  
## 10 1 3.23 11.11  
## 11 1 3.23 11.11

**Commentary**: From the listed Linked Open Datasets (LOD), Pleaides and Trismegistos are by far the most popular, being used in 88.89% of all participating projects. The EAGLE vocabularies are used in 55.56% of all participating projects. The onomastic and prosopographic data, represented by EDH People, PIR, and LGPN are used only sporadically, namely in 22.22% or 22.22% of all participating projects, suggesting there is a great space for improvement and potentially great benefit in creating onomastic and prosopographic LOD.

# Digital epigraphy projects

This section summarises the results of the online survey *Digital epigraphy in 2022: scoping survey* aimed at digital projects listed under the Digital Epigraphy Projects on the Digital Classicist Wiki page (<https://wiki.digitalclassicist.org/Category:Projects>). The survey was originally sent to 83 projects. We have received 25 responses to the survey, with a response rate of 30.12%.

The digital projects represent a wide range of projects from well-established projects to short-term mostly PhD projects, with the average duration of a project being 6 years. The shortest participating project reported their duration as 1 year and the longest 117 years.

**Question**: *Is the project still active?*

## # A tibble: 3 x 3  
## `Is the project still active?` n ratio  
## <chr> <int> <dbl>  
## 1 Currently not, but we are considering a re-start 4 16  
## 2 No, the project is closed 3 12  
## 3 Yes 18 72

**Commentary**: 72% of participating projects are still active, while 12% of projects are permanently closed and do not consider restarting in the future. 16% of projects are currently not active, but might be reactivated in the future.

## Language coverage

**Question:** *What is the predominant language of epigraphic data in your project (for mixed collections or collections where other languages are predominant provide details in Other)*

## language n ratio  
## 1 Greek 15 37.5  
## 2 Latin 10 25.0  
## 3 Phoenician 2 5.0  
## 4 Akkadian 1 2.5  
## 5 Ancient Languages of the Mediterranean area 1 2.5  
## 6 Arabic 1 2.5  
## 7 Aramaic 1 2.5  
## 8 Hattian u.a. 1 2.5  
## 9 Hittite 1 2.5  
## 10 Hurrian 1 2.5  
## 11 Luwian 1 2.5  
## 12 Neopunic 1 2.5  
## 13 Other 1 2.5  
## 14 Palaeo-European 1 2.5  
## 15 Palaeo-Hispanic 1 2.5  
## 16 Punic 1 2.5

**Commentary:** The language coverage of the participating projects consisted predominantly of Latin and Greek projects (representing 62.5% of the answers). The Other category encompassed a substantial part of the surveyed projects, documenting the need to expand beyond the traditional Latin and Greek focus of the epigraphic discipline. The languages listed as Other consisted of Akkadian, Ancient Languages of the Mediterranean area, Arabic, Aramaic, Hattian u.a., Hittite, Hurrian, Luwian, Neopunic, Other, Palaeo-European, Palaeo-Hispanic, Punic.

## IT infrastructure

**Question**: *Does the project have a website?*

## # A tibble: 2 x 2  
## Website n  
## <chr> <int>  
## 1 No 1  
## 2 Yes 24

**Commentary**: The vast majority of the participating projects maintains an online presence (as of February 2022).

**Question**: *Does you project have an IT specialist(s)?*

## # A tibble: 9 x 3  
## IT\_spec n ratio  
## <chr> <int> <dbl>  
## 1 N/A 7 28  
## 2 No 6 24  
## 3 Yes, equivalent of part-time (<1.0 FTE) position 5 20  
## 4 Yes, equivalent of full-time (1.0 FTE) position 2 8  
## 5 depending on development steps; expertise and experience transfer~ 1 4  
## 6 We are in cooperation with an IT specialist (equivalent of full-t~ 1 4  
## 7 We do not have an IT specialist permanently assigned to the proje~ 1 4  
## 8 We had 1 4  
## 9 We have the support of two IT specialists for maintenance and sma~ 1 4

**Commentary**: Only 8% of projects have an equivalent of 1.0 FTE or more at their disposal. The 32% of digital projects have an IT specialist available for at least several hours per week or share them with other digital projects within their institution. Several projects report difficulty with finding financial resources to support further development and long-term sustainability of the project or even day-to-day support. 28% of the participating projects report currently does not have any access to any IT support. An additional 28% of projects did not indicate whether they have access to IT support (most likely because they are no longer active).

