February 21, 2020 Methods & Tools Working Group Call

Attendees: Mike, Alexey, Mingkai Jiang, Jake, Jody

Agenda and Notes
Notes from Feb 20 call
1. The EFI Task View list of resources is a useful resource but isn't ready for public consumption
    a. Note from the last RCN Steering call - it would be good to have a running list of the software needs of the community (which is what the Task List provides) and then think about what tools are needed for people to get up and running with making forecasts with NEON data and then what are the tools that are needed more generally for forecasts
    b. Can we assign sections of the EFI Task View for editing to make it ready for public? We can link to it on the website
    c. On November call - we talked about creating blog posts around EFI Task Views resources
    d. Should we clean up the EFI Task View, create blog posts or do something in between?
        i. Mike's idea was to take [Cran approach](#) - link to packages with bullet point outline with explanations of general areas. Annotated bibliography to tools and packages rather than a vignette
        ii. Task View approach is nice for quick reference for which sets of tools people want to dive into and what might meet people's needs. Then the vignette comes after that.
        iii. Could imagine a short blog that is like an annotated bibliography. Or could make it more vignette
        iv. Task View - is something that is maintained and added to. Alexey suggests creating the Task View with the list of resources and then have blog posts follow-up on that that could focus on certain tools (this would allow us to not need the whole Task View list ready all at once).
        v. Roll out the Task View in chunks.
        vi. **Jake has offered to do this for Workflows. Could lump Pipelines and Reproducibility in there. Mingkai/Rob/Kenton may be able to review or help with this section. Alexey as well.**
        vii. On the Task View document the categories that fit together are:
            1. Interface falls in with Data Vis
            2. Data Ingest, Data Cleaning, Data Management
            3. Uncertainty Quant paired with Stats
        viii. One idea is to have an initial blog post that lays out the 7 categories and how they fit together. The concepts behind the categories. This could be done by Jake and focus on Workflows/Reproducibility
    e. EFI Standards discussion overlaps with this in that if we come up with standards then it will be easier to build/identify re-usable tools

     f. An example of where there is a gap is on the model/Stats side of things where there is a gap between stats tools (R, Python) and Bayes tools (Jag, Bugs, NIMBLE) and between the niche data assimilation (e.g., Dart). It isn't easy to transition between Jags and Dart. Argument for more accessible DA tools that apply for ecological data compared to NOAA's data that they process every day.

     g. Do a targeted gap analysis for the RCN forecasting challenge once the data sources are determined, then in that narrow sense we can make specific identification of small tools (e.g., NEON API R package is or is not suitable for the forecasting challenge, another example is to have a location for code to download/process weather data that lots of people use).

2. Mingkai is new to the group - works with EUCFAC and is thinking about building a platform for the group. Mingkai's background is in modeling. He has constructed C budget. Benchmark all the efforts and benchmark the model skills. This is complementary to the work Mike/Alexey have done with data-model

3. Alexey is on the OakRidge/DAAC working group and another group. In those 2 places he will be pushing heavily on some of these needs that will overlap nicely with components of the Task View

4. Another task to think about is updating the Methods Working Group text on the EFI website

This is a combination of notes from previous calls and notes from the Methods call today on 2-21-20:

- Decide Tasks/Activities for the Methods Group
  - Purpose of the group: The group gets to decide! **What low hanging fruit to get info out for what is available and what is missing to spur development of methods for forecasting and then matchmaking with potential partners**
  - We have the EFI Task View list of resources used for forecasting that could be organized around steps to forecasting
  - From the last RCN steering call -> need for a running list of software needs (overlaps with task view effort)
    - 1st priority - tools to help people get forecasts for NEON data
      - 2nd priority - tools for forecasts more generally
    - Script to create Forecast standards (i.e., Carl's EML creation script)
    - Little tools that we have or wished we had. Bucket list for Hackathon/pre-Hackathon
      - e.g. Download and processing of NOAA GEFS
  - Set of Blog posts revolving around the steps to forecasting Jake laid out in his GLEON presentation:
    - And summarized into these 9 steps here:
  - Jake offered to lead a blog post about Step 3 (see notes above where we fleshed this out more)
  - To support the RCN - have brief tutorials that can be useful for students who are going to participate on the forecasting challenge
    - Put on Slack and Email

- - - Goal to get groups that already have participated. Have you already done stuff or have things that are related to this and will you be willing to help with this effort?
    - We don't have to cover every single step in the forecasting process
  -
- Slack post from Mike 11/18/19:
  - On a recent STC call I mentioned to @Leah Johnson and others that I think we should take a closer look at how the new ensemble variational data assimilation approaches work, compared to things like EnKF. Here's a recent ecological application by Pinnington et al, which includes citations of the original lit in this area: https://www.geosci-model-dev-discuss.net/gmd-2019-60/

