

For announcements on ESS-DIVE activities (i.e. webinars, publications, new feature announcements)...

Follow ESS-DIVE on Twitter! [@ESSDIVE](https://twitter.com/ESSDIVE)

Join ESS-DIVE's Community Mailing List! <https://groups.google.com/a/lbl.gov/g/ESS-DIVE-Community>

Polling Questions: pollev.com/essdive

Welcome session notes

Talks in this session:

- Welcome Address
- Intros and Demos
- BER ESS Data Management
- ESS-DIVE Features
- Data Management Plans
- Reporting Formats and Standards
- Project Demos

Q: Is there going to be a recording of the meeting , or at least tutorials?

A: There will be no recordings, but all presentations, including the tutorials will be available in the shared Google Drive and the Community GitHub repository linked above. We also have short video tutorials on our ESS-DIVE YouTube channel.

Q: Can you submit real-time data?

A: I noticed someone was interested in submitting real-time data. That's an interesting problem but not really something we've looked at in a lot of detail. But I'd be curious to hear requirements so if it comes up in one of your breakouts, feel free to ping me and I can try and hop on.

Q: Do you have to register as a new contributor each time, even if you already have data published on ESS-DIVE?

A: At this time, you only need to register once.

Q: Is ESS-DIVE hosting data from ESS-relevant projects outside DOE? Should we recommend it to collaborators and colleagues from anywhere in the world?

A: Great question. ESS-DIVE is specifically for DOE ESS data. For data from other programs, we might be able to store it if it is ESS relevant but we have to request permission. I would not recommend broadly advertising us as available for storage outside the ESS program.

Q: How are you distinguishing between portal vs project space?

A: The portal is a feature in support of projects where you can gather all the data on a theme or project. The project spaces are the broad set of features we are building to support projects (portals are one of these).

Q: Where do we add acknowledgment about reading the contribution terms?

A: It is up to you as to if/where you want to include your acknowledging that you have read the contributor terms as part of your proposal. If you indicate that your data will be stored in ESS-DIVE, please also indicate that you have read the ESS-DIVE terms and conditions. One common location is in your PMP.

Q: Do you have to worry about copyright issues? e.g. images that were included in a publication. can they be included in ess-dive?

A: Yes, we try to look for copyright issues. For e.g. we flag any files that are papers where the copyright may not be held by the author. We don't do a comprehensive copyright check. It is your responsibility as a data submitter to make sure you have the copyright - it is part of the terms you agree to when you submit data to be published. If there is a pre-existing publication where the figure was published, please provide a citation to the figure/image along with the image/figure in the dataset.

Q: What if we as authors don't have copyright, but DOE does? Does this make any difference?

Example: *When I publish a paper (as a contractor at Argonne), I don't have copyright, but we include a statement when submitting the paper that the DOE/US govt. does retain copyright. "The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. <http://energy.gov/downloads/doe-public-accessplan>"*

A: If you don't have the rights to your data figures/products, you can link out to it in your data package if it has a citation. We recommend getting a final decision from your Principal Investigator (PI) or your Intellectual Property Office (IPO).

Q: Will ESS-DIVE be archiving and making available open source tools to help PI's add metadata to their data?

A: Certainly something on our wishlist to provide, still providing mainstream functions. We've been able to augment the capabilities of the community through community funds and reporting formats. We like to work with members of the community - we don't currently have funds available but eventually we would like the community's help, so if you have any ideas or want to get involved please reach out and once we have community funds available we will reach out to those that are interested.

Q: Is there a preference for all data to be zipped?

A: It is your preference whether to zip files or not. ESS-DIVE does not store file hierarchy within a folder, if it is necessary for data reuse for the files to be in a hierarchy, you will need to zip your files. If your zip file exceeds the Submission Form Upload limit (10 GB), you will either need to (a) upload the data files individually or (b) use the Package Service Application Programming Interface (API) to upload the files.

Follow up comment: oh....I think preserving a structure would be really important for the future. Putting files inside a compressed archive means the data is far less accessible, e.g. from an API or remote direct URL download.

That is, of course 100s of files in one folder is a mess but if we can preserve a file hierarchy then that means end users can transport through archives and find the files they might want in their workflows

Q: In the contributor terms of use it says "Discuss data archiving needs: data quantity, interaction of project data systems and personnel with ESS-DIVE, and formats with the ESS-DIVE team (ess-dive-support@lbl.gov)." is this cover when I submit a request to be a data contributor or it is a separate process/questions?

A: We will update that text, thank you for pointing that out. You do not need to contact ESS-DIVE with your data archiving needs outside of your contributor request.