**Question**: *Does your project store epigraphic data in the following formats…?*

## # A tibble: 12 x 3  
## format n no\_format  
## <chr> <int> <dbl>  
## 1 N/A 7 1  
## 2 Epidoc XML 5 1  
## 3 SQL or similar 4 1  
## 4 CSV, SQL or similar 1 2  
## 5 Epidoc XML, CSV 1 2  
## 6 Epidoc XML, JSON, RDF, CSV 1 4  
## 7 Epidoc XML, SQL or similar 1 2  
## 8 JSON, SQL or similar, the xml version of the data is availab~ 1 3  
## 9 None - we use analog systems (printed), 3d viewers 1 2  
## 10 RDF 1 1  
## 11 SQL or similar, We are working on providing also an Epidoc X~ 1 2  
## 12 XML adapted from Epidoc XML 1 1

**Commentary**: Epidoc XML represented one of the two main output data formats with 36% of participating projects, while the data stored in SQL and similar database languages represented 32% of participating projects. 4% of projects indicated the use of analogue data, such as printed materials and 3D format. 4% of projects indicated using their own version of Epidoc XML, adapted to their specific needs. JSON (8%) and RDF (8%) formats are used only by a small number of projects and mostly as complementary data formats to more popular formats such as Epidoc XML or SQL. A relatively large portion of projects did not indicate any format of the data (28%, a number corresponding with the number of no longer active projects).

20% of projects use only one type of data format, while 28% two or more data format types (such as Epidoc XML, SQL or similar, CSV, JSON, RDF).

## Data sharing

### Active projects

**Question**: *Do you share your data outside of your project?*

## # A tibble: 10 x 3  
## share n ratio  
## <chr> <int> <dbl>  
## 1 Yes, under a Creative Commons license 7 38.9   
## 2 Not currently, but we are thinking about it 3 16.7   
## 3 so far without explicit license 1 5.56  
## 4 Under demand 1 5.56  
## 5 we periodically share our data with the Europeana platform 1 5.56  
## 6 Yes, publishing contributions with link to the Catalogue of the ~ 1 5.56  
## 7 Yes, under a Creative Commons license, and also French Etalab Li~ 1 5.56  
## 8 Yes, under a Creative Commons license, by login through guest pa~ 1 5.56  
## 9 Yes, under a Creative Commons license, We are linked with other ~ 1 5.56  
## 10 Yes, without any license 1 5.56

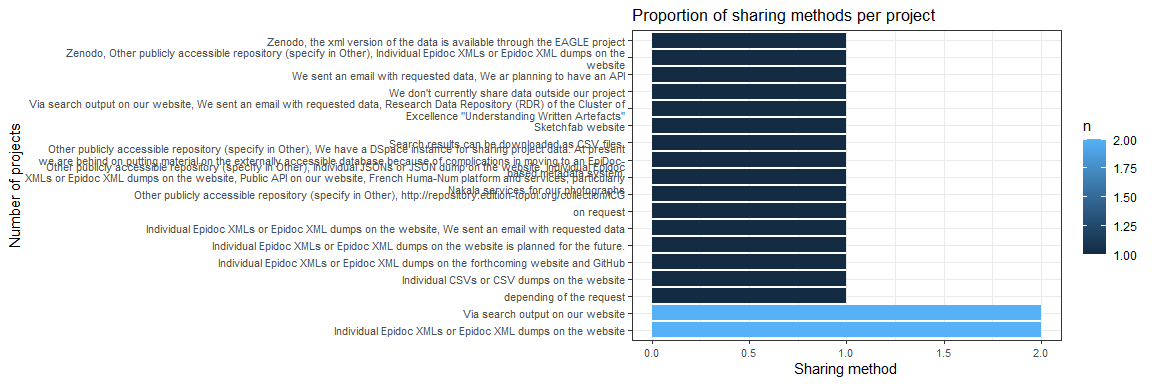
**Commentary**: As of February 2022, 18 projects participated the survey as active projects. The majority of active projects are willing to share their data, representing 83.37% of participating projects. 55.57% of participating projects share the data under a Creative Commons license, which is the preferred mode according to the FAIR data principles. 11.12% of participating projects share the data without any specific license, 5.56% of participating projects provide the data only on demand.

