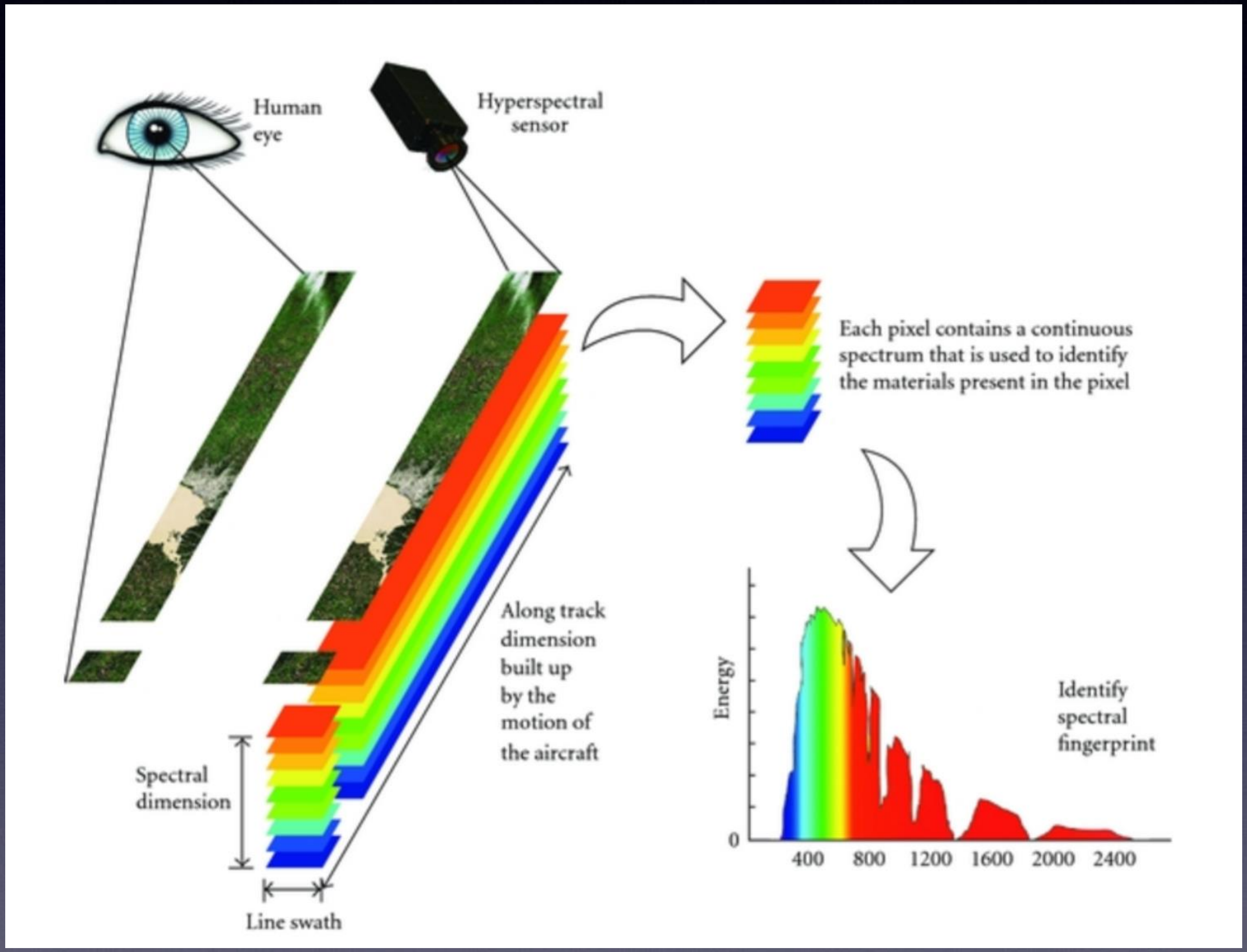
Hyperspectral Image Analysis via a Functional Data Model

