Report to AAAC on Euclid/LSST/WFIRST Coordination

Jason Rhodes (JPL)

January 26, 2017

Starting point: Summary Slide from Steve Kahn Presentation to AAAC 1 year ago

* LSST, Euclid, and WFIRST are all moving forward toward the onset of operations on similar timescales.

* While there is strong overlap in the science planned for these three separate facilities, their designs are highly complementary.

* A combined analysis of the data from all three will provide a significant enhancement in scientific return. For reduction of systematics, this will probably require joint processing at the pixel level.

* That form of joint analysis is outside the current scope of all three projects in the US. It will thus require some additional funding.

* We have formed a Tri-Agency, Tri-Project Working Group to explore this. The initial reports from the technical subgroups will occur this Spring.

What is the TAG?

Euclid: ESA-led with NASA and Canadian contribution
LSST: DOE and NSF funded, plus international partners
WFIRST: NASA with potential international partners

- Tri-Agency: DOE/NASA/NSF
- Tri-Project: Euclid/LSST/WFIRST
- Meet to discuss commonalities, coordination, optimization of data, simulations, software
- Three aspects to TAG:
  1. the agency program managers talking to each other, often as part of wider program communication
  2. the agencies plus project leads (what we usually mean by TAG)
  3. the project leads gaining input from their collaborations, e.g. informal “task forces”
- Informal group started in ~2012
- Discuss coordination on these three projects, primarily for dark energy
  - Other science areas are considered, but haven’t driven discussions
- Telecons every ~2 months, in person meetings as needed (~yearly)
- Facilitates open communication about issues that affect all the projects and require joint or coordinated efforts of the agencies

Who is the TAG

- HQ Reps
  - Eric Linder and Kathy Turner (DOE)
  - Dominic Benford and Linda Sparke(NASA)
  - Nigel Sharp (NSF)
- Project Reps
  - Jason Rhodes (Euclid, also a WFIRST Deputy Project Scientist and on LSST)
  - Steve Kahn (LSST Project, also on Euclid)
  - Rachel Bean (LSST Dark Energy Science Collaboration/DESC Spokesperson, also on Euclid and WFIRST)
  - Neil Gehrels (WFIRST Project Scientist)
  - David Spergel (WFIRST Adjutant Scientist, also on Euclid and LSST); added in 2016

Philosophies

- The best dark energy constraints in ~2030 will come from a combination of Euclid/LSST/WFIRST data
- The teams that are processing those data best understand the data and should be involved in that combination
- We should maintain US leadership or co-leadership in dark energy science using these combined data sets
- Coordinating the cadences and survey footprints can increase science output
- The optimal combination of the data sets may need to be done at the 'pixel' rather than 'catalog' level
  - This requires coordination and goes beyond the scope of the projects
  - This needs to be done with forethought in order to be done correctly
- Simulations of the universe are computationally intensive and should be used by multiple groups where possible

LSST/Euclid coordination

TAG-informed activity, but done at project level

- Requires international agreements and careful examination of data rights
- Requires agreement of Euclid Consortium, LSST Project and LSST DESC
- Previous attempts at coordination have not yet reached an agreement on what would be shared, how it would be shared, and when it would be shared
  - Both projects have data they do not want to share immediately without safeguarding their core science
- In 2015, the TAG endorsed an effort at defining the scientific and technical benefits of coordination (see next slide)
- Euclid Consortium, LSST DESC, LSST Project produced a joint charge that led to an in-person meeting of key scientists in July 2015

LSST/Euclid White paper

TAG-informed activity, but done at project level

- One day meeting organized by Rhodes in July 2016 in Oxford, UK
- ~20 participants from Euclid and LSST (many in both)
- Agreed to write a white paper that did not concern itself with the politics of data sharing or cadence coordination
- White paper has progressed in fits and starts since July, but expect a compete draft by end of February
- Plan to publish white paper in a refereed journal for use as a guide for future MOU discussions. Outline:
  1. Conceptual Rationale
  2. Benefits for Cosmology
  3. Benefits for Other Science (solar system, galaxy clusters, etc)
  4. Technical aspects of Coordination (shear, photometric redshifts, etc)
  5. Implementation Plan
  6. Gap Analysis/Resources Needed
  7. Summary

Attendees:
Mario Juric
Andy Connolly
Steve Kahn
Ian Shipsey
Robert Lupton
Jean Charles Cuillandre
Marc Sauvage
Bob Nichol
Tom Kitching
Rachel Bean
Anja von der Linden
Peter Capak
Richard Massey
Jason Rhodes
Andy Taylor
Will Percival
Dominique Boutigny
Eric Aubourg
Hendrik Hildebrandt
Rachel Mandelbaum
Jeff Newman

LSST/WFIRST Coordination

TAG-informed activity, but done at project level

- Primarily requires agency and inter-team discussions
- WFIRST data has no proprietary period; simplifies data rights issues
- First steps of coordination are cosmology-based but other science areas considered
- WFIRST (coordinated by Rhodes) contributed 3 sections (wide survey, supernova, microlensing) to LSST observing strategy white paper*
- 3 day workshop organized by Doré (relevant WFIRST Science Investigation Team lead), Bean, Kahn, Rhodes in Pasadena in September

