Tephra Fusion 2022
Feb. 10, 2022
Session 1 - Motivations and Perspectives

Questions and Answers (Q&A) and Links from Zoom Chat

Question: Will recordings of the session be made available?
- Yes, a link will be distributed via the Google Group and will be added to the Tephra Fusion 2022 website: https://tephrochronology.org/cot/Tephra2022/#

Question: Where can I find the best practice documents?

- The best practice spreadsheets can be found on Zenodo. These are currently at version 3.0. https://doi.org/10.5281/zenodo.3866266
- Specific implementations developed for SESAR and EarthChem can be found at https://www.earthchem.org/communities/tephra/
- Two example datasets which implement the best practice recommendations include June Lake Tephra Dataset https://doi.org/10.5281/zenodo.4074289 and Summer Lake Pliocene Tephra Dataset https://doi.org/10.5281/zenodo.4072461
- An example EPMA-SEM method descriptor can be found here: https://doi.org/10.26022/IEDA/112102
- The tephra Fusion 2022 website has more links: https://tephrochronology.org/cot/Tephra2022/#

Question: What are IGSNs, and how do IGSNs work?
1. IGSN stands for International GeoSample Number. See https://www.igsn.org
2. IGSN are persistent identifiers like DOIs (DIgital Object Identifiers https://www.doi.org).
3. IGSNs are resolvable (you can put them into your browser, for example http://igsn.org/IEZHP0001 and it will lead to a persistent metadata profile).
4. IGSNs are globally unique (can not be duplicated) and persistent (no matter where the metadata profile moves to, the resolution will lead to that page).
5. The benefit of the IGSN is that there is an online accessible set of metadata about the sample (it can be kept private if necessary), which helps sample management and sample reuse. The IGSN can be embedded as a link into bar codes or QR codes that will; make it possible to link from a sample label directly to all the metadata and data for a given sample.
6. USGS is also implementing IGSN for samples.

Question: Somebody mentioned that IGSN was moving to Data Cite. Where can we find more info about this?

1. IGSN is partnering with DataCite. You can read the announcement here, https://blog.datacite.org/igsn-and-datacite-partnership/ and see more about it in our AGU poster https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/992144. For a copy of the poster, email us at info@igsn.org

Question: What is machine actionable?
1. Depending on the context - machine actionable can mean that a machine (computer system and software) is able to process or use the information without additional input or support from a human. The data and metadata are structured in a way where the machine can process it and take an action.
2. "This term refers to information that is structured in a consistent way so that machines, or computers, can be programmed against the structure." https://ddialliance.org/taxonomy/term/198
3. Machine actionable means that computers can directly access, retrieve, and interpret data without human interaction.

Question: We have heard a lot about FAIR and Open data from the US perspective but how are FAIR data principles enacted around the globe?
1. There are some international organizations and a need to engage with professional societies. Many major societies like AGU, EGU have sections which focus on data. Some nations have national data infrastructure involved with FAIR data.
2. Belmont forum is one of the organizations that helps to coordinate internationally.
3. Funding agencies can also coordinate proposal solicitations with other funders.
4. Please also view discussion of this topic in the session video.

Question: How do these efforts fit into ICON science principles? - i.e. Integrated, Coordinated, Open, and Networked (ICON) Science to Advance the Geosciences https://www.essoar.org/doi/10.1002/essoar.10508554.1 ICON also incorporates the human dimension of scientific research. In this context, networked refers to networks of people, not computers.
1. NSF acknowledges efforts to a networked and open research enterprise and how support for the human element is important.

Question: Field work and a lot of lab work are disproportionately time and resource intensive compared to modeling studies, and they have much higher risk of delays or deficits due to practical issues not controlled by the scientists. If scientists are still hired, evaluated and awarded grants based on how many papers and successful studies they publish - then how do we as a community ensure that it is attractive to do the field- and

lab-work in a world of freely accessible data that allows for much faster and low-risk turnaround studies focused on the data of others?

