

2020 ESIP Summer Meeting (July 14th-24th)

<https://2020esipsummermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

## **Drafting Guidelines For Vocabulary Selection By Data Curators and Repositories**

Wednesday, July 22 • 4:00 pm - 5:30 pm ET

**Organizers: Margaret O'Brien** (0000-0002-1693-8322), **Mark Schildhauer** (0000-0003-0632-7576)

**More info & slides on Sched Session Page:** <https://sched.co/clw7>

---

**Participants (Please add name and or ORCID):** [Mark Schildhauer](#), Margaret O'Brien, UCSB, ORCID: 0000-0002-1693-8322, [Charles Vardeman](#), [Kai Blumberg](#), [Kathe Todd-Brown](#) (UFlorida) , Steve Aulenbach (ORCID: [000-0002-0172-6538](#)), Ge Peng, [Ruth Duerr](#), [gpeng@ncsu.edu](mailto:gpeng@ncsu.edu), [0000-0002-1986-9115](#), Karen Stocks 0000-0002-1282-300X, Mikki Johnson (0000-0001-6133-0247), [Pier Luigi Buttigieg](#), Gastil Gastil-Buhl [0000-0001-9177-5722](#), [Brian Westra](#), [Robert J. Rovetto](#), [Stephanie Holmgren](#), [Teddy Gelabert](#), [Gary Berg-Cross](#), Frank Nitsche (ORCID: 0000-0002-4137-547X), Hannah Sweets, LDEO (ORCID: 0000-0002-3708-3917) Tyler Stevens (NASA/GCMD), Maria Stoica, CU Boulder, Jaclyn Saunders (WHOI/BCO-DMO; ORCID: 0000-0003-1023-6239), Jay Pearlman Four Bridges, Don Collins (NOAA/NCEI), Denise Hills, Geo Survey of Alabama (ORCID: 0000-0001-9581-4944), Matt Jones (0000-0003-0077-4738), John Graybeal (0000-0001-6875-5360), Kristin Vanderbilt, Yang Xia

**Agenda** (Session leads, please modify this as desired)

## Resources

- **Presentations (if not linked in Sched)**
  - <https://docs.google.com/presentation/d/1OwlnYH3d3Ndzw6hUwdTv78o1ZBtMNmYnglS7V1V9vP4>
- **Other supplementary materials**
  - **Directory with spreadsheets (5), notes**
    - <https://drive.google.com/drive/folders/1B2mqmylhxzoDipYt7vv0GpHZYKNAh4C2>
  - **Breakout synthesis**
    - [https://docs.google.com/document/d/1wWwcuVeVO3EB9v5B6zwwYXTwtSaPL7xrkZJbgJtrG\\_U/edit#](https://docs.google.com/document/d/1wWwcuVeVO3EB9v5B6zwwYXTwtSaPL7xrkZJbgJtrG_U/edit#)

## 2020 ESIP Summer Meeting (July 14th-24th)

<https://2020esipsummermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

## Agenda:

- **Welcome - ERIN**
  - Orient attendees to tech (Zoom and QiqoChat)
  - Invite attendees to share their name/affiliation/location in the Zoom chat
  - Invite others to take notes in this document (if desired)
  - Assign any other roles needed
- **Community Participation Guidelines Reminder - ERIN**
  - By participating in this session, you agree to adhere to the ESIP [Community Participation Guidelines](#)
  - Report an Issue: <https://www.integritycounts.ca/org/esip>
- **Goals of this Session - MARK, MARGARET**
  - Goal: Collect materials to begin a draft of guidelines to assist users in choosing vocabularies.
  - Agenda Overview ( [go to DETAILED AGENDA, INSTRUCTIONS](#))
    - Intro and instructions
    - Crowdsourc: answer 7 questions about your use of vocabularies
    - Breakout 1: synthesize crowdsourced material
    - Plenary discussion
    - Breakout 2: Follow up
    - The last few minutes should be reserved for recording 3 key takeaways from the session in the Takeaways Document.

## DETAILED AGENDA AND INSTRUCTIONS

Session link: <https://qiqochat.com/breakout/13/GSIBQqAfUSqebTadpHLwUFOBy>

1. Introduction (20 min):
  - a. Session info/techy stuff (5 min) - Erin
  - b. Intro - Margaret and Mark
    - i. Mark: Contextualization vocab landscape
    - ii. Margaret - Goals of this working session: start a draft of guidelines
2. Individual exercise (10 min) Crowdsourc experiences
  - a. Groups of <~20: work in a spreadsheet to list your experience with vocabs. Include benefits, limitations,
  - b. Spreadsheets to collect info (based on your initial) [FOLDER w/all docs](#)