*https://github.com/LSSTScienceCollaborations/ObservingStrategy

WFIRST/LSST Workshop

TAG-informed activity, but done at project level

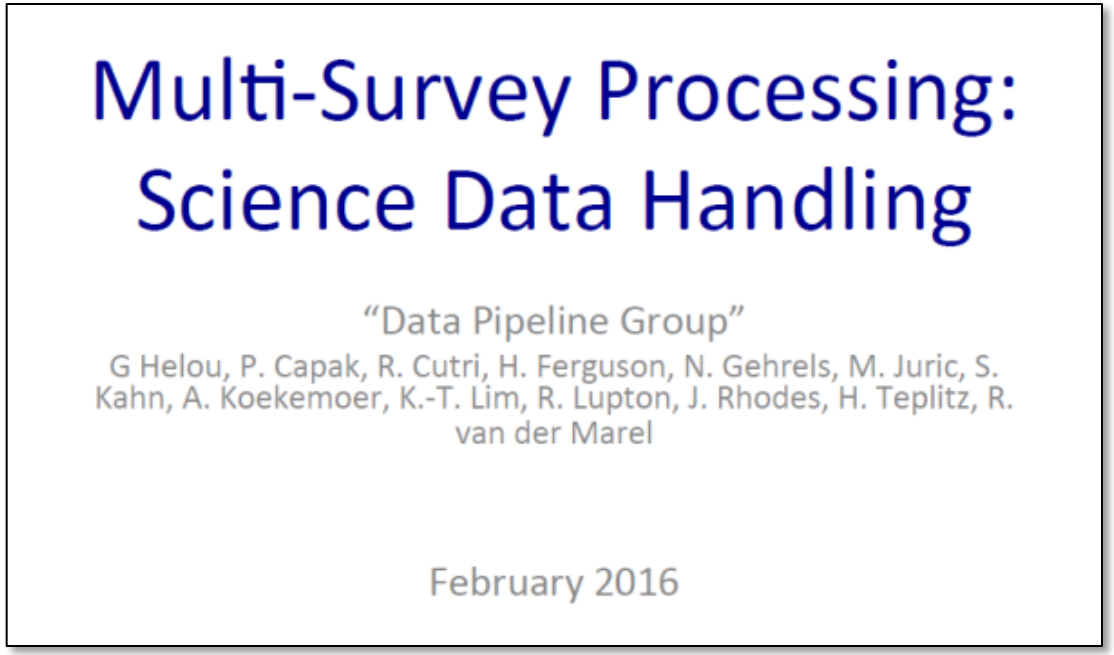
- Workshop in September 2016; Science and data processing synergies discussed
- Cosmology focus, but future meetings will branch out further
- Can be a guide for future collaborative meetings
- Doré plans to start a white paper in March based on the ideas developed at the workshop

- First in a series?

Registrants:
Lee Armus (IPAC)
Rachel Bean (Cornell)
Andrew Benson (Carnegie)
Phil Bull (Caltech)
Peter Capak (IPAC)
Tzu-Ching Chang (ASIAA/JPL)
Andy Connolly (Washington)
Roc Cutri (IPAC)
Olivier Doré (JPL/Caltech)
Richard Dubois (Stanford)
Tim Eifler (JPL)
Harry Ferguson (STScI)
Ryan Foley (UC Santa Cruz)
Neil Gehrels (GSFC)
Daniel Gruen (Stanford)
Salman Habib (Argonne)
Katrin Heitmann (Argonne)
George Helou (IPAC)
Chris Hirata (Ohio State)
Shirley Ho (LBL)
Eric Huff (JPL)
Bhuvnesh Jain (Penn)
Steve Kahn (SLAC)
Alina Kiessling (JPL)
Anton Koekemoer (STScI)
Elisabeth Krause (SLAC)
Rachel Mandelbaum (CMU)
Dan Masters (IPAC)
Peter Melchior (Princeton)
Alex Merson (JPL)
Hironao Myatake (JPL)
Jeff Newman (Pittsburgh)
Peter Nugent (LBNL)
Saul Perlmutter (UC Berkeley)
Andres Plazas-Malagon (JPL)
Graca Rocha (JPL)
Jason Rhodes (JPL)
Michael Schneider (LLNL)
Sam Schmidt (Davis)
Melanie Simet (JPL)
Masahiro Takada (IPMU)
Michael Troxel (OSU)
Anja Von den Linden (Stony Brook)
Yun Wang (IPAC)
Risa Wechsler (Stanford)
Hao-Yi Wu (Caltech)
Joe Zuntz (Manchester)

Joint Processing

- TAG commissioned a report on how joint pixel-level data processing could be done in the US and what work is needed to make that happen
- George Helou (IPAC) coordinated the effort and produced a PPT report in February 2016
- Defined a 4-stage approach to joint processing to maximize science return
- Report recommended the next step be conducting a scoping study of work to be done to fulfill joint processing



