

# Education and Outreach:

## **The summary notes from the Ed & Outreach group's meeting:**

### **Topic 1: Undergraduate DS education, including pedagogy, project-based learning, and data science ethics**

A number of issues came up around the mere existence of DS programs at the undergraduate level:

- do we need an undergrad DS degree?
- how is it different than computational science in CS or applied statistics in Stat?
- What's the best way to ensure students get experience applying DS to a domain area and how will they learn enough about the domain area?

Where should DS live? And where should the intro to DS live?

- Should intro DS be taught in DS or in courses in domain departments. What will students miss based on where the course lives? Note that intro DS being taught in multiple domain departments not likely work in the smaller college setting.
  - but DS is inherently interdisciplinary which should be built in to any curriculum from the start
- How do we get DS worked into the institutional budget model
  - related questions arose in morning session — seems like some of the institutional tension is about whether DS exists just so that people within the institution can get necessary analysis support, or whether it is a legitimate educational area that needs to have all the same support as any other academic area. This raises the issue of the glacial pace of change within the academy which is completely incompatible with the fast rate of change that is happening in DS, AI, ML etc.

Content of a DS program:

- should there be a move toward development of foundational materials that can be shared
- there are already materials from some schools that can be adapted/adopted (Berkeley intro, Berkeley Data8)

Where do students go?

- seems like many go direct to industry first, then some go to grad school
- which leads to.....

Faculty hiring crisis

- few applicants
- small LA schools are at a huge disadvantage

- increasingly have to hire folks from domain areas, but need them to teach DS and/or CS and be tenurable in DS/CS

The community needs to build resources that would support the range of DS education activities

- a portal of curated datasets for various domains – with emphasis on datasets that adhere to FAIR/CARE principles
- Open source python/Jupyter hub notebooks that use provided datasets, can be incorporated into modules that could be made available to researchers from outside HDR as well, allowing folks to build those materials into courses
- templates for courses
- modules for gen ed courses
- Modules from DSC project that could be made available to broader HDR community (which can then be used to help domain experts get up to speed)
- related course content that faculty can adapt and adopt
- Materials for foundation courses
- Professional development for faculty so they can retool for teaching DS
- use DSC projects as models for workforce training that can be incorporated into the educational setting

## **Topic 2: Sustainable and reproducible data science research, including open source code, collaborative shared community building, and training programs**

In this case, we tended to stray a bit more from the topic. Things discussed were:

1. again, need to build resources that support DS education activities — this could give students direct experience with reproducible data science activities. Need data, curriculum, lessons.
2. question of how to prepare domain experts to do effective data science work without forcing them through complete DS curriculum — need exposure to linux, command line, git, etc.
3. Suggestion about making HDR related REUs easy to find so we can let our students know about them.
4. Suggestion about having time in the next PI meeting for the DSCs to start work on distilling out successes, challenges, etc. with an eye toward assembling materials that would enable other schools to spin up successful experiential learning data science programs at their institutions.
  - a. Also should see materials collected from the various DSCs that involve community college students (either as participants in the projects or as part of pathway to 4-year degree)

REU (Data) Symposium organized by one REU at UMBC:

<https://bigdatareu.umbc.edu/reu-symposium/>

Research:

Data and Cyberinfrastructure:

We should consider taking [nsf.hdr.org](http://nsf.hdr.org) as a potential website to host everything