1. You are correct that reusable data will facilitate many kinds of research. There also will continue to be a need for new data. For example, in producing an eruption history and volcanic hazards report for poorly-studied volcanoes, new research may be required. For applied work like developing chronologies for sediment cores, new lab work will be needed whenever that hasn't been done previously.
2. Great questions, if only that were a major problem in the tephra world. It is far more common that you cannot find people's data. So people redo fieldwork or delay their research waiting on others' results. So the goal here is to improve data quality and sharing and ultimately more timely publication of data. One way to encourage employers to start new work including field and lab studies is to say you are following holistic best practice guidelines and will build a fundamental dataset that is needed and will be used extensively.
3. It is important to note that FAIR data are citable, and implementation of data citation will help researchers get credit for sharing data.
4. Usage licenses can also be applied which specify what may or may not be done with a data set. (See also Guy Jones' discussion of this topic in the video.)
5. Another part of the FAIR data movement is crediting researchers for producing data, not just for publishing papers. Making data sets citable with DOI is a key part of facilitating this. Many have argued that hiring and professional evaluation practices need to change to acknowledge the production and citation of data.

Question: Is the "data descriptor" publication subject to the same standards of self-plagiarization evaluation that would be applied to a research article? Because, I could see that there could be a lot of overlap between data descriptors for papers presenting similar datasets for different geographic areas.

1. Yes. At Scientific Data, data and method publications also go through technical peer review (via exactly the same process as 'regular' article publications). For papers discussing areas already covered by other research articles, then similar rules apply. We will consider: Is this a necessary addition to the literature? Is this data or discussion available elsewhere? Does this need a data descriptor, or does the dataset stand on its own? (See also Guy Jones discussion of this topic in the video.)

Question: Are there any initiatives that focus on making or keeping historic data FAIR as well? I'm wondering about it in the broadest sense like "rescuing" or securing data from databases that are no longer curated as well as previously published datasets etc.

1. There have been some general data rescue projects in the past through ESIP, RDA, and CODATA. There are projects like PetDB and GEOROC where curators are ingesting data from publications.
2. The EARThD (East African Rift Tephra Database) project is creating a searchable database of previously published tephra datasets from the East African Rift with funding

from the U.S. NSF. So far we have standardized and entered data from 400+ dat sets, and 100 of these have been ingested by the Interdisciplinary Earth Data Alliance (IEDA) and are available on PetDB (IEDA Petrological Database). The EARThD group is always on the lookout for published datasets we may have missed.

3. In the US Geological Survey, we know this is a valuable need. There are efforts to accomplish some of this over time. We have a database of "data rescue" candidates, but unfortunately the funding and resources are not always available to attend to them. It's an important need, however, as legacy data provides a wealth of information!

Question: How do you find geochemical data and tephra samples submitted to USGS?

1. This is a good question - the USGS has many different programs, and there is no single tephra data repository at this time. Data from the Alaska Volcano Observatory (USGS funded) are contained in GeoDiva: https://www.avo.alaska.edu/geochem/ which is a growing Db and only published data are accessible. Please email Kristi Wallace for more information - kwallace@usgs.gov. Other data may be in Science Base or in supplements to papers. Sorry that this is not a very satisfying answer. Accessibility of USGS tephra data is a work in progress.

Question: Many of the repositories seem quite focused on geochem data, is it also possible to add physical data (e.g. grain size, componentry data)?

1. Excellent question. Yes, we do have good repositories for geochemistry, samples, etc. Grain size and componentry is a current area that needs to be addressed better. Ideally, this data would be formatted in a consistent way to make it easier to re-use. One option to explore could be StraboSpot as max clast data, for example, is often done in the field. A conversation has been opened with StraboSpot, but that needs some follow up.
2. In the context of physical data/samples, SESAR - the system for earth sample registration is a metadata repository for samples. If you register the sample in SESAR you can capture information about the physical data like grain size. IGSNs issued by SESAR can be included in other datasets (for example in submission to EarthChem) in order to connect back to this information.
3. The IEDA data facility focuses on sample-based data, not only chemical data, but also petrography, physical properties, etc. The EarthChem Library already hosts grain size datasets.

Question: Beyond all the good physical parameters you have mentioned, juvenile textures are also very important to distinguish different events.

