Needs assessment

To assess the interest of the scientific community for an open and fair lead isotope database with global coverage and its expectations to the proposed infrastructure, a survey was designed. The survey was split in five parts which aimed to (1) gather information about the scientific background of the participants and some demographic data related to their (academic) age, (2) their current approaches in working with lead isotope data, (3) their experiences with already existing databases and repositories and their willingness in contributing to a new data infrastructure, (4) their interest in a web-based interface, and (5) general comments. The survey was widely announced (mailing lists, personal networks, social media) before any part of the GlobaLID prototype was published. The overwhelming majority of the participants filled it in before its publication. 46 colleagues participated in the survey between August 13 and October 22, 2021. Although 46 participants seem to be a small number at the first glance, the number of researchers that is working with lead isotope data in archaeological research on an everyday basis is estimated to be between 200 and 300 worldwide. Consequently, the survey represents about a 22.5 % to 15 % of the global community, which is a very good coverage taking into account that the networks of the project team are mostly limited to Europe.
The overall results of the survey clearly shows that the community is looking forward to such an infrastructure or even deem it long overdue because current approaches are cumbersome at best. Alongside the survey, personal communication to the project members consisted of positive feedback and expressions of interest throughout, including an invitation to a conference.

Background of the participants

Figure 1 presents some basic demographic data of the participants. As expected, the majority of the participants have more than 12 years experience in research (counted from the master degree on) and most of them are between 31 and 50 years old. This indicates that a lot of well-experienced researchers participated in the survey and will hopefully interested to provide thoughtful feedback to the planned infrastructure in the course of its development. At the same time, it can be expected that these persons stay active in the field long enough to commit to the planned infrastructure on a long-term basis.
The country of residence mirrors the network of the project team with a clear dominance of Europe with Germany and France in particular. At the same time, it can be regarded as symptomatic of the community in general. Despite recent efforts and increased interest in e. g. African metallurgy and the strongly increased research activities in China, researchers from these areas are still not that well connected to Western researchers and their networks as it would be preferable from our perspective (e. g. there are hardly any contributions on the respective mailing lists). Hence we plan to put a special focus on the inclusion and training of researchers from these regions within the GlobaLID project.

The participants were also asked to provide some information about the type of institution they are currently working at and their academic background (Figure 2). As expected, the vast majority works in a research context and there mostly at a university. However, a few participants currently