Anne Wilson
Odele Coddington
Peter Pilewski
Doug Lindholm
The LASP Web Team

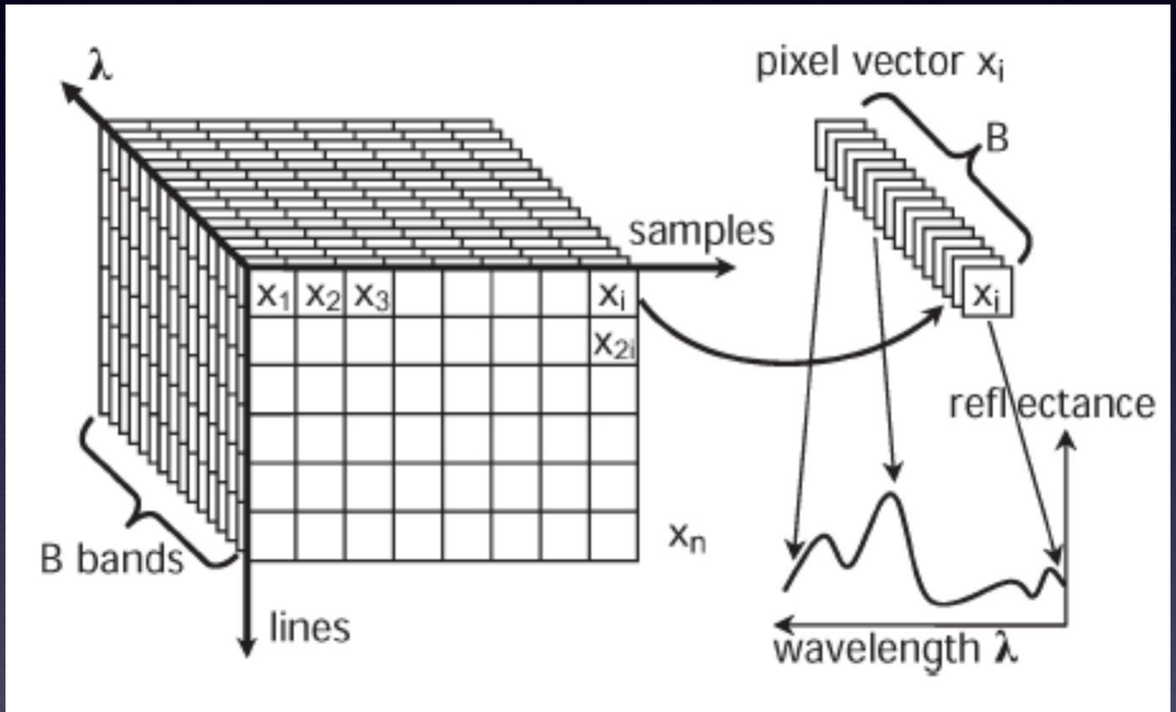
Laboratory for Atmospheric and Space Physics (LASP)
University of Colorado, Boulder

Presentation to ESTF 2018, June 14, 2018

Hyperspectral cubes



Spectral Distortion in Loss Compression of Hyperspectral Data, Aiazzi, et al, Journal of Electrical and Computer Engineering, 2012(5):1817-1819, Aug 2003, DOI: 10.1109/IGARSS.2003.1294260



Bodkin Engineering, http://www.bodkindesign.com/products-page/hyperspectral-imaging/hyperspectral-imaging/

Continuous coverage of an area

Multi versus hyperspectral imagery

- Multi: lower resolution band coverage
- Hyper: continuous resolution with much higher spectral resolution
  - much higher information content
  - ability to identify unique components
  - more suitable for machine learning algorithms
- The shape of the two are the same
  - We treat them the same

Hyperspectral imagery drivers

- Improving instrumentation provides increasing resolution
- Growing commercial development of imagers
- Increased flying options, including UAVs, allow even greater deployment
- Cloud computing, commodity parallelization, big data engines

