July 9, 2019 Cyberinfrastructure Working Group Call

Agenda
• Introductions / record attendance
• What are the short term goals for this working group -- what could be accomplished (or balls that could get rolling) this year and how do we get there?
• What are the longer term goals we are working toward
• Are there goals where a small amount of seed funding from EFI (<$5K) would help move us forward
• Leave with a plan for a next call (either exact date or data range to poll) and hopefully some discrete, assigned next steps.
• Links to EFI2019 Conference Notes: [LINK REMOVED]
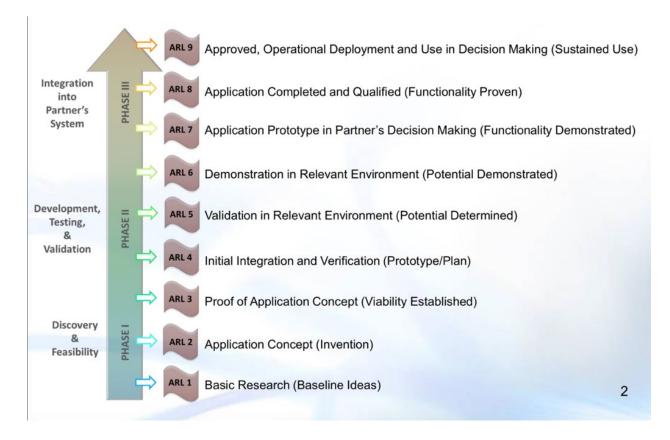
Notes From the July 9 Call
Introductions/Attendance

Mike
Quinn
Elise Zipkin
Carl Boettiger

What's our low-hanging fruit:

- Standards
    - CI should lead, but work with others to meet their needs
    - E.g. with Theory, do we need all the ensemble members or just summary stats
    - EDU group may ask for the ability to see the MODEL in an easy to understand way
        - Could every forecast be a case study that a class could learn from? What would we need to do to get it there
- What do need standards for
    - **Output formats (#1)**
        - Existing standards?
            - schema.org/Dataset (see https://developers.google.com/search/docs/data-types/dataset)
            - EML
            - EDI
    - Model metadata
        - Existing standards / tools?
        - Link to code / code base (what version number / commit tag of the model was used)
    - Metadata about software (e.g. https://codemeta.github.io)
        - Documentation of larger workflow?
    - Metadata about inputs

- Metadata about cross-forecast covariates
    - Taxonomy
    - Scale, resolution, forecast horizon (need to develop list with Theory group)
    - DA techniques used
    - **Uncertainties considered**
        - Mike showed example slide
        - Could be an interesting review paper on the current state of EF
    - **Dynamic or not?**
    - **NOAA/NASA ARL**

	→ ARL 9	Approved, Operational Deployment and Use in Decision Making (Sustained Use)
Integration into Partner's System — PHASE III	→ ARL 8	Application Completed and Qualified (Functionality Proven)
	→ ARL 7	Application Prototype in Partner's Decision Making (Functionality Demonstrated)
	→ ARL 6	Demonstration in Relevant Environment (Potential Demonstrated)
Development, Testing, & Validation — PHASE II	→ ARL 5	Validation in Relevant Environment (Potential Determined)
	→ ARL 4	Initial Integration and Verification (Prototype/Plan)
	→ ARL 3	Proof of Application Concept (Viability Established)
Discovery & Feasibility — PHASE I	→ ARL 2	Application Concept (Invention)
	→ ARL 1	Basic Research (Baseline Ideas)

2

- Examples will also be a critical piece
    - Being able to look at what other people have done
    - Not everyone will do the same thing
    - Having living document
    - Illustrate advantage of that -- able to compare iterative forecasts on the same data
    - Should compare existing way that different have been outputting / saving their forecasts

- - - ■ Organizing a list of folks that are producing, or have produced, forecasts that have some internal standard; evaluate how different they are and what the key things are that people want/need to save?
  - - - ■ Interface as much as possible with existing standards
  - - ○ If we look at our current forecast catalog, we could evaluate some of these aspects (what is the current state of archiving?)
    - - ■ E.g. are the forecasts quantitative (or just pictures of maps or graphs)
    - - ■ Are the forecasts being archived
    - - ■ Is there an identifier other than a paper (URI / DOI)
  - - ○ Is the shortest term goal to assess the state of the forecasts we know about?
  - - ○ Elise: is the goal to be inward or outward?
    - - ■ Bit of both
  - - ○ Some things might really vary by subfield
  - - ○ May want to start with very general standards, very grassroots, can evolve
    - - ■ What would be really easy and basic to get buy in on?
    - - ■ E.g. Actually output the forecast and provide a link to it
  - - ○ Carl: example on Zenodo you can identify a Community; trivial metadata
    - - ■ Can apply that community tag
    - - ■ Moderator can review content and then get approved for that tag
    - - ■ Low bar way to group things
    - - ■ Many of the metadata are field specific
      - - ● Want the output to be similar to the input (e.g. NOAA netCDF vs NEON formats)

When?
  Week of Aug 19th?

Tasks:
- Beer @ ESA: timing relative to Stats section
  - Think about the key metadata that we need (form folks could fill out and populate a table)
  - Will turn people off is too heavy; what's the real minimal set for a forecast to be useful
  - Helps if it's fun!!!
- Examples of output formats
  - Collect from folk
- Application of uncertainty classification matrix to some existing forecasts

- Ask existing members to 'archive a forecast on Zenodo'
- Ask existing members to fill out basic Google Dataset description a la: https://developers.google.com/search/docs/data-types/dataset
- Then ask for their feedback on whether this was easy/hard useful/not?