**Question**: *How do share your data with users outside your project?*

## # A tibble: 16 x 3  
## share\_all n share\_method  
## <chr> <int> <dbl>  
## 1 "Individual Epidoc XMLs or Epidoc XML dumps on the websit~ 2 1  
## 2 "Via search output on our website" 2 1  
## 3 "depending of the request" 1 1  
## 4 "Individual CSVs or CSV dumps on the website" 1 1  
## 5 "Individual Epidoc XMLs or Epidoc XML dumps on the forthc~ 1 1  
## 6 "Individual Epidoc XMLs or Epidoc XML dumps on the websit~ 1 2  
## 7 "on request" 1 1  
## 8 "Other publicly accessible repository (specify in Other),~ 1 2  
## 9 "Other publicly accessible repository (specify in Other),~ 1 6  
## 10 "Other publicly accessible repository (specify in Other),~ 1 2  
## 11 "Sketchfab website" 1 1  
## 12 "Via search output on our website, We sent an email with ~ 1 4  
## 13 "We don't currently share data outside our project" 1 1  
## 14 "We sent an email with requested data, We ar planning to ~ 1 3  
## 15 "Zenodo, Other publicly accessible repository (specify in~ 1 3  
## 16 "Zenodo, the xml version of the data is available through~ 1 2

**Commentary**: As of February 2022, all active projects provide at least one way of sharing the data (whether it may be currently accessible to the public or not, or it is intended to be accessible in the future). The average (median) number of sharing methods per project is 2, while the maximum number is 6 (e.g.,Other publicly accessible repository (specify in Other), Individual JSONs or JSON dump on the website, Individual Epidoc XMLs or Epidoc XML dumps on the website, Public API on our website, French Huma-Num platform and services, particularly Nakala services for our photographs).

**Question**: *What is the proportion of sharing methods across projects?*



**Commentary**: The majority of digital projects do not provide any indication of the method for sharing their data, which severely limits any attempts of discipline-wide integration of data. From those who share the data, the Epidoc XML format is the most popular format for data sharing, as well as search output on the project’s website. Open Science practices do not seem to be a popular choice in digital epigraphy, such as sharing via public repository on GitHub or Zenodo, as well as providing raw data in the CSV (comma-separated value) format, or JSON (JavaScript Object Notation) files. Computer-automated access to data is rare and manual human interaction, such as manual selection, manual download of files prevails, potentially hindering any quantitative and reproducible studies, or linking of datasets via automating processes. For example, an API access point is currently available only to a very limited number of projects.

### Closed projects

**Question**: *Is the data created by your project accessible?*

## # A tibble: 3 x 3  
## share n ratio  
## <chr> <int> <dbl>  
## 1 Yes, under a Creative Commons license 5 71.4  
## 2 Not currently, but we are thinking about making it available 1 14.3  
## 3 Yes, without any license 1 14.3

**Commentary**: As of February 2022, 7 of the participating projects are closed. 71.43% of them provides access to their data under a Creative Commons license even though the project is no longer active, 14.29% of closed projects provide access without any license and 14.29% does not currently provide access to the data they have created during the duration of their project, but they are considering to make the data available.

**Question**: *Is the data created by your project accessible?*

## # A tibble: 8 x 3  
## service n ratio  
## <chr> <int> <dbl>  
## 1 Individual Epidoc XMLs or Epidoc XML dumps on the website 4 57.1  
## 2 Public repository on GitHub 3 42.9  
## 3 Other publicly accessible repository (specify in Other) 2 28.6  
## 4 https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.50~ 1 14.3  
## 5 https://open.library.ubc.ca/collections/squeezes 1 14.3  
## 6 ILC4CLARIN Repository 1 14.3  
## 7 Via search output on our website 1 14.3  
## 8 We don't currently share data outside our project 1 14.3

**Commentary**: As of February 2022, 7 of the participating projects are closed. Out of these projects, 57.14% provide their data in the Epidoc XML format on their website. 42.86% provide their data via public repository on GitHub, 28.57% via other publicly accessible repositories, such as ILC4CLARIN Repository. 14.29% of closed projects don’t currently share data outside the project.