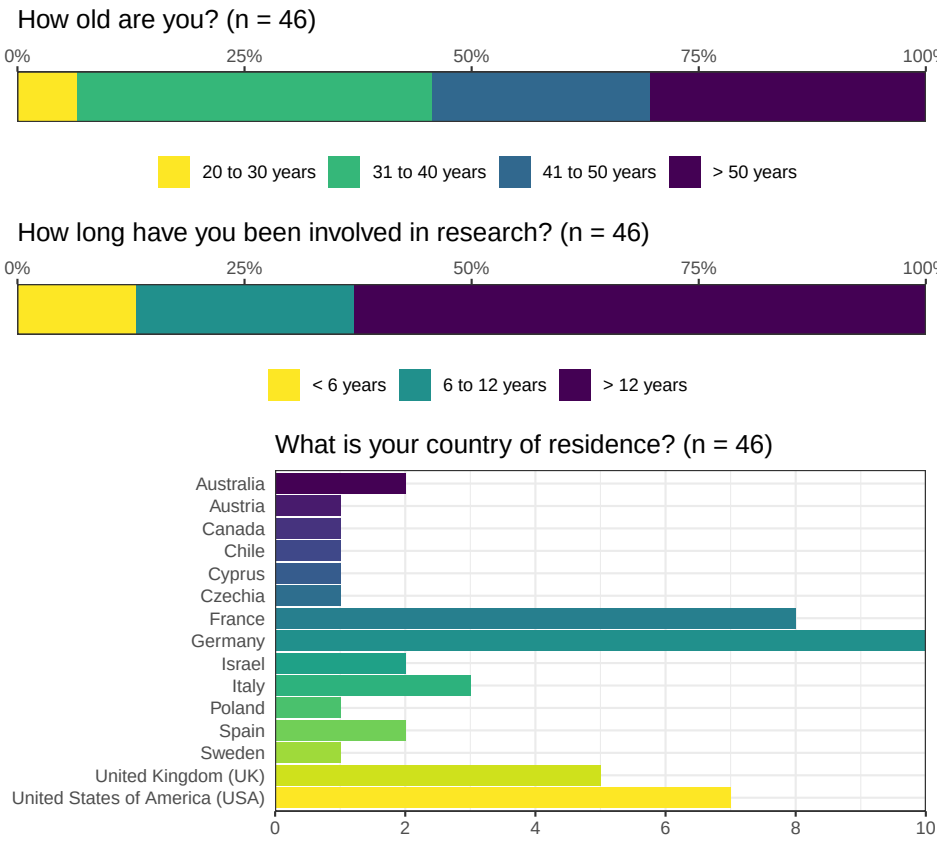

Figure 1: Basic demographic data of the participants.

work in the commercial sector, which proves that our project is also interesting for researchers outside Academia. Likewise, all relevant research environments are represented among the participants like museums and geological surveys.
Similarly broad is also the academic background of the participants. Archaeometrists and archaeologists clearly dominate, followed by geoscientists. However, the lead isotope method is also of relevance for researchers with expertise outside these fields as the survey shows. This holds particularly true for numismatics, where the high chronological resolution of coins provides the perfect conditions to reconstruct the evolution of supply networks and economic processes.

It might be for these reasons that the reconstruction of networks for e. g. exchange or supplies was named by over half of the participants in their answers to the question “Which topics are you investigating with lead isotopes?” (Table 1). Even more mentioned provenance analysis as a topic, which was stated by 11 participants as the only topic they use lead isotopes for. This was to be expected because the provenance of raw materials is the central research question to which lead isotopes are applied to and the starting point of further investigations like the reconstruction of exchange networks. Single mentions of e. g. recycling, geochronology, or pollution demonstrate the high interdisciplinary character of the method and the participants.

Current work practices with lead isotope reference data

Nearly all participants use Excel sheets to store analytical data and also their lead isotope reference data (Figure 3). Besides the widespread use of Excel to handle spreadsheets and the higher comfort is offers compared to character-separated spreadsheets, their preference for the storage of the reference data can be explained by the availability of the OXALID data as Excel spreadsheets

Table 1: The different research topics provided by the participants (n = 43) and how often they were mentioned.

Topic	n
Provenance analysis	36
Network reconstruction (exchange, supply)	20
Metal processing reconstruction	4
Ore-forming processes	2
Recycling	2
Ancient mining	1
Arabia and Turkey	1
Archaeometallurgy	1
Changing patterns	1
Circulation of metals	1
Dating	1
Earth sciences	1
Geochronology	1
Grouping of alloys	1
Human-environment relations	1
Isotope mixing modelling	1
Metal characterisation	1
Metals	1
Mining technology	1
Mixing	1
Mobility reconstruction	1
North africa	1
Ore	1
Pollution	1
Resource exploitation	1
Silver and lead metallurgy	1
South-east Asia	1
Spread (or lack of) of innovation of ideas	1
Technological analysis	1
The organisation of metal industries on different scales	1