Q: How will ESS-DIVE be integrating with existing and future data archives? Are their plans for data federation (like) capability?

A: There are 2 cases: 1. The data is meant to be archived with ESS-DIVE because it is ESS data. You need to include at least the metadata on ESS-DIVE and link out to the external repository where the data is currently stored.

2. In the case of uploading data that is not ESS data but is relevant - In the near future, we will have the capability to have metadata stored in ESS-DIVE but will link to other repositories/external archives. In the long term, we are working on integration aspects that will allow for connections with repositories that are of significant value (more on a case by case basis). There is a fair amount of integrative search currently, though, and data on ESS-DIVE is going to DataONE nodes.

Q: oh....I think preserving a structure would be really important for the future. Putting files inside a compressed archive means the data is far less accessible, e.g. from an API or remote direct URL download. if we can preserve a file hierarchy

then that means end users can transport through archives and find the files they might want in their workflows

A: At the moment we don't have a way to preserve hierarchical file organizations.

Q: Is ESS-DIVE a good place to provide analyses files which are not suitable for supplementary material in a paper but are needed to reproduce the analyses/figures provided in the paper. For example, raw outputs from Markov Chain Monte Carlo analysis.

A: In ESS-DIVE we are happy to take any data that is non-proprietary. We encourage you to include any files which you think are relevant and helps users interpret, read, and utilize your data-- it should absolutely be included.

Q: Some data have existing, community accepted repositories (e.g. NCBI for 'omics data). How should these data be archived if they are collected for an ESS project?

A: We require that sufficient metadata for ESS data be published on ESS-DIVE. If the data is already publicly available elsewhere with a DOI, we just ask that

Q: Sending you to another archive, is this automated? In that in a script would it automatically allow the code to take you to the actual data location? If not on ESS-DIVE already.

A: We encourage linking out to data files in other repositories if they are also reliable long term data repositories.

Q: Good question about analysis file (I am thinking small python scripts for plotting/manipulating data). Is that appropriate or should that go on GitHub or somewhere else?

A: Analysis and processing scripts are encouraged for upload if you think the scripts are relevant and helps users interpret, read, and utilize your data-- it should absolutely be included.

Q: What about using DOECODE too?

A: If you do have major software that is a place you go to register code. If you have small scripts most people prefer not to register it. If your institution requires

that you publish your code you are welcome to store with ESS-DIVE. DOECODE is also a place that you can link to. We will treat the code as data, so if you have requirements to license your software though you should go through DOECODE and deal with the licensing

Q: What is the procedure to add descriptive data packages such as title, abstract, keywords, authorship, and funding information that correspond to two different projects that produced two different publications but used the same data?

A: You could potentially split your data packages depending on whether they are different enough. We put the project name as the publisher so that it shows up in your citation. Right now, we ask you to choose one primary project to put in the citation.

Q: Can data submissions be updated with corrections or extended/ongoing data submissions once published? Or do changes require a new submission/DOI? For example, if manuscript reviewers request new analyses or changes?

A: We recognize that that is a common case. You are given the choice - minor revisions you are able to keep the same DOI, major revisions maybe get a new DOI. This is a choice you make when you make the update.

Q: I am assuming that mined, digitized, and manipulated data can be stored on ESS-Dive as long as it is appropriately referenced? I am thinking of digitized historical data that are being used for a EESSD project

A: Depends on the licensing policy of the existing dataset. If it allows derivative projects then yes we can store. If it is an original dataset (eg. published by NASA) there may be restrictions so we would need to deal with that on a case by case basis.

Q: I would be interested in learning more about the data fusion concept. Will it be covered during the tutorial session tomorrow? When would it be available for the community to use?

A: In development and it was one of the efforts that will need to be supported by reporting formats. The current prototype can handle some amount of parsing but

without standardized data files, there are limitations. We will be having user interviews in the future, please reach out to us if you are interested in participating.

Q: If ESS-DIVE is the primary repository for our data, can we deposit copies in another repository? What if those data are published in the other repository in a slightly different manner (e.g., different units or a combination such as aggregating across different depth increments)?

A: You can do that. If the data that is going to be published could fit within a relevant reporting format and you are not using that reporting format in the original publication, then we would prefer that you put the data into the reporting format before submitting to ESS-DIVE. If the data is well documented and described in a long living repository, then you can link to the original and reference between the two data packages, and we leave that up to you to decide.

Q: We have a lot of phenocam images stored in a phenocam database. What should we do if phenocam loses funding in the future? How will we preserve this high volume of data?

A: ...