

# Science Platforms

Thursday 2 PM

# Potential soundbites

Careful:

“Science platforms are the natural evolution of today’s archives for the next decade of large surveys. They will let you bring code to the data for those cases where transferring the data is impractical. They will leverage new technological developments from the fast-paced software industry to increase the scientific return of large survey datasets.”

# Potential soundbites

Ambitious:

“Science platforms will be the “Google Office Suite” for astronomical data analysis—collectively, they will have all of the major survey catalogs, images that cover most of the sky, millions of spectra, and a full-blown computing environment, all at your fingertips. You can also upload your own data and your own software and share whatever you like with your collaborators, and see what others do with their data as well. You could do all your work on your own hardware, but why bother—science platforms will make your work available anytime, anywhere.”

# What science is greatly facilitated by science platforms?

- 1) Joint pixel processing (identified as critical for weak lensing)
- 2) Multi-catalog access (multi-wavelength and multi-measurement)
- 3) Joint pixel-catalog queries (must be fast)
- 4) Monte Carlo simulations
- 5) Support for facilities/experiments that generate so much data that the key to the generation of science is democratizing effective access to the data and its analysis, making it available to as broad a population of researchers as possible
- 6) Large-scale statistical studies (including machine learning etc.)

# What do you get from science platforms?

- 1) Easy access to big computing, just a step away from the interactive environment or (e.g. Dask) integrated with it
- 2) A curated software environment tuned to the users' needs; a well-tested set of tools providing access to a broad range of data (flexibility for users)
- 3) An environment that supports educating users on the tools, and that supports capturing the knowledge acquired by users; software re-use
- 4) An inherently collaborative environment, with storage for intermediate results and the ability to share them
- 5) Richer interfaces to the data
- 6) Access to expertise in handling complex back end systems like big clusters, GPU hardware, database administration

# What do we want from decadal survey?

- Continued support for maintenance and evolution of archives?
- Endorsement of an ambitious vision for science in the Cloud?