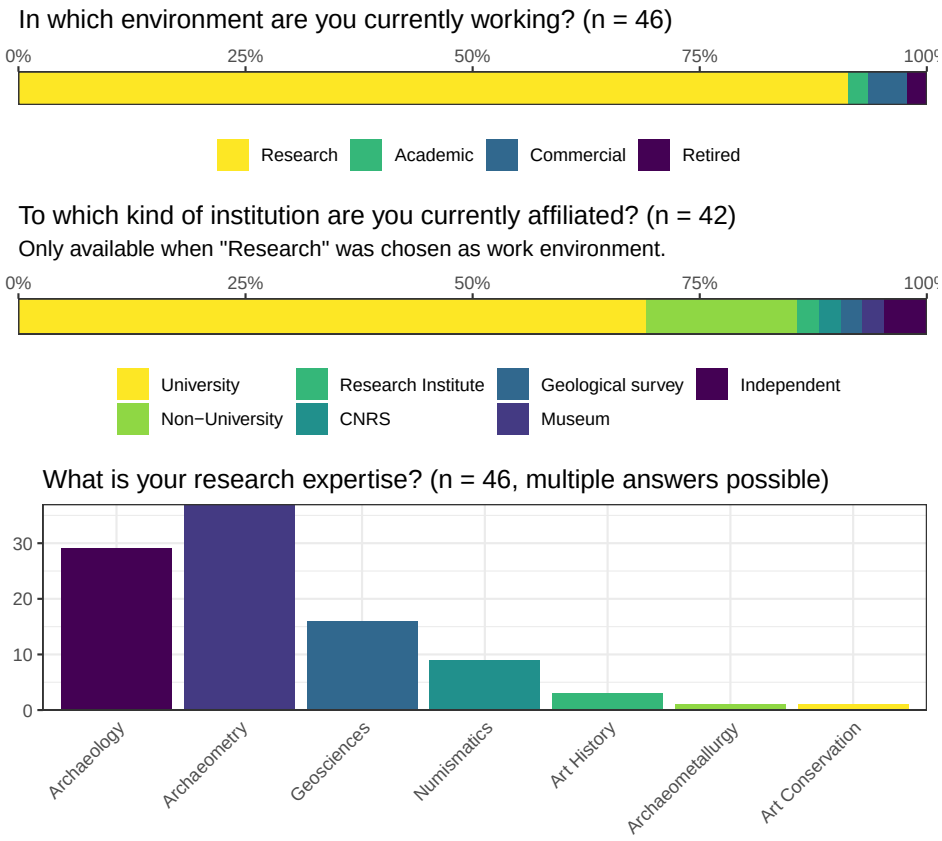

Figure 2: Type of affiliation and academic background of the participants.

which can be downloaded ready-to-use. Similarly, the higher proportion of character-separated spreadsheets for the storage of own analytical results might be because the analytical instruments usually store or export the data in this file format. Additionally, 8 participants already use a database for their reference data and 5 participants store their analytical data in other file formats than common spreadsheets. In conjunction with the problems Excel files can create because of different character encoding, this highlights the need for the common use of a universal and open file format. In the best case, lead isotope data are saved in a standardized structure to facilitate the sharing of them between researchers and the integration of new data in local data collections.

The most often used software type is spreadsheet programs (Figure 4), probably because it is easy to handle tabular data with them and they allow a first look on the data at the same time. Over a quarter of the participants (additionally) use more advanced software that was specifically designed for data analysis or plotting of the data. Admittedly, the differences between the software types are often not that clear-cut as it might seem from the response options. Programming languages often offer functions for the statistical treatment of data as well as elaborated plotting functions. The programming language R for example was initially developed for statistics but is now also widely used to create elaborated plots. The trend towards the use of specialized programs at least for the visualization of data can also be observed in the publications, where only rarely the typical Excel plots can be seen anymore.