Laboratory for Atmospheric and Space Physics

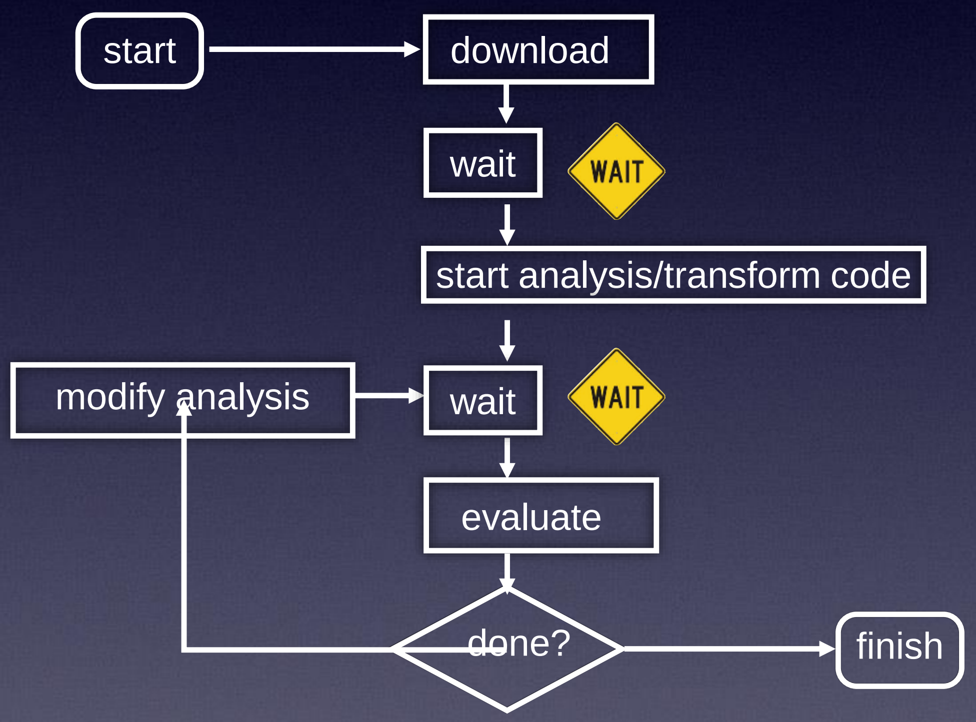
Hyperspectral imagery applications

- Solar variability, effects on climate
- Cloud property retrievals
- Agriculture, agriscience
- Biodiversity, ecology, environmental monitoring
- Food processing, safety
- Medical imaging, biotechnology
- Drug identification, counterfeit and foreign material detection
- Reconnaissance, surveillance

Laboratory for Atmospheric and Space Physics

Multi and Hyperspectral Imagery Challenges

- Huge data volumes, cost to manage
    - 100TB for an analysis
- Complex preprocessing
- Coordinate system transformations



A typical workflow (assuming space is not an issue)

Scientists want to

- Match data from different sources by time, geolocation,
- Select pieces of spectral imagery by time, location, wavelength,
- Apply analysis algorithms to the data,
- Create new data products that are the ‘fusion’ of multiple data sources,
- Work with both multi and hyperspectral datasets
    - cross calibration, find common areas, etc.
- … in a responsive, performant fashion.

Laboratory for Atmospheric and Space Physics

Hylatis: a platform for hyperspectral image analysis

- Toolset for multispectral and hyperspectral datasets in the cloud
- Data model and framework to support reusable code, dataset integration, and interoperability
- Most generally, a domain agnostic platform for science data representation and analysis, that
  - Any domain can write to
  - Supports operations such as transforms in a structured, organized way, rather than writing one off code for each transformation needed

NASA

LASP
Laboratory for Atmospheric and Space Physics

Hylatis: a software research project

- How to structure software to support modeling and analysis of spectral data
- Will not produce new data products
- Will not stand up a repository of hyperspectral imagery
- Building a software layer in front of hyperspectral data
    - could be used to support a repository

Core project principles

- Commitment to principles of mathematics to model and operate on data
- Use of functional programming style, whose benefits include:
  - More thoughtful, rigorous, up front development produces code that is more correct, more easily parallelizable
  - Compiler can check reasoning
  - Capture of generalizations without domain specificity, enabling code reuse
  - Composition of functions
- Modeling data mathematically, for flexibility in representation and thus easier interoperability of disparate datasets