## Institutional policies

*Question:* *Does your institution or funding body require your project to comply with any data policies (e.g., FAIR principles, data storage, data sharing, Open Science)?*

## # A tibble: 8 x 3  
## policies n ratio  
## <chr> <int> <dbl>  
## 1 No 10 40  
## 2 N/A 7 28  
## 3 Yes 3 12  
## 4 Neither our grant funding (NEH), private funding, nor institution~ 1 4  
## 5 Not with an official request, at the moment 1 4  
## 6 Policies are on the way, but not yet established. 1 4  
## 7 The French National Centre for Scientific Research strongly encou~ 1 4  
## 8 We don't work for any institution 1 4

**Commentary**: 12% of projects are required to comply with data related policies, while an additional 16% of projects are encouraged to comply with FAIR data principles but no rules are enforced. 40% of projects do not have to explicitly follow any policy and 28% of projects did not disclose any information.

**Question**: *If you have answered YES in the previous question, please specify what are the policies, or provide a link.*

## [1] "https://www.uio.no/english/for-employees/support/research/research-data-management/fair-data/"   
## [2] "All : French \"Plan national pour la science ouverte:Open Science\", https://www.ouvrirlascience.fr/plan-national-pour-la-science-ouverte/; FAIR principles, Mandatory deposit of our publications on the open archive HAL, https://hal.archives-ouvertes.fr/"  
## [3] "Creative Commons"   
## [4] "data sharing"

**Commentary**: Digital policies in the field of digital epigraphy are still being implemented, which does not reflect yet on past and current projects. There is a variation between national policies amongst our responses, with France providing a vocal example in the implementation of Open Science in digital epigraphy.

## Open Science Practice

**Question**: *Are you familiar with the FAIR data principles?*

## # A tibble: 3 x 3  
## policy n ratio  
## <chr> <int> <dbl>  
## 1 Yes 18 72  
## 2 Vaguely 6 24  
## 3 No 1 4

**Commentary**: The majority of projects (72%) is familiar with FAIR data policy, however, 24% of participating projects are familiar only vaguely and would benefit from clear guidelines customised for the epigraphic community. Only 4% of projects are not familiar with FAIR data principles.

**Question**: *Standardized terminologies: The project uses the following systems:*

## standard\_terminologies  
## 1 Internal authority lists  
## 2 EAGLE vocabularies as provided at https://www.eagle-network.eu/resources/vocabularies/  
## 3 Own version of EAGLE vocabularies (edited for our project)  
## 4 We don't use any standardized lists  
## 5 https://epigraphie.mom.fr  
## 6 The project suggests the use of vocabularies in digital projects dealing with ancient writing cultures  
## 7 We created our own thesaurus with OpenTheso tool (EpiVoc) https://thesaurus.mom.fr/opentheso/?idt=th61 and we aligne with existing vocabularies (work still in progress)  
## 8 We generated a system for metadata based on the UBC library's ability to categorize objects (it was very limited for ancient objects)  
## 9 We use standard Mycenological terms but the community does not yet have standardized lists.  
## 10 We use the data provided by Konkordanz der Hethitischen Keilschrifttafeln (www.hethiter.net/hetkonk)  
## n ratio  
## 1 12 36.36  
## 2 7 21.21  
## 3 5 15.15  
## 4 3 9.09  
## 5 1 3.03  
## 6 1 3.03  
## 7 1 3.03  
## 8 1 3.03  
## 9 1 3.03  
## 10 1 3.03

**Commentary**: 9.09% of projects don’t use any standardized lists or vocabularies. 36.36% of projects use their internal authority lists. EAGLE vocabularies in their original form are used by 21.21% of projects, and in its edited version by 15.15% of projects. Several projects focusing on languages other than Greek and Latin have created their own systems including thesauri, e.g. the response:We created our own thesaurus with OpenTheso tool (EpiVoc) https://thesaurus.mom.fr/opentheso/?idt=th61 and we aligne with existing vocabularies (work still in progress) or We use standard Mycenological terms but the community does not yet have standardized lists..

**Question**: *Are you willing to share the standardized terminologies used in your project with us (e.g. type of inscription vocabularies, type of material etc.)*

## # A tibble: 2 x 3  
## policy\_share n ratio  
## <chr> <int> <dbl>  
## 1 Yes 22 88  
## 2 No 3 12

**Commentary**: Vast majority of participating projects (88%) is willing to share any standardized terminologies used in their project, such as terminologies covering the type of inscription vocabularies, the type of material etc.