OXALID is the most used database (Figure 5), which was to be expected because it is the oldest and most famous one, and the Mediterranean is still one of the major areas for research. It is followed by the database of Killick and his colleagues, most likely because it is the only one that covers a wide range of non-European countries. However, the most common approach is to compile

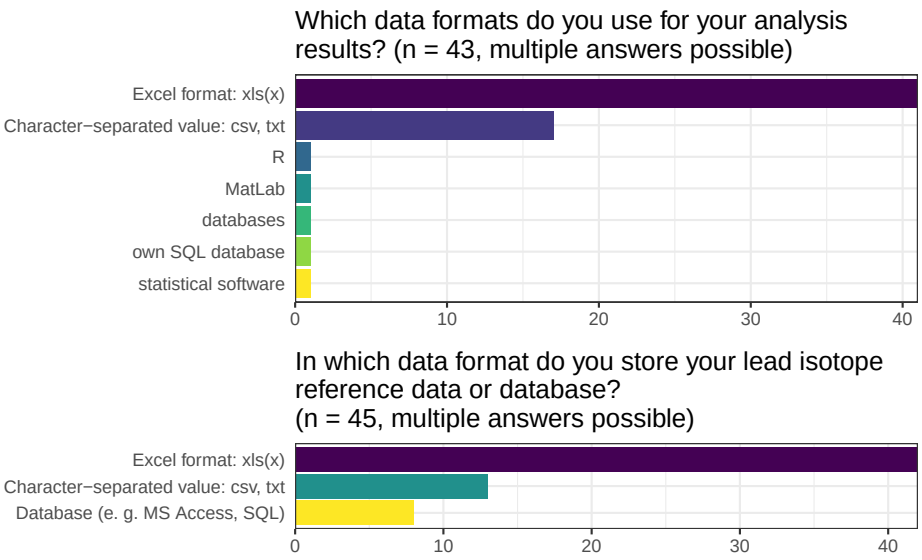

Figure 3: Preferences in the use of file formats for the storage of own analytical data and lead isotope reference data.

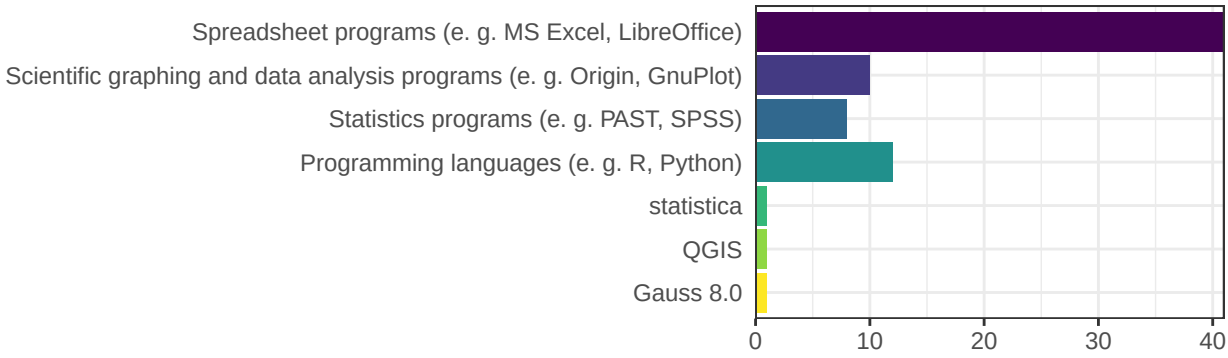

Figure 4: Which data processing software do you use to work with lead isotope data? (n = 45, multiple answers possible)

data from scattered literature. This is a time-consuming but necessary task because all the listed databases are static and hence do not contain the newest data, and there is none that does this, yet. The use of unpublished data is also widely spread although this is highly problematic from a scientific point of view – especially if conclusions are built upon them.