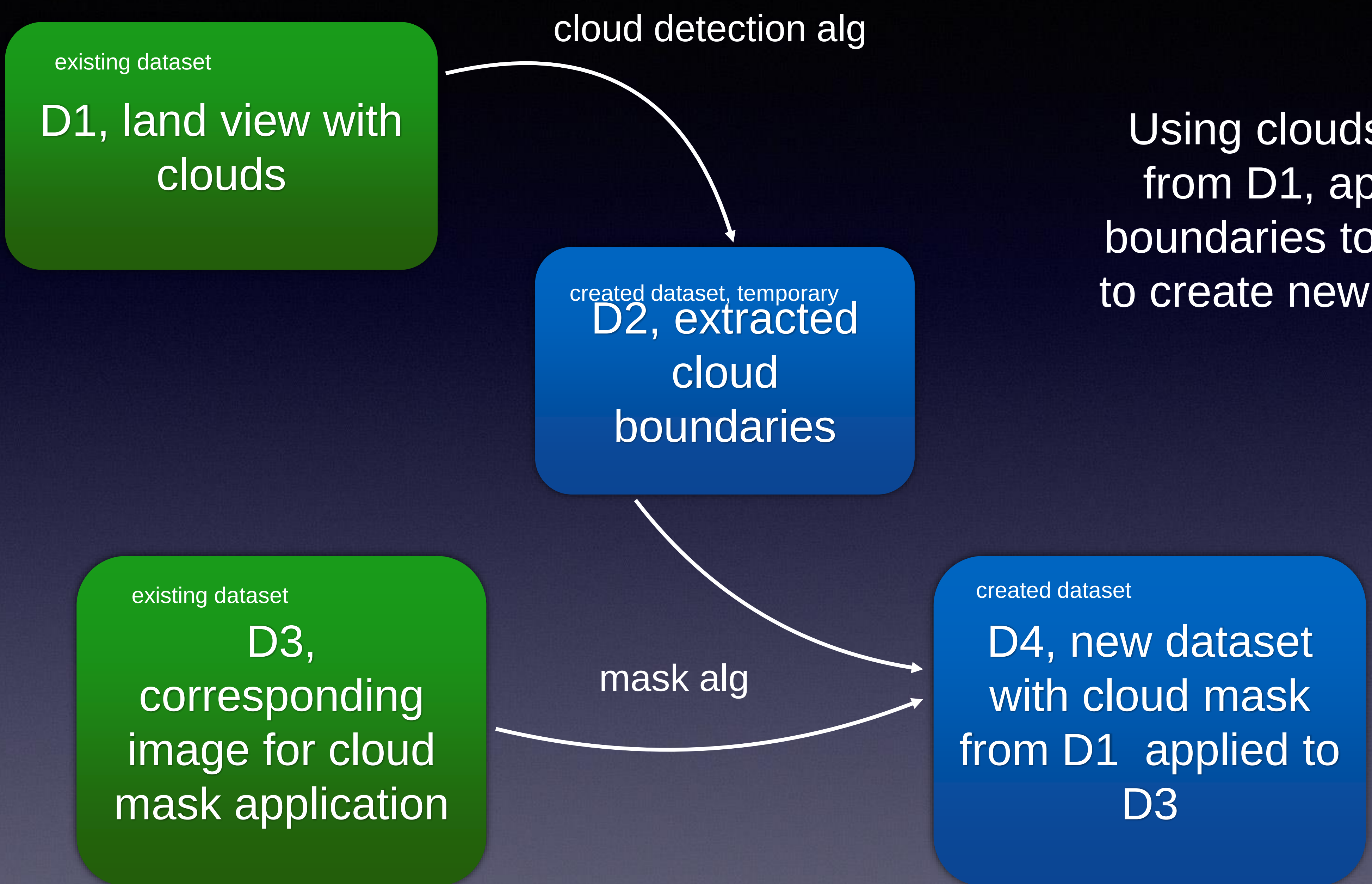
Laboratory for Atmospheric and Space Physics

Using math at the foundational level

- Data model is simply a mathematical function of independent and dependent variables
  - no domain semantics
  - temperature = f(time, latitude, longitude)
  - same as: x = f(a, b, c)
    - interpreter uses pattern matching

