(Slides for Tue start here.)

Science with Large Samples

3:30-5:00, Tue Feb 20

Chairs: Knut Olsen & Melissa Graham

Please sit roughly by science interest.
Form small groups of 6-8. ***Assign a scribe.***
Multiple/blended science groups totally OK.

Cosmology (LSS, BAO, lensing)	Galaxies	Extragalactic Transients
Local Group (dwarf galaxies)	Milky Way	Variable Stars
Interstellar Medium	Exoplanets	Solar System

Goal: Create a prioritized list of capabilities to enable science with large surveys.

Agenda:	15 min	discussion of capabilities table*
	25 min	small group brainstorm discussion
	15 min	group reports, fill in capabilities table
	20 min	small group discuss priorities
	15 min	group reports, integrated priorities

*table to be introduced on next slide

Science with Large Samples

Table contents to look like this:

Topic	Science Motivation	Data Sets	Capabilities	Examples	Priority
transients and variable stars	optical transients in NIR-bright hosts; known variable stars	optical source catalogs	fast catalog coordinate cross-matching		medium
cosmology	weak lensing systematics reduction	multi-band survey images	platforms for pixel-level data modeling		
all		Images, catalogs, and spectra	data delivery systems with embedded analysis tools	NOAO DataLab	high

(Slides for Wed start here.)

Breakout: Science with Large Surveys

Charge to Participants:
Organize into small groups by science area, for efficient brainstorming.

Assume that the data you need has been acquired. Create a prioritized list of capabilities to enable science with large surveys.

Highest-Priority Capabilities

Either one group marked as “highest”,
or many groups marked as “high”.

Science	Capability
Cross-disciplinary: optical follow-up of GW events (ground-space diffs); finding optical transients in NIR-bright hosts; galaxy evolution studies and SED measurements; weak lensing shape measurement, deblending, and/or photometry.	Joint pixel-level alignment and analysis for multiple, multi-wavelength, heterogeneous imaging datasets; pixel level data access for spectra (e.g., Gaia); and platforms for pixel-level data modeling.
Cross-disciplinary: optical follow-up of GW events; finding known variable stars; Milky Way and Local Group studies of resolved stellar populations; galaxy evolution; data-mining	Fast catalog coordinate cross-matching (in SQL) with easy probabilistic matching; co-location of astrometry, photometry, spectral data; make code/products sharable.
Transients and variable stars: both to understand events and use as tracers (of MW, or as standard candles).	Development of Alerts broker and target observation manager (TOM); and servers for processing and storage.
Cosmology: e.g., all archival LSS science.	Mechanisms to publicly curate existing data sets (at NOAO/NCOA?), preferably with embedded analysis tools.
Milky Way / Local Group: science with resolved stellar populations (covers many science goals).	Algorithms for global parameter fits from all data, via Deep Learning; include mixed sources, heterogeneous surveys.

Above: approximately ordered by priority based on level of discussion

High-Priority Capabilities

Any group marked as “high”.

Science	Capability
Cross-disciplinary: galactic archaeology; stellar astrophysics, galaxy and star formation physics, SMBHs, stellar feedback; obtaining photo-z’s at the S/N limit; searches for transient signatures in spectra from MOS-type data.	Tools for extracting information from large samples of spectra, spectral model fitting, and extracting information from spectra with low S/N and/or resolution.
Time-domain: known transients; odd/new variables; sudden changes to long-term variables; exoplanets (e.g., in S/LMC in LSST DDF); blazars turning on/off.	Tools for monitoring of alert stream/long time series to find variables with unexpected changes in behaviour (light-curve classification algorithms).
Time-domain: optical follow-up of GW events; known variable stars.	Fast photometry on image cutouts and coordinate-based comparisons with existing catalogs.
Milky Way / Local Group: galactic archaeology; stellar astrophysics, galaxy and star formation physics, SMBHs, stellar feedback.	Algorithms for optimized star/galaxy classification and crowded field photometry.

Above: approximately ordered by priority based on level of discussion

Any group marked as “high”.

High-Priority Capabilities

Science	Capability
All	Education: for "pipeline" software development; for training students/researchers in effective machine learning techniques and other analysis/computing methods for large datasets; cross-disciplinary workshops for new ideas, best practices
All	Software and Computation: community access to Cloud computing; support for scale up of existing astronomy-driven tools; policies for recognition/credit for software dev contributions; compression methods for information preservation when storing raw/processed data products (e.g., Chad Schafer's presentation)
Cosmology Theory: combining large-scale structure probes	Algorithm development for joint likelihood and/or bias posterior extraction
Cross-disciplinary: Milky Way science and cosmology	Explicit overlap for imaging/spectroscopic surveys.

Above: approximately ordered by priority based on level of discussion

Medium-Priority Capabilities

Any group marked as “medium”.

Science	Capability
Time-domain: target prioritization for further follow-up of transients, GW events optical counterparts.	Automatic, rapid reduction pipelines for commonly-used spectral setups.
Milky Way / Local Group: science with resolved stellar populations (covers many science goals).	Techniques for the deblending of spectra (using imaging when it helps).
Milky Way / Local Group: science with resolved stellar populations (covers many science goals).	Spectrophotometric calibration.
Time-domain: light-curve classification, e.g. GRBs, for fast spectroscopic follow-up.	Data science platforms with built-in machine learning.
Cosmology: cross-correlation calibration of photo-zs.	Efficient combining and correlation techniques for multi-wavelength and training catalogs.
Local Group: detecting low surface brightness features in local dwarf galaxies.	Algorithms for custom image stacking, differencing and processing to extract low-surface features; creation of residual images.
Milky Way / Local Group	Forced photometry on original pixel data

Above: all related to “data analysis”.

Medium-Priority Capabilities

Any group marked as "medium".

Science	Capability
Milky Way / Local Group	Artificial star pipelines
Milky Way / Local Group	Tunable galaxy model predictions
Galaxies	Techniques for forward modeling galaxy properties
Milky Way / Local Group	Simulations of comparable size & complexity to the new datasets
Cosmology: theory modeling of small scales	Efficient hydro codes and sub-grid scale physics codes, mock observation generation (galaxy populations, for example)

Above: all related to "modeling".

All	Statistical techniques: error determination (systematics and random)
All	Selection function tracking
All	Crowd-source friendly environment
All	Ability to host and serve user-derived published data products (i.e., replace online journal article tables for really large samples)

Above: "other medium-priority capabilities".

Links to the brainstorming organizational sheets

Original spreadsheet of science goals and needed capabilities (view only):
https://docs.google.com/spreadsheets/d/1NoVsJyM5EZ-YOKreTFbMDMnFisnXEubEaZKiASa7wpY/edit?usp=sharing

Organizational spreadsheet with priority-sorted list of amalgamated capabilities:
https://docs.google.com/spreadsheets/d/1ajJ4ASws6VD6Pswz3dpIHLI_MFuz2sAtcTy8cM0DEEg/edit?usp=sharing