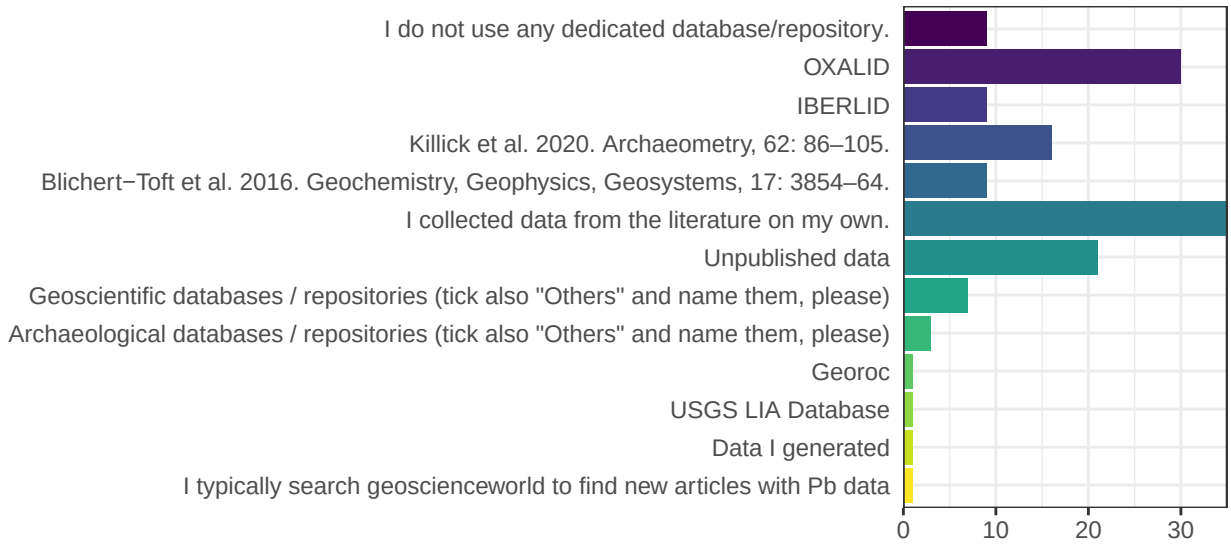

Figure 5: Which databases / repositories do you already work with? (n = 45, multiple answers possible)

27 commented on their experiences with these databases ("How has your experience been with these services?"). Several of them commented negatively on the comparability of the data they contain (e. g. "Irregular reporting of raw data, irregular reporting of precision/accuracy…", "… data not always compatible or in the same format.", "Inconsistencies in data entry; issues with data quality") and their usability (e. g. "cumbersome", "valuable as a fundamental source of reliable raw data, but not really easy to use", "Inefficient doing it this way."). Particularly criticized was the often missing or non-satisfactory amount of meta-data (e. g. "lack of geospatial components, lack of geological context information, oftentimes no accompanying elemental/mineralogical data", "issues with where actually and what actually was sampled and analyzed", "The most difficult is to connect geographic location with geological information."). "Data are spread out everywhere, which requires much time to organize them" as one participants remarked. This was even mentioned in a comparably positive comment: "good, generally it is time consuming to collect and process the data". It became clear that there is a high level of frustration about the current state of scattered, inconsistent and incomplete data, and colleagues are longing for an infrastructure like GlobaLID. This is also confirmed by the next question – only one person is not interested in actively contributing to GlobaLID (Figure 6). This is a very positive sign for GlobaLID because only with the active and sustained support of the community the database can grow quickly large enough to become a viable alternative.

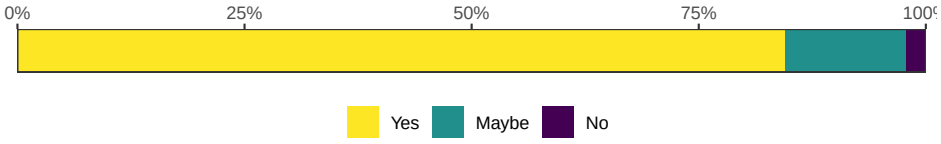

Figure 6: Could you imagine to contribute to a global lead isotope data database? (n = 45)

A web application as interface to GlobaLID

Parallel to the database, a web application is planned to facilitate access to the database and the work with lead isotope data. All of the participants could imagine to switch to such a web application (Figure 7), if it meets their expectations (see below). All of the features suggested by the project team are regarded at least as “important” by a strong majority of the participants (Figure 8). Additional features suggested by the participants and not already included in the prototype of the web application were:

- The inclusion of archaeological materials with respective meta-information (date, culture, locality, context)
- Activity times of mines
- Error bars and/or other indicators for the quality of the data
- Lead content
- Specific plot types
- Possibility to save and resume work at a later stage

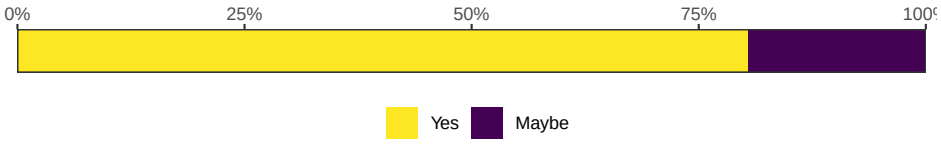

Figure 7: Could you imagine to switch to a web application to access and use lead isotope reference data? (n = 46)

The majority of the 32 participants that commented to the question “How would a web-based database and application infrastructure likely impact your work?” said that it would speed up their workflow considerably especially because there would be no need anymore to browse through the literature (e. g. “It would be extremely valuable not to take time to do the work of collecting the data and not to miss data I could not have access to”, “It will save me a lot of time”, “would speed up analysis times, with more focus on interpretation and less on laborious data collection, formatting and cleaning, etc.”), leaving more time to focus on the scientific work with the data (“Speeds up data processing = speeds up publishing of results”). Other participants also see the dangers of such an low-level approach (e. g. “I can imagine, on the other hand, that some people may be more prone to treat the results with less critique (beginners - like me, students etc.), so in my opinion a warning should be added for the users, to treat the results as a step towards the interpretation and should be assessed with necessary critique”) or see a potential problem in the coverage of the database and how quickly it can be updated (e. g. “Very useful, time saving, fear it would not be updated quickly enough”, “If enough data is available, it could drastically reduce the time spent looking for reference data…”). Some participants are aware that such a database could open up new ways of research and collaboration (“… it would also facilitate a variety of”big-data” approaches to interregional analyses.“,”I also see a great opportunity for the community to extend networking.“). One participant even commented “It is something that everybody in the field has dreamed of. It would be great to have a constantly updated database, with easy access/use. … We did not do it because of missing time/resources.“, and another emphasized that an involvement of the community is necessary (”Some wrong data must be discharged but this needs consensus based on experts criteria”). GlobaLID aims to provide exactly this: the consensus should be reached at the first workshop, and the provided funding in combination with a committed institution will allow to invest in an infrastructure for which support and investment is not available otherwise.