Laboratory for Atmospheric and Space Physics

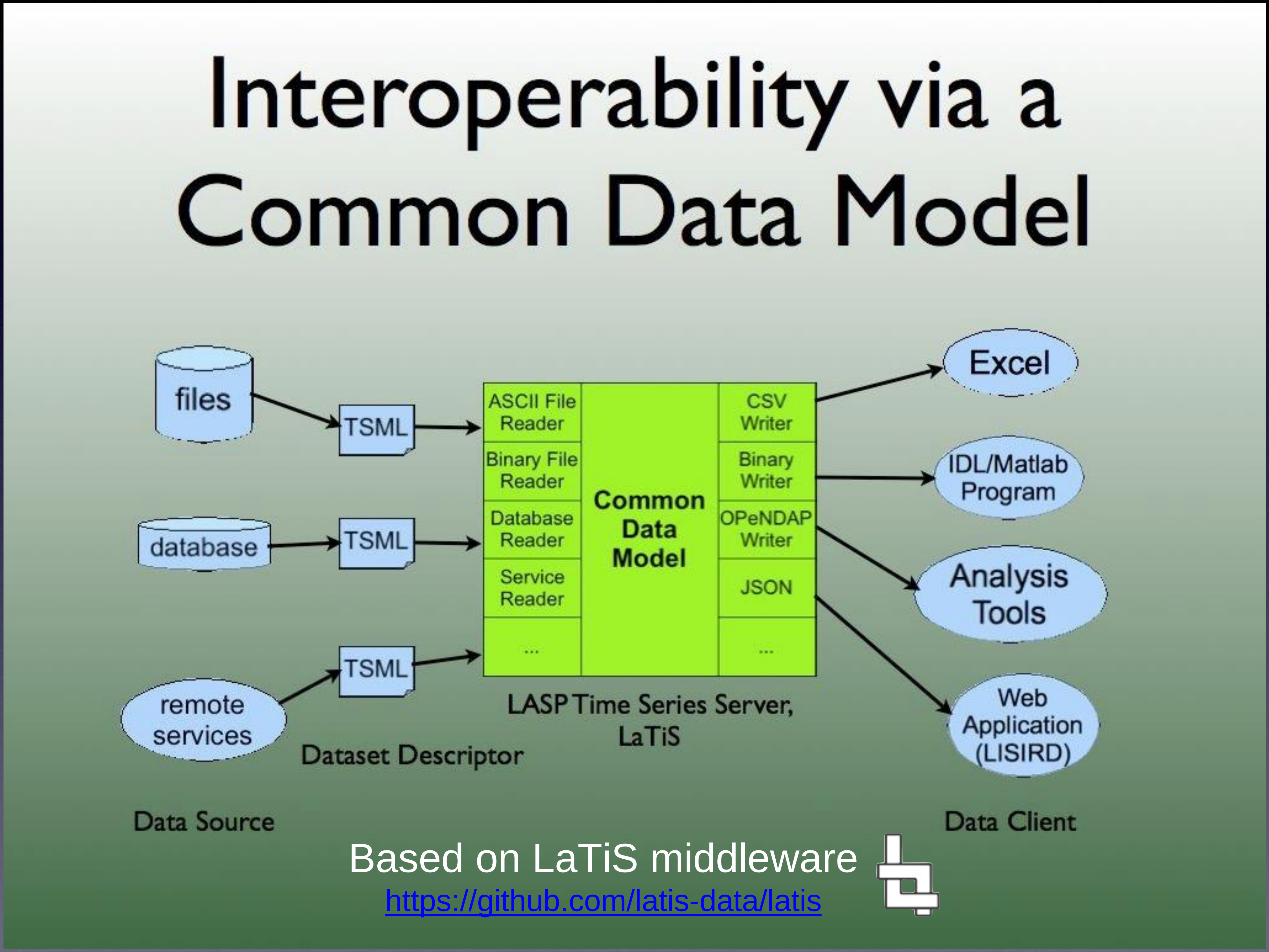
Hypothetical composition problem



Using clouds detected from D1, apply cloud boundaries to image D3, to create new dataset D4

D4 = mask((cloud_detection(D1)), D3)

Hylatis architecture



LaTiS supports:

- code reuse
- server side computations
- e.g., on demand reformatting

Initial Hylatis task

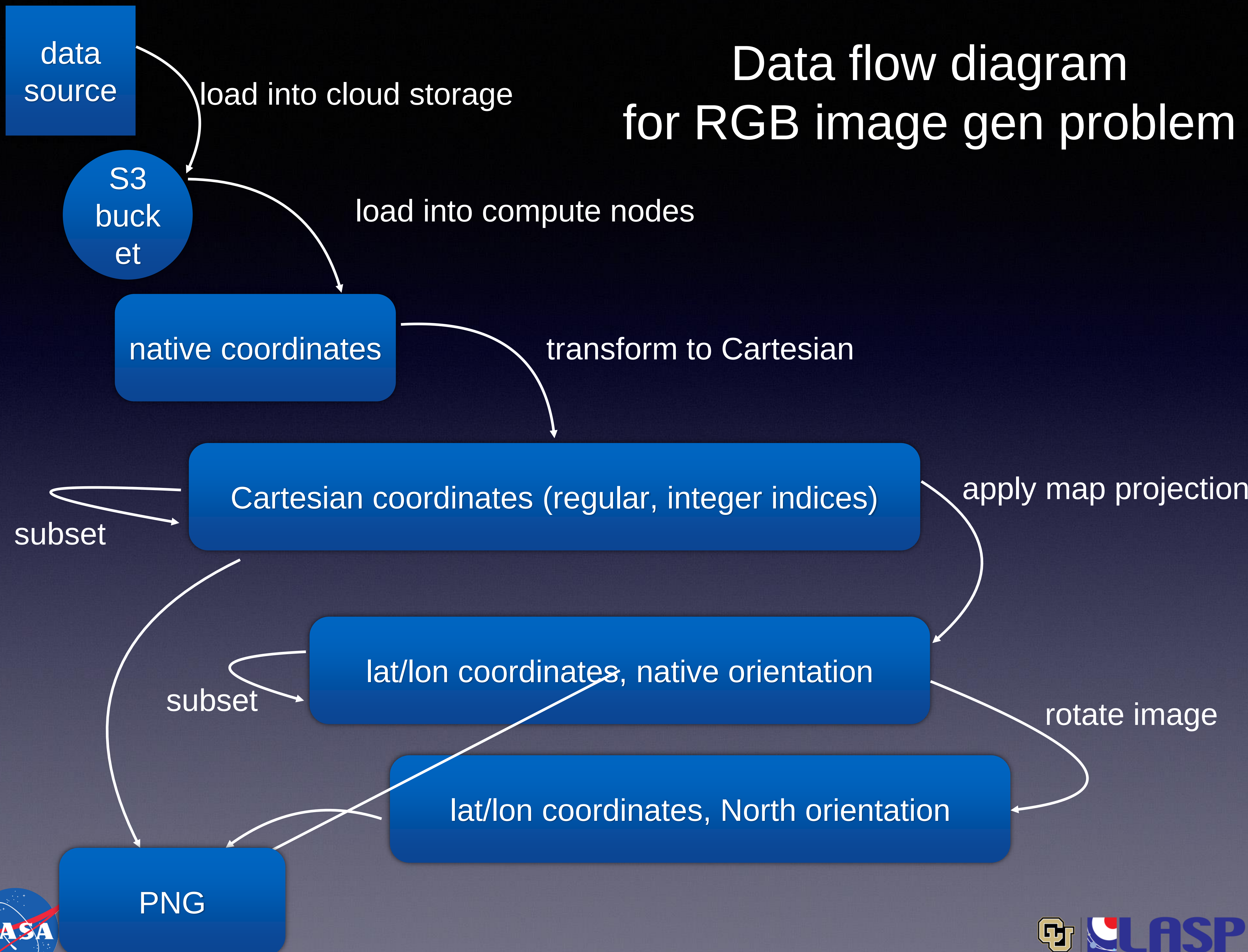
Subset a hyperspectral cube by geolocation,

select 3 bands,

display them as RGB image.

Laboratory for Atmospheric and Space Physics

Data flow diagram for RGB image gen problem



Laboratory for Atmospheric and Space Physics

Selected datasets

1. HySICS: LASP hyperspectral instrument to fly on CLARREO
   - using calibration data from a balloon flight
   - each image is 480 x 640, and cube has 4200 images
2. GOES-R: multispectral, 16 bands
   - 1 - 6 measure outgoing radiance at the top of the atmosphere
   - 7 - 16 are digital maps of outgoing radiance values at the top of the atmosphere
3. MODIS: multispectral, 36 bands
   - Level 1b, 1km spatial resolution, 5 minute temporal resolution
4. POLDER: includes polarization data, 242 x 548
   - an extra dimension in the shape of the data

Laboratory for Atmospheric and Space Physics

Specific project term goals

- Load 4 datasets into cloud: HySICS, GOES, MODIS, POLDER
- Via web page, allow simultaneous users to select datasets, subset on geolocation, bandwidth, pixel
- Generate RGB images, for browsing purposes
- Fuse datasets into new products using simple interpolation