1. Very true we will make sure to include this in the best practices framework!
2. The physical properties and textures definitely deserve to have a dedicated database or /repository - either stand alone or included into another DB like GeoDiva.

Question: The physical properties include more than just grain size. Do these repositories allow for density, 3D volume model reconstructions, componentry, porosity, isomass/isopach data, etc? I would be happy to help develop a framework if these items are not readily addressed in a repository. To my knowledge, there isn't a good physical properties data repository, but I will look forward to Strabospot next week, and I will look into SESAR too.

1. Great question. I think databases to store other datasets are rare and built for purpose, and there is NO single Db/repository that does all of this. AVO's GeoDiva is growing modules 1 at a time to incorporate. It would be good to hear from EarthChem on this topic too.

2. I forgot to mention that we have developed templates for upload to EarthChem that include these data (simplified as well as comprehensive) but they are not mapped to the EarthChem Db (i.e. fields do not exist within the database), but stand alone data files and collections of files can still be uploaded and discoverable by others. It is just that EarthChem does not have Db fields for all of the tephra data that we need at this time, but you can upload a spreadsheet with everything.

3. From Lucy the EarthChem curator: EarthChem accepts multiple types of petrographic observations. We currently do not have the capacity to support 3D volume models but we are always open to discussions about new data types that are required by the community.

Links Shared in the Chat

Tephra Fusion 2022 first circular and agenda documents which list the speakers:
https://www.tephrochronology.org/cot/Tephra2022/pdfs/TephraFusion2022FirstCircular.pdf
https://www.tephrochronology.org/cot/Tephra2022/pdfs/TephraFusion2022Agenda.pdf

IAVCEI Commission on Tephrochronology Membership Form
https://forms.gle/xDCYcByMPrfehJVc9

IAVCEI Commission on Tephrochronology Facebook page:
https://www.facebook.com/IAVCEICOT/?business_id=10152592499697447

IAVCEI Commission on Tephrochronology Twitter page: https://twitter.com/IAVCEI_COT

Best practice spreadsheets on Zenodo https://doi.org/10.5281/zenodo.3866266

StraboSpot tephra help document: https://www.strabospot.org/files/StraboSpotTephraHelp.pdf

Tephra Community page at EarthChem: https://earthchem.org/communities/tephra/
12:03:45 From Stephen Kuehn to Host and Panelists:

Registry of Research Data Repositories https://www.re3data.org

CrossRef - find research objects https://www.crossref.org

Links from the NSF presentation:
- NSF Proposal & Award Policies & Procedures Guide (PAPPG): https://www.nsf.gov/publications/pub_summ.jsp?ods_key=pappg
- NSF GEO data policies: https://www.nsf.gov/geo/geo-data-policies/index.jsp
- NSF GEO Cyberinfrastructure Opportunities: https://www.nsf.gov/geo/geo-ci/index.jsp
- NSF EarthCube program page (NSF 21-515): https://beta.nsf.gov/funding/opportunities/earthcube
- NSF CSSI program page (NSF 21-617): https://beta.nsf.gov/funding/opportunities/cyberinfrastructure-sustained-scientific-innovation-cssi
- NSF Geoinformatics program page (NSF 21-583): https://beta.nsf.gov/funding/opportunities/geoinformatics-gi
- National Academies' "Earth in Time" report (2020): https://www.nap.edu/catalog/25761/a-vision-for-nsf-earth-sciences-2020-2030-earth-in
- NSF Petrology and Geochemistry program page (NSF 22-560): https://beta.nsf.gov/funding/opportunities/petrology-and-geochemistry-ch-0
- NSF FAIROS RCN program page (NSF 22-553): https://beta.nsf.gov/funding/opportunities/findable-accessible-interoperable-reusable-open-science-research-coordination
- NSF Public Access Repository (NSF-PAR): https://par.nsf.gov
- NSF GEO Quarterly Webinar https://nsf.gov/events/event_summ.jsp?cntn_id=304260&org=NSF
- NSF EAR Listserv sign-up page: https://www.nsf.gov/geo/ear/ear-listserv.jsp