    -
  - Also interesting new paper from Jordan Read and collaborators (inc. @Jake Zwart) on process-guided deep learning #papers
  - https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019WR024922
- This may not be a Methods project, but for now, Jody is leaving this in the notes from the Nov 2019 call as a reminder about Elise's idea: for RCN or something down the road - does everyone have their own workflow, or are there other concepts that aren't settled/worked out? Put out a series of podcasts where people do 10-15 minutes on a specific topic. Panel discussion of addressing specific topics
  - Ideas:
  - Coordinate with podcasts that already exist. Freshwater podcast (Making Waves https://freshwater-science.org/education-outreach/making-waves)
  - Major Revisions podcast (https://www.majorrevisionspodcast.com/)
    - Expand the idea of who ecologists are and what they do. Alexey has listened to Major Revisions and knows the host
  - Definition of forecasting in ecology could be a good topic
  - Explain more to the ecological community what we are doing with forecasts
  - Take different parts of the forecasting process and the people doing those (maybe with Partners?) How forecasts are connecting with stakeholders
    - Propose ideas to existing podcasters, not making a new one
  - **VT water work can make a good podcast story**
  - **Think about a podcast that highlights the person who wins the ESA Forecasting Award**

- This came up in the Sept 2019 call, but didn't have time to get to in Nov 2019 call:
  - Mixing of infrastructure and methods is this discussion. Where is a good fit for broad approach discussions such as ensemble modeling vs. trend analysis of big data?

- What is the breakdown between CI and Methods? Task View list has Methods tools, but a number of resources also falls into CI
- Originally, the CI - is more informatics focused, while Methods is more tools focused. But in STC proposal, they were combined into one larger theme
- We could combine Working Groups if want and then if at a later point it makes sense to split, we can.
  - **JP - add to CI agenda to see if we want to merge CI with Methods**

- Here are the notes about the Long-term goals from Sept 26 call:
  - Mike: Intersection between CI/Methods & Tools. Call with EFI & NCEAS about putting a forecasting/ecological modeling, reproducibility mid-scale CI proposal together. Matt Jones asked - what would the stack look like, what technologies? Ask NSF to support NCEAS working group/workshop proposal to develop roadmap/strategic plan statement of what is needed. Build on what we have started with the Task View doc and get into what the technologies are available, what are the unmet needs, and where are the bottlenecks? Have generic workflow (here are the components that go in) and this will allow others to plug into the workflow. Lay out the needs of the community. Be clear about standards and protocols involved.
    - Do workshops at NCEAS. Bring in modeling community, CI community, NSF super computing centers. Think about a long term plan for interoperability
    - Goal would be to produce a white paper/road map for a major infrastructure proposal down the line
  - PEcAN has become large and sprawling. Has become hard for folks outside of PEcAn development team to drop modules in. Need a master plan on how it all works so others can figure out how to make updates they need
  - There doesn't need to be 1 technical solution. But if there is a common ground or enough people are building tools that are interoperable with others.
  - Long term goal **interoperability**
  - PEcAn desires moving forward:
    - What is current landscape of tools available that allows provenance tracking and transparency without imposing that strong connection to the central database
    - Alexey thinks it would be useful to build each piece and component and then bring in provenance at the end instead of starting with the provenance tracking. Pull a few pieces out of a few key places.
    - EFI efforts/RCN is good motivation for doing this
    - Wish list is to get a workflow to get driver data. PEcAn has a lot of that already built that would be useful, but it isn't currently being used independent of PEcAn.

- Thinking of met workflow and using the PEcAn standard for that met workflow. Use the PEcAn standard for a number of different workflows. Standards are key. Start with the right set of Standards.
- Tricky part in Ecology is the long-tail data. Things that don't fall neatly in the space/time grid. With the RCN short term goal NEON has a specific set of variable names/units so on the short term that is what we will go with. Then if we can adapt/enforce the NEON standard then it could make everything easier moving forward.
- **Concrete goal: Forming a subgroup within intersection between Methods/CI interested in spec'-ing out what a general workflow looks like.**
  - Had been thinking of scripts for weather data, but could think about expanding it to grid data (anything that falls within a netcdf format).
  - Expect from May meeting that there will be Terrestrial C, Aquatic, and Population forecasts. Will see if there will be any non-NEON data products that will be needed for those forecasts.
- Moving forward:
  - Jake to wrangle outline for first blog Intro/Workflow. Jody to create Google doc to share with Jake/others to pull together ideas.