**Question**: *Linked Open Datasets: The project uses the following systems:*

## linked\_data n ratio  
## 1 Pleiades 13 22.41  
## 2 Trismegistos 12 20.69  
## 3 EAGLE vocabularies 8 13.79  
## 4 LGPN 7 12.07  
## 5 None 3 5.17  
## 6 PIR 3 5.17  
## 7 diacritical marks from Leiden (CIL) 1 1.72  
## 8 Geonames 1 1.72  
## 9 GODOT: https://godot.date/home 1 1.72  
## 10 I can't remember (sorry!) 1 1.72  
## 11 iDaiGazetteer 1 1.72  
## 12 idRef 1 1.72  
## 13 None were yet available: a new edition will want to use all 1 1.72  
## 14 Pactols 1 1.72  
## 15 Period.O 1 1.72  
## 16 ToposTexts 1 1.72  
## 17 under demand 1 1.72  
## 18 We periodically ask to Trismegistos an ID for our records 1 1.72

**Commentary**: Pleiades is the most popular LOD dataset, being used in 22.41% of all participating projects, followed by Trismegistos with 20.69%. EAGLE vocabularies are represented in 13.79% of participating projects, while LGPN in 12.07% of projects. Only 5.17% of participating projects do not use any LOD.

# Future needs of digital epigraphy

This section covers the wishes of partner projects as well as all participating digital epigraphy projects. The responses were anonymised so no individual or project can be identified but otherwise presented as submitted in the survey.

## Partner projects

**Question**: *Our project would like to be able to use within the next three years:*

## lod\_f  
## 1 Bibliographical references to all epigraphic publications with stable URI (e.g. Zenon)  
## 2 EAGLE vocabularies (revised and extended with clear structure + eliminated duplicates + multi-language support)  
## 3 Roman Prosopographical data with stable URIs  
## 4 Greek Onomastic data with stable URIs (e.g. LGPN with stable identifiers)  
## 5 One domain specific repository for epigraphic data  
## 6 Open and accessible RDF Triplestore  
## 7 We are not sure what is meant by "epigraphic data" in the preceding entry. If something like a papyri.info for inscriptions then no. If a basic aggregator like Humanities Commons for epigraphy then that would be nore useful.  
## n ratio ratio\_all\_proj  
## 1 8 22.86 88.89  
## 2 7 20.00 77.78  
## 3 7 20.00 77.78  
## 4 4 11.43 44.44  
## 5 4 11.43 44.44  
## 6 4 11.43 44.44  
## 7 1 2.86 11.11

**Commentary**: The most popular is the option Bibliographical references to all epigraphic publications with stable URI (e.g. Zenon) with 22.86% of responses representing the wishes of 88.89% of all projects. The great interest in onomastic and prosopographical LOD for both the Greek and Roman world is supported by 31.43% of positive responses from 77.78% and 44.44% of projects respectively. The improved EAGLE vocabularies are wished for by 77.78% of participating projects. The domain-specific repository for epigraphic data or the open and accessible RDF Triplestore do not seem to be the highest priority of participating projects, but still relatively popular as 44.44% of responses wishes for one of the two. One participating project wishes specifically for the following: We are not sure what is meant by "epigraphic data" in the preceding entry. If something like a papyri.info for inscriptions then no. If a basic aggregator like Humanities Commons for epigraphy then that would be nore useful..

**Question**: *Potential ideas that our project would benefit from:*

## lod\_i n  
## 1 Set of guidelines for FAIR and Linked Open Data in epigraphy 9  
## 2 Practical scripted examples on how to use LOD in epigraphy 7  
## 3 Workshop on FAIR principles in epigraphy 6  
## 4 Set of guidelines/resources for quantitative analysis of epigraphic data 5  
## 5 Workshop on how to use LOD in epigraphy 4  
## ratio ratio\_all\_proj  
## 1 29.03226 100.00  
## 2 22.58065 77.78  
## 3 19.35484 66.67  
## 4 16.12903 55.56  
## 5 12.90323 44.44

**Commentary**: 100% of all projects would benefit from A set of guidelines for FAIR and Linked Open Data in epigraphy. There is a general interest in practical examples and workshop(s) on how to use LOD and FAIR Principles in Epigraphy, as well as resources for quantitative analysis of data in epigraphy.