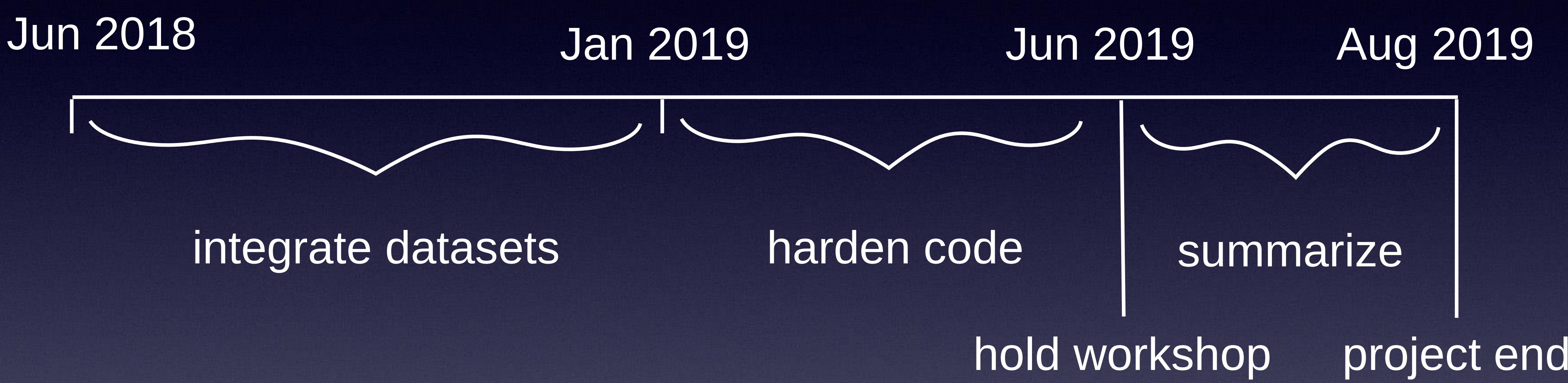
Laboratory for Atmospheric and Space Physics

Demonstrating…

- Handling different multi and hyperspectral datasets in the cloud via unified API and generalizable framework
- Multiple users sharing very large datasets
- Taking computation to the cloud

Laboratory for Atmospheric and Space Physics

Project timeline



Workshop for invited community members, June 2019

- Steer, harden development
- Get community feedback

Thank you!

Laboratory for Atmospheric and Space Physics

Extras

General Science Drivers

Dr. Peter Pilewski studies solar spectral variability and effects on cli

1. Retrievals of geophysical variables, for example, cloud and aerosol micro-physical properties. This is highest priority, making most efficient use of contemporaneous measurements, finding overlap in sampling volume and time.

2. Inter-calibration: transfer of calibration from high-accuracy sensor to operational sensor (this is related to CLARREO).

3. Model testing. For example, testing climate model output, developing and testing Climate Observing System Simulation Experiments (OSSEs)."

Odele's science application

Dr. Odele Coddington, studies cloud phase transitions

"I … plan to use the fused GOES-R, MODIS, and POLDER datasets to evaluate the added information content from the fusion of polarized cloud reflectances (i.e. POLDER) with non-polarized cloud reflectances (GOES, MODIS) for cloud thermodynamic phase."

Hylatis tools applied to Odele's problem

Currently using MODIS bands as proxy for PACE

Currently using POLDER data as proxy for PACE polarimeters

- Subsetting capabilities needed for these

When PACE is in orbit

- Subset PACE to get 2 micron channel, and others
- Subset PACE to get polarization data
- Fuse the above

RGB image generation of above as part of workflow

(Future: Generate RGB images that outline dataset coverages, so overlap

Laboratory for Atmospheric and Space Physics

Use of functional programming in industry

- Amgen
- AT&T
- Bank of America Merrill Lynch
- Barclays Capital Quantitative Analysis Group
- Bloomberg
- Credit Suisse
- Facebook
- Google
- Intel
- Microsoft
- MITRE
- New York Times
- NVIDIA
- Qualcomm
- Twitter
- Walmart
- Bloomberg

Laboratory for Atmospheric and Space Physics

Functional programming in the financial industry, why?

- Lexical analysis and parsing over rich data
- Easy to code AI rules
  - immutability reduces error rates
- Scaling across compute units
  - lack of state information eases parallelization
- Prototyping of complex algorithms over large data volumes