Joint Processing Summary

Summary (1)

- Joint analysis will target specific science goals. A common-use data legacy from co-processing gets to the science faster and amplifies pay-off well beyond targeted science papers
  - Well-designed, well-documented public-release products are best use of limited resources
  - They attract new users, enable more science
  - Science center expertise can help generate and preserve them efficiently, along with tools, docs, ancillary data, simulations, etc

Summary (2)

- This plan is agnostic as to who does what
  - Existing expertise at centers and projects is more than adequate
  - "Processing tiers" are one example of resource usage tuned for maximizing the science
  - Centers and projects do lower tiers, community competes for science exploitation
- U.S. agencies and community should be aware of similar planning in Europe, leave open a path to collaboration, but avoid being left behind

Joint Processing: 4 Stages

Long-Term Program: Overview

- Phase 1 [05/2015-02/2016] Scouting the terrain & preparing this report
- Phase 2 [~2 years] Defining (a)_Requirements, (b)_Algorithms and (c)_Architecture
- Phase 3 [~3 years] Developing Software and Systems; Preparatory Community Research
- Phase 4 [~2021-TBD] Conducting Survey Operations; Community Research

Phase 1 Complete. Projects ready to start Phase 2 in 2017

Joint Simulation Efforts

- TAG commissioned a report on cosmological numerical simulations and the required computing resources
- Alina Kiessling (JPL) coordinated the group that wrote the report
- Report to TAG in March, 2016

Tri-Agency, Tri-Project Task Force: Simulations

Julian Borrill (LBNL), Andy Connolly (UW), Salman Habib (ANL), Alina Kiessling (JPL/Caltech), Rachel Mandelbaum (CMU), Peter Nugent (LBNL), Michael Schneider (LLNL)

Date: 03/15/16

1 Charge

Investigate the simulation requirements (including cosmological, image and instrument simulations) for the Large Synoptic Survey Telescope (LSST[1]), Euclid[2] and the Wide-Field Infrared Survey Telescope (WFIRST[3]) – see Table 1 for a high-level overview of the surveys. Provide guidance on where investments should be made to ensure the three projects have timely access to the required simulations. Pay particular attention to the areas in which the three projects could share resources (computing and personnel) and simulations.

Simulation Report Key Findings/Recommendations

- Sharing simulations makes sense (it's the same Universe)
- Sharing computing resources makes sense- High Performance Computing (HPC) infrastructure is expensive
- Open access to simulations makes scientific sense
- Data sharing infrastructure needs work (moving large amounts of data is challenging)
- Common data formats help
- Covariance matrix calculation may be computationally challenging
- Modeling of baryonic physics needs work
- Recommend setting up Tri Agency Cosmological Simulations (TACS) Task Force to explore options for joint simulations, HPC resource coordination, data sharing

Simulation Data Sharing Opportunity (1)

- Recently, a data sharing experiment was carried out and showcased at the International Conference for High Performance Computing, Networking, Storage and Analysis. Data from a large simulation run at one supercomputing facility was moved to a different facility for longer term hosting and storage.
- They obtained transfer speeds of ~0.5PB+ per day between the Argonne Leadership Compute Facility (ALCF) and the National Center for Supercomputing Applications (NCSA, in Illinois) without doing anything special other than optimizing file sizes with the ~100GB transfer link. Note that both facilities already have very good hardware for data transfer purposes.
- Large simulations are being produced by a number of groups involved in LSST, WFIRST, and Euclid. The facilities that run the simulations are not necessarily the best facilities to host and serve the Level 1 (particle data snapshots) and Level 2 (e.g., lightcone and mock catalog) data.

Simulation Data Sharing Opportunity (2)

- Currently, there are also difficulties with long-term storage of very large datasets because appropriate policies with supercomputing facilities are not in place.
- In order to ensure that valuable data are not being deleted, it would be beneficial to have a facility that can both store and host these Level 1 and 2 cosmological simulation data, including staff who could set up and maintain an efficient database for serving the data.
- Kiessling and Heitmann have played a key role in initiating discussions of a possible arrangement where simulations are run using predominantly DOE supercomputing facilities and are then stored and hosted at a NASA facility. VERY PRELIMINARY.
- The goal would be for all simulations hosted by this facility to be fully public.

Tri Agency Cosmological Simulations (TACS) Task Force – Starting Soon

- Projects (Rhodes, Kahn, Gehrels) have asked Alina Kiessling (JPL) and Katrin Heitmann (Argonne) to draft a charge for the TACS and recommend a TACS advisory board and task force members
- Draft charge produced, will be sent to the TAG next week
- Focus on:
  - Common infrastructure to share simulation products
  - Base cosmological simulations (info from projects)
  - Investigation of systematic effects
  - Large simulation campaigns

Conclusions

- TAG is providing excellent inter-agency and inter-project communications
- Dark energy cosmology is driving current activities but other science areas may benefit and drive future activities
- LSST/Euclid and LSST/WFIRST coordination meetings have engaged the community
- Long term joint processing plans have been laid out
- Sharing of simulations and high performance computing resources will be explored in the near future

Additional slides