**Question**: *Additional digital needs*

## [1] "Further development of a single research portal to interrogate multiple epigraphic databases; development of a specific API to use the standardized common vocabularies"   
## [2] "- Further collaboration and development of concepts for vocabularies. - Getty vocabularies crosswalks where they apply - In doing all this work, we hope that FAIR Epigraphy will use as many different applications of the EpiDoc schema as possible, so as to accommodate the ways different projects mark up documents and metadata."  
## [3] "Sustainable common platform of all digital epigraphic editions (a Vision)"   
## [4] "Advisory Board for new Digital Epigraphy projects, guidelines for FAIR epigraphy"

**Commentary**: This section covers the additional needs of partner projects. Partner projects would like to see a platform linking epigraphic data from multiple sources, including a stable reference point or an API for improved epigraphic vocabularies. Partner projects would also like to be able to use guidelines of FAIR practices in epigraphy, that currently do not exist.

## Digital epigraphy projects

**Question**: *Our project would like to be able to use within the next three years:*

## lod\_f  
## 1 Bibliographical references to all epigraphic publications with stable URI (e.g. Zenon)  
## 2 EAGLE vocabularies (revised and extended with clear structure + eliminated duplicates + multi-language support)  
## 3 Greek Onomastic data with stable URIs (e.g. LGPN with stable identifiers)  
## 4 One domain specific repository for epigraphic data  
## 5 Roman Prosopographical data with stable URIs  
## 6 Open and accessible RDF Triplestore  
## 7 None  
## 8 Geolocation of inscriptions and searches related to geography  
## 9 In the case of our project most of the options are not applicable  
## 10 LGPN does not yet contain Mycenaean names but I would be happy if that changed  
## 11 This project is currently closed  
## n ratio ratio\_all\_proj  
## 1 16 20.78 64  
## 2 16 20.78 64  
## 3 12 15.58 48  
## 4 11 14.29 44  
## 5 10 12.99 40  
## 6 6 7.79 24  
## 7 2 2.60 8  
## 8 1 1.30 4  
## 9 1 1.30 4  
## 10 1 1.30 4  
## 11 1 1.30 4

**Commentary**: The most popular is the option Bibliographical references to all epigraphic publications with stable URI (e.g. Zenon) representing the wishes of 64% of all participating projects. The great interest in onomastic and prosopographical LOD for both the Greek and Roman world is supported by 52% and 40% of positive responses from participating projects. The improved EAGLE vocabularies are wished for by 64% of participating projects. The domain-specific repository for epigraphic data (44%) or the open and accessible RDF Triplestore do not seem to be the highest priority of participating projects (24%), but still a relatively popular response. One participating project wishes specifically for the following: Geolocation of inscriptions and searches related to geography.

**Question**: *Potential ideas that our project would benefit from:*

## lod\_i  
## 1 Set of guidelines for FAIR and Linked Open Data in epigraphy  
## 2 Practical scripted examples on how to use LOD in epigraphy  
## 3 Set of guidelines/resources for quantitative analysis of epigraphic data  
## 4 Workshop on how to use LOD in epigraphy  
## 5 Workshop on FAIR principles in epigraphy  
## 6 In the next three years we planned a few Digital Epigraphy workshops in the frame of the French School at Athens  
## 7 None  
## n ratio ratio\_all\_proj  
## 1 21 25.609756 84  
## 2 16 19.512195 64  
## 3 16 19.512195 64  
## 4 16 19.512195 64  
## 5 11 13.414634 44  
## 6 1 1.219512 4  
## 7 1 1.219512 4

**Commentary**: 84% of all participating projects would benefit from A set of guidelines for FAIR and Linked Open Data in epigraphy and 84% from Workshop on FAIR principles in epigraphy. There is a general interest in practical examples (64%) and workshop(s) on how to use LOD in epigraphy (84%), as well as resources for quantitative analysis of data in epigraphy (64%). There might be potential synergy in organising workshops in digital epigraphy between the participating projects, e.g. the French School at Athens.