## 2020 ESIP Summer Meeting (July 14th-24th)

<https://2020esipsummermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- i. 1 A-D:  
[https://docs.google.com/spreadsheets/d/1VEfNae0Zp2BVBaM3ciTIGINyGRhICM8I-51\\_0Nmqxm8](https://docs.google.com/spreadsheets/d/1VEfNae0Zp2BVBaM3ciTIGINyGRhICM8I-51_0Nmqxm8)
  - ii. 2 E-I:  
<https://docs.google.com/spreadsheets/d/1oIMsR56CwcQ0CZc6CPLmd7vSBIDnsGOyV1veYyLhcug>
  - iii. 3 J-N:  
[https://docs.google.com/spreadsheets/d/10sIN8Mgy15spDMSNUbOchN78Fbq\\_OBpBbjyrOJteQA](https://docs.google.com/spreadsheets/d/10sIN8Mgy15spDMSNUbOchN78Fbq_OBpBbjyrOJteQA)
  - iv. 4 O-S:  
<https://docs.google.com/spreadsheets/d/1Ls4mUxaT2ixzejHRVftOXr9OUG3GWGt42wLW4pFTB6Q>
  - v. 5 T-Z:  
[https://docs.google.com/spreadsheets/d/1dwlALf-QMe8oJV1KjBrdMMogM3kFAAXTbwce\\_3n8iG8](https://docs.google.com/spreadsheets/d/1dwlALf-QMe8oJV1KjBrdMMogM3kFAAXTbwce_3n8iG8)
- c. Info you will enter
- i. Personal: Name, ID, Theme of Interest and Role
  - ii. Vocab you use (Choose from the list)
    1. URLs are on the vocab tab
- d. What benefits do you gain from using your Vocab?
- e. What features are missing from your Vocab?
- i. Can you contribute to or correct your Vocab?
- f. How do you “use” the Vocab
- g. Other (for notes col):
- i. At what granularity can you access terms in your Vocab-- e.g. term level?

### 3. Breakout A: Synthesizing the Crowdsourcing (15 min)

[https://docs.google.com/document/d/1wWwcuVeVO3EB9v5B6zWuYXTwtsaPL7xrkZJbgJtrG\\_U/edit?ts=5f173df8#heading=h.7hjr9uw8o7dw](https://docs.google.com/document/d/1wWwcuVeVO3EB9v5B6zWuYXTwtsaPL7xrkZJbgJtrG_U/edit?ts=5f173df8#heading=h.7hjr9uw8o7dw)

- a. RANDOM ASSIGNMENT, 5 breakouts (1/spreadsheet) moderators:
- i. Room 1- Pier Luigi Buttigieg 1:A-D  
[https://docs.google.com/spreadsheets/d/1VEfNae0Zp2BVBaM3ciTIGINyGRhICM8I-51\\_0Nmqxm8](https://docs.google.com/spreadsheets/d/1VEfNae0Zp2BVBaM3ciTIGINyGRhICM8I-51_0Nmqxm8)
  - ii. Room 2 - Kai Blumberg 2:E-I  
<https://docs.google.com/spreadsheets/d/1oIMsR56CwcQ0CZc6CPLmd7vSBIDnsGOyV1veYyLhcug>
  - iii. Room 3 - Karen Stocks 3:J-N  
[https://docs.google.com/spreadsheets/d/10sIN8Mgy15spDMSNUbOchN78Fbq\\_OBpBbjyrOJteQA](https://docs.google.com/spreadsheets/d/10sIN8Mgy15spDMSNUbOchN78Fbq_OBpBbjyrOJteQA)

**2020 ESIP Summer Meeting (July 14th-24th)**

<https://2020esipsummermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- iv. Room 4 - Kathe Todd-Brown 4:O-S  
<https://docs.google.com/spreadsheets/d/1Ls4mUxaT2ixzejHRVftOXr9OUG3GWGt42wLW4pFTB6Q>
- v. Room 5 - Shelley Stall 5:T-Z  
[https://docs.google.com/spreadsheets/d/1dwlAlf-QMe8oJV1KjBrdMMogM3kFAAXTbwce\\_3n8iG8](https://docs.google.com/spreadsheets/d/1dwlAlf-QMe8oJV1KjBrdMMogM3kFAAXTbwce_3n8iG8)

- b. CHARGE TO GROUPS: Collectively examine/discuss contents. distill individual perspectives into commonalities, priorities.
  - i. Each group: add a note-taker
  - ii. Take notes in a single google doc. [LINK HERE](#)
    - 1. Use the section for the “alphabetical” Spreadsheet you are reporting on
      - a. Prepopulated empty bullets to reduce text bouncing
      - b. Use “+1s” or similar to denote agreement/priorities
    - 2. Pick a bullet, write your thoughts, or talk them thru and summarize
    - 3. Moderator tries to keep written ideas together (if similar) and coherent

4. Plenary/preliminary discussion (20 min) of Synthesized Outcomes

- a. Moderator: Mark,
- b. Note taker: Margaret
- c. Report-backs from each Group Moderator (5 x 2 mins)

5. Breakout B - Rebuttals, Omissions and Harrumphs (10 breakouts?)

- a. Everyone contributes! Smaller breakout groups!
- b. What additional questions/concerns do you have that are missing from the Synthesized Outcomes? NOW’s the TIME TO SPEAK UP!
- c. Add these to “Additional Concerns” column of which ever Spreadsheet you need
  - i. ([go to Crowdsourcing activity](#))