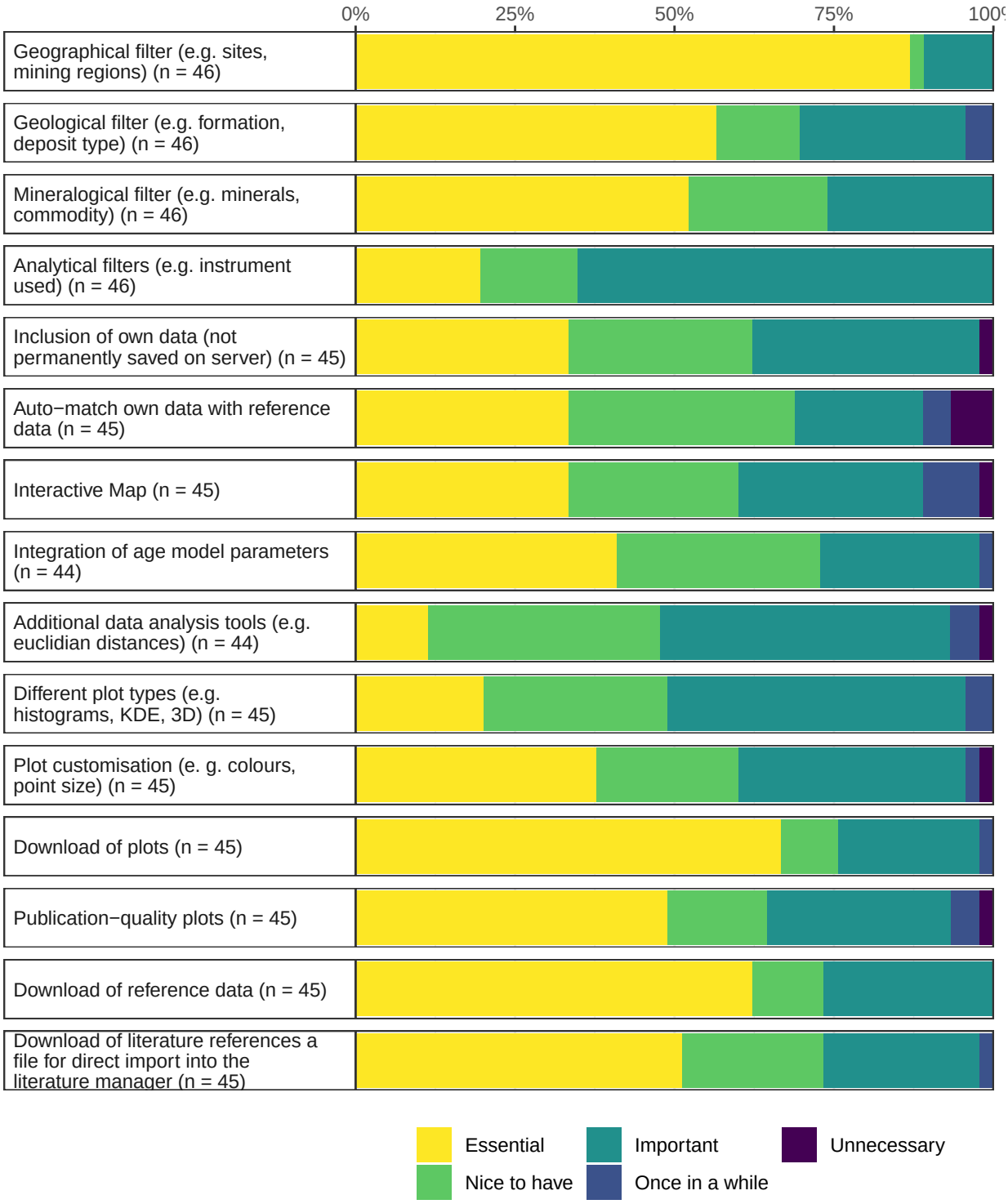

Figure 8: What features should an attractive interactive web application have?

That the participants are very aware about the quality of the data is also visible from their comments on the question “What would prevent you from using a web-based database interface?” (n = 32), where several participants mentioned a “Lack of quality control” or “If data proved unreliable”. Similarly, a low usability of the web application would keep the participants from using it (e. g. “Difficulties of use, non user-friendly interface, etc.”, “If it is only usuable by experts in geochemistry. The interface would need to be understandable for archaeologists/numismatists too.”, “a poor interface; poor instructions / support”). Other important aspects mentioned are the reliance on an internet connection (“a blackout. No, seriously, we also work in areas with sparse internet signal (Arabia for instance) and necessity to be online to work with the database could be an obstacle. An off-line solution might be a welcome option.”), if it is behind a pay-wall or if it results in not citing the original publications anymore (“Not being credited for every use of data I have bust my guts trying to produce!”).

General questions

Being asked “What would be your alternative?” (n = 20) all participants said in one way or another to carry on as they do already. This is also supported by comments received on the question “Time constrains and technical limitations aside, how would your optimal workflow with lead isotope data look like?” (n = 22), where most participants described what they are currently doing. The main potential for them seems to be that “Automatisation of the analytical processes would be useful, not only because of all the available data being at one repository (what a relief!), but also the knowledge, that the processes offered by the app are standard, justified and approved by a scientific community.”.