**Question**: *Additional digital needs*

## [1] "Digitalization of Roman Inscriptions for dissemination and research"   
## [2] "A workshop on integrating Mycenaean data into epigraphy?"   
## [3] "Data retrieval also on spatial base: for example: from maps of the single archaeological sites and single complexes (as plans or 3d scans of catacombs and churches...). Links with the existing geographical and georeferenced resources. Controlled and shared vocabulary about palaeographical features; Storage, search and analysis of the 'aberrant forms' (not to be 'corrected') for Late Latin and Late/Byzantine Greek words (and names)."  
## [4] "The most important for me would be 1/ to have a more complete view of real FAIR epigraphic projects and 2/a sustainable \"common place\" where to find resources + tools and help + let's call it an improved EAGLE + and more \"international\""   
## [5] "It would be very nice (but I might be a bit biased!) if FAIR Epigrahy would like to help develop EFES (EpiDoc Front-End Services). For example by helping to make the existing RDF data export functionality really usable even by less experienced people."   
## [6] "I would love to see it revitalized and improved with FAIR and Linked Open Data guidelines and other resources."   
## [7] "Unicode for Punic"   
## [8] "help to act in a shared dedicated academical environment and help in spreading our results"   
## [9] "FAIR Epigraphy's team can help us by providing advice on specifical topics"

**Commentary**: This section covers additional needs of participating digital projects. Some of the wishes might be beyond scope of the FAIR Epigraphy project but the responses provide valuable guidance and hint to some of the challenges the epigraphic discipline will be facing in the near future. The responses may inspire other projects with similar needs to join forces and potentially develop the solution together.

# Summary

The present report demonstrates a great variation of the epigraphic discipline in 2022. Although the majority of participating projects record inscriptions in Latin and Greek, we see a diverse array of projects expanding beyond the traditional boundaries of the discipline. Participating projects involve well-established projects that exist over several decades, regional or thematic corpora or more specialised short-term PhD projects. We have noticed a clear distinction between projects with a long tradition and most importantly institutional support, that have access to institutional repositories, policies and IT services and the small-scale projects with limited support and access to resources and training. One of the missions of the FAIR Epigraphy project is to support projects with limited access to resources by providing accessible and comprehensible training and guidelines for FAIR and Linked Open Data principles in epigraphy.

The established projects mostly follow the FAIR principles, although to a variable extent. Most established projects share their data under a Creative Commons license in one or more widely accepted formats (with Epidoc XML being the most popular format for all types of projects irrespective of their status). In general, the more established projects provide more access points to the data as well as more data formats than the projects with less institutional support. The use of standardized terminologies is still limited and project-specific, mostly due to the lack of uniformly accepted standards. On contrary, the adoption of Linked Open Datasets and connecting providing datasets with stable identifiers seems to be fairly advanced, especially in the case of established LOD such as Pleiades or Trismegistos, and to some degree EAGLE vocabularies.

As to the current and future needs of digital epigraphy, there is a growing demand for more LOD, especially for bibliographical references of standard epigraphic corpora, standardisation of discipline-specific vocabularies, and onomastic and prosopographic LOD, all supported by training and providing accessible resources and sets of guidelines for FAIR and Open epigraphy. The need for an accessible and open platform connecting and linking various epigraphic resources into one is generally supported, building on the experience of the EAGLE Project.

# Bibliography

Bodard G., Roueché C. & Reynolds J. (2007). Inscriptions of Aphrodisias (IAph2007). Unknown Publisher.

Roueche C. (2022). Inscriptions of roman tripolitania 2021. The Society for Libyan Studies.

Roueche C., Reynolds J. & Bodard G. (2020). Inscriptions of roman cyrenaica (2020). The Society for Libyan Studies.

Tupman C. (2021). Where can our inscriptions take us? Harnessing the potential of linked open data for epigraphy. Archaeopress, Oxford, p. 115–128.

1. Johannes Gutenberg University in Mainz, [petra.hermankova@uni.mainz.de](mailto:petra.hermankova@uni.mainz.de), <https://orcid.org/0000-0002-6349-0540> [↑](#footnote-ref-20)
2. Johannes Gutenberg University in Mainz, [horster@uni.mainz.de](mailto:horster@uni.mainz.de), <https://orcid.org/0000-0003-1434-224X> [↑](#footnote-ref-23)
3. Oxford University, [jonathan.prag@merton.ox.ac.uk](mailto:jonathan.prag@merton.ox.ac.uk), <https://orcid.org/0000-0003-3819-8537> [↑](#footnote-ref-26)