6. Plenary/Wrap-up (15 min):

- a. Moderator: mark,
- b. Note taker: margaret
- a. Discussion
  - i. Point out Semantic Tech clusters, especially Semantic Harmonization
  - ii. Work with outcomes from Cox/Weyborn to develop *ESIP Guidelines for Vocabulary Choice*
  - iii. Consider activity for existing (or new) ESIP Cluster on Vocabulary

## 2020 ESIP Summer Meeting (July 14th-24th)

<https://2020esipsummermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

## Before session close:

- **Capture 3 Takeaways & Shared insights**
  - What did we discover?
  - How can we help each other?
  - What knowledge and resources do we have to share?
  - Important remaining questions?
- **Closing circle**
  - Share any reminders
  - Shout-outs & thanks
  - Requests for peer assists

## 5 ways to stay involved with ESIP

- Join the conversations on our [ESIPFed Slack Channel](#) and say hello on #general.
- Keep up with the latest from in and around ESIP by signing up for the [ESIP Monday Update](#) or the mailing list of this group.
- Share your feedback, questions, and suggestions with our Twitter community [@esipfed](#) and [#esipfed](#).
- Hop on an [ESIP Telecon](#) that interests you - no need to RSVP!
- At this meeting, pitch an idea for FUNding Friday and check out the Research Showcase.

## Feedback

Please take a moment to share your thoughts on this session in Sched by going to the Sched session page linked at the top of this document.

## Zoom Chat Notes

**Paste in Zoom Chat notes that you would like to save here**

## 2020 ESIP Summer Meeting (July 14th-24th)

<https://2020esipsummermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

### Notes:

- As a data manager, I continue to be challenged to understand the difference between vocabularies and ontologies. And how to apply each. We work mostly with vocabularies as they make sense. Have terms we can map to our problems.
  - As a data collector/analyzer and occasional data manager - I completely agree with Shawn.
- @Shawn: part of the issue is that the terminology development community aren't consistent either, and some resources are mislabelled as one or the other. I think we need community-level consensus building here.
- FROM SIMON COX: It would be good if NASA would host GCMD properly! With a decent UI . I find it easier to access GCMD Keywords through Research Vocab Australia than from NASA
- The interoperability between terminologies (vocab, thesauri, ontologies) is key here - data managers should be able to use what works for them, but what they use should be carefully linked to other resources.
- Yes Lesley! That is exactly what I was thinking. If the goal is to have linked data then shouldn't the ontologies be linked themselves so a user can just search once across all of them?
- What do you mean by It would be good if NASA would host GCMD properly! ?
- Karen has nailed the issue - it is so much easier to develop your own 'bag of terms'
- +1 that the term adding process is too slow, but we have to balance that with QC/QA. training more people to become co-editors would help, as would more middleware to make the process easier.
- Silos are highly efficient for those who built and live in the silo
  - not silos, 'cylinders of excellence'!
  - @Steve: very true, but then the "I" bit of FAIR suffers
- Has any work been done to see if those vocabulary lists are FAIR?
- I'm surprised that ENVO was identified as easy to navigate : ) It's often too big for general use, and the subsets we generate are only used by a few. We really rely on the portals and services (OntoBee, OLS, OntoPortal resources...) to expose the content we develop in a usable way.
  - Yeah - that was not a feature I had experienced ;- ) (SIMON)
- @Ge: yes, some, but most of it is sort of "survey based" - we need more automated testing, but that means more conventional cross-mapping, which is weak/missing. We're trying to support this in the Semantic Harmonisation Cluster / SemTech Committee
- Corinna, Hanna: here's the YAMZ vocabulary tool we were just talking about: <http://yamz.herokuapp.com/> from the CDL
- I totally agree this has been going on for 15 years!

## 2020 ESIP Summer Meeting (July 14th-24th)

<https://2020esipsummermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- From Karen Stocks to Everyone: (2:31 PM)
- Good point John
- From Simon Cox to Everyone: (2:32 PM)
- How to get funding for generic work, infrastructure ...?
- From Lesley Wyborn to Everyone: (2:32 PM)
- Or also rationalise the funding that is going on the silos and put it together to achieve something HAs someone copied the chat!
- From Mark Schildhauer to Everyone: (2:32 PM)
- there is much greater clarity about how to move forward now though; and the growing groundswell of interest helps;
- From steve.aulenbach to Everyone: (2:33 PM)
- Which google doc?
- From Pier Luigi Buttigieg to Everyone: (2:33 PM)
- @Simon: Especially because we do so much of this voluntarily - why would a funder fund this?
- 
- Well, we do it to prevent poor quality stuff from having free-reign
- From Karen Stocks to Everyone: (2:34 PM)
- +1 on the great session design
- From Kathe Todd-Brown (UFlorida, Gainesville) to Everyone: (2:34 PM)
- Bye!
-