On the last question (“Is there anything else you would like to mention?”, n = 16) we received a lot of very positive comments:

- Great initiative!! Let me know if I can offer support.
- I would like to mention that this is really an excellent initiative and I would be happy to help - contribute, assist in any stage and form!
- It must be done
- Congrats for this initiative that all Pb isotope users are waiting since long time
- fantastic project idea, I love it! Hopefully some dinosaurs will finally provide their raw data…!
- I would be happy to contribute with data in French publications of PhD thesis that would be hard to access from abroad. This is a great project, I am looking forward to seeing it work!
- Good luck! This is a fantastic project

But there were also three aspects that were critically commented on, two of them were already mentioned previously:

- Who will commit to keeping this database online and updated forever? My past experience is that these tend to run a few years, and then disappear, unless a government agency, professional society, or well-funded museum commits to maintain the database in perpetuity
- Needs designated person to maintain in long term rather than short term funded for set up only Archaeological material is also needed
- Similar to [redacted]' reply, I do worry that some individuals may not grasp the complexity of the data or the iterative/integrative nature of analyzing lead isotopic data. Perhaps there could also be a training module to help individuals understand the steps necessary to properly contextualize lead isotopic results, rather than just simple pattern matching.

- I see a little problem for younger researchers who need - having in mind our actual grant system - publications in high ranked journals AND citations of these articles. If the data is part of a online database, the user will just cite the database and not the specific article where the data was published. This means the (young) researcher looses citations which is not acknowledge by some evaluation systems.

We fully agree with the apprehensions of the participants and we already planned to tackle them within the project. The second workshop should provide the required training. The Deutsche Bergbau-Museum and the project team are firmly committed to maintain the infrastructure and to create the involvement of the community that is necessary to keep the database up-to-date as long as possible. The last aspect, that especially young researchers might loose citations because the database is cited instead of the original publication is nothing that we can actively engage with. However, the users will always be provided with a full reference list and we encourage them to cite the full list instead of the database. This is also necessary to signify which reference data were actually used.

Conclusions

The results of the survey show very clearly that the suggested data infrastructure GlobaLID will be a welcomed and well-received addition to the lead isotope community worldwide. For many of the participants, such an infrastructure will be a huge step forward that is long overdue. The overwhelming majority of the participants is interested in the web application and, even more important, it is willing to contribute to the database. Additionally, two colleagues already asked the project team how they can submit their data. Based on the feedback in the survey, there are more that stayed anonymous for the moment.

The suggestions made by the participants will be taken into account when developing the database and redesigning the web application. The overall positive feedback encourages us to make the effort in including archaeological data as well. They will add a whole new level of complexity to the data but based on the survey results this part of the database will also become quickly populated. The participants expect the web application to provide an easy access to this complexity in the data, including a suitable filter structure and matching function. Implementation of these features with all their consequences on e. g. visualization will be of priority for the redesign of the web application. Most of the features deemed as “essential” or “important” and also some of the additionally suggested ones are already included in the prototype of the web application. Thus it provides a good starting point. If these features are implemented and the usability of the web application is high, all participants are willing or even keen to include the web application in their workflow. This clearly shows the longing for a more standardized and comparable workflow by the community and the potential of such an application. Consequently, the survey impressivley shows the potential of the proposed GlobaLID data infrastructure to become the major platform for handling and managing lead isotope data worldwide – a true Global Lead Isotope Database.

The results of the survey also made clear that the community expects to be actively involved in the project. For these reasons, two workshops will be organised: one to find a common standard on how to report lead isotope data and which meta-information should be included, and the other to provide training in the lead isotope method. Additionally, feedback from the colleagues will be actively sought during conferences and when testing the different components of the infrastructure. And of course, colleagues will be encouraged to provide feedback at any time.