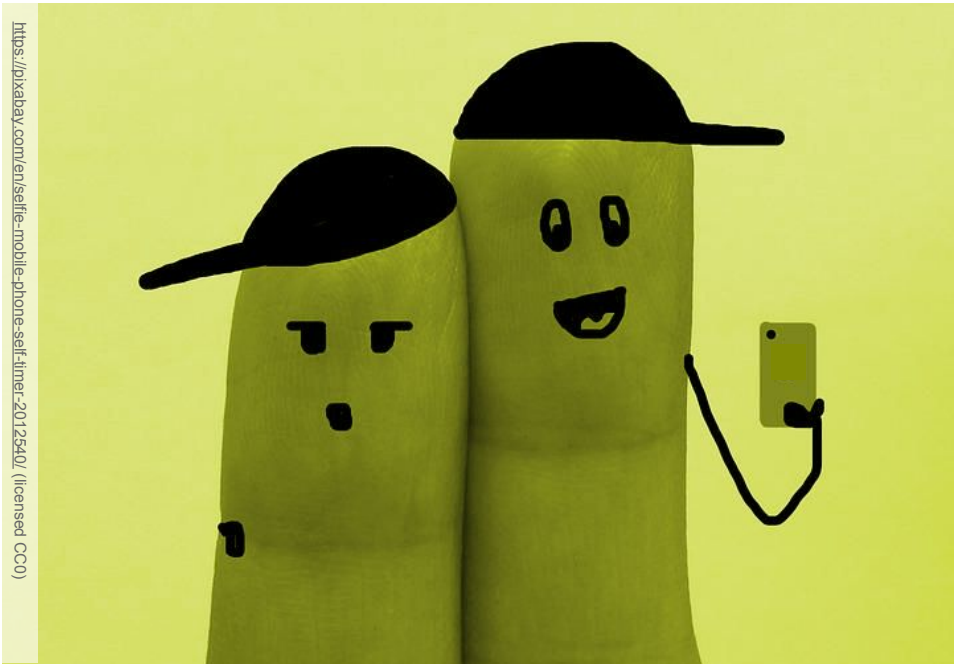
Embracing Open Source for NASA's Earth Science Data Systems

Katie Baynes, Dan Pilone, Ryan Boller, David Meyer, NASA Goddard Space Flight Center, Greenbelt, MD
Kevin Murphy, NASA HQ, Washington, DC

First of all, thanks for having me. I am so excited to be here!

Look guys! I am on stage!



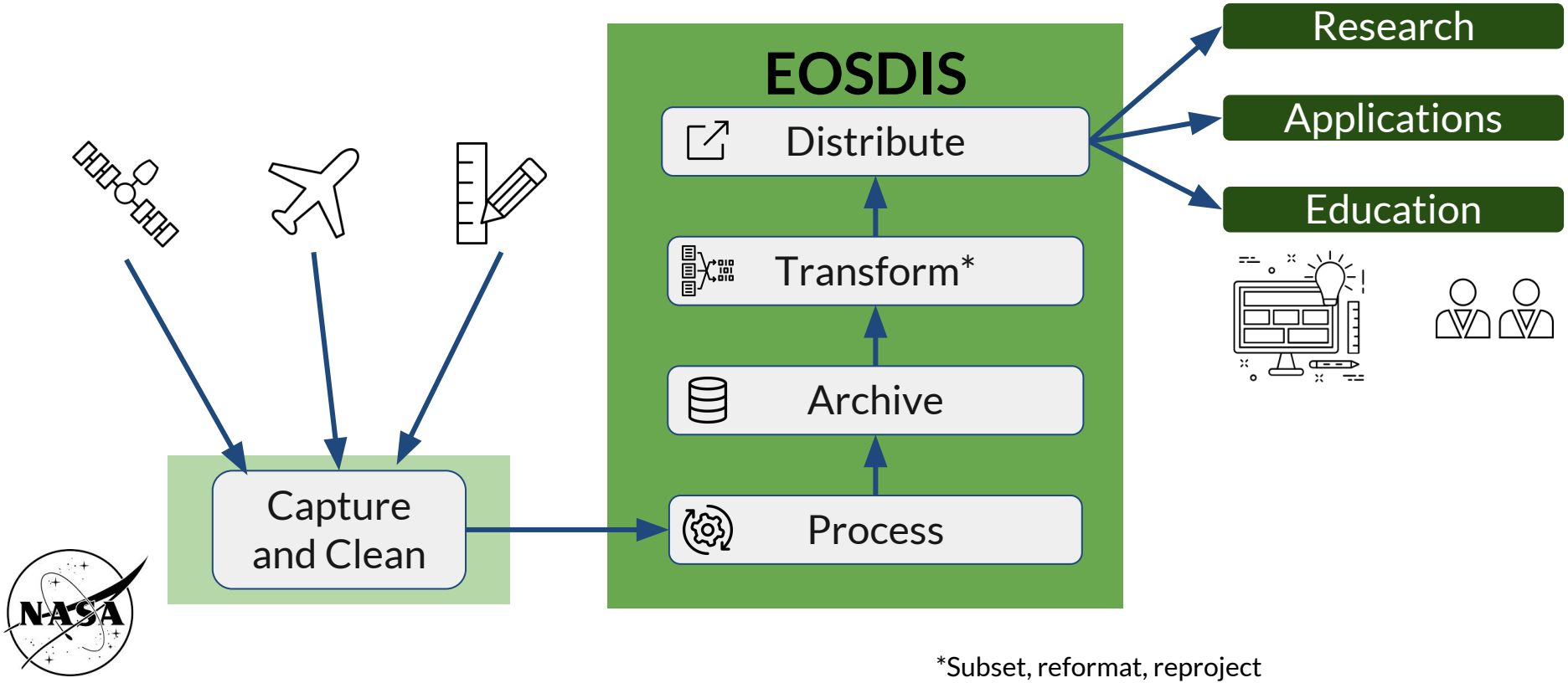
https://pixabay.com/en/selfie-mobile-phone-self-timer-2012540/ (licensed CC0)

First, an EOSDIS* Overview

(so where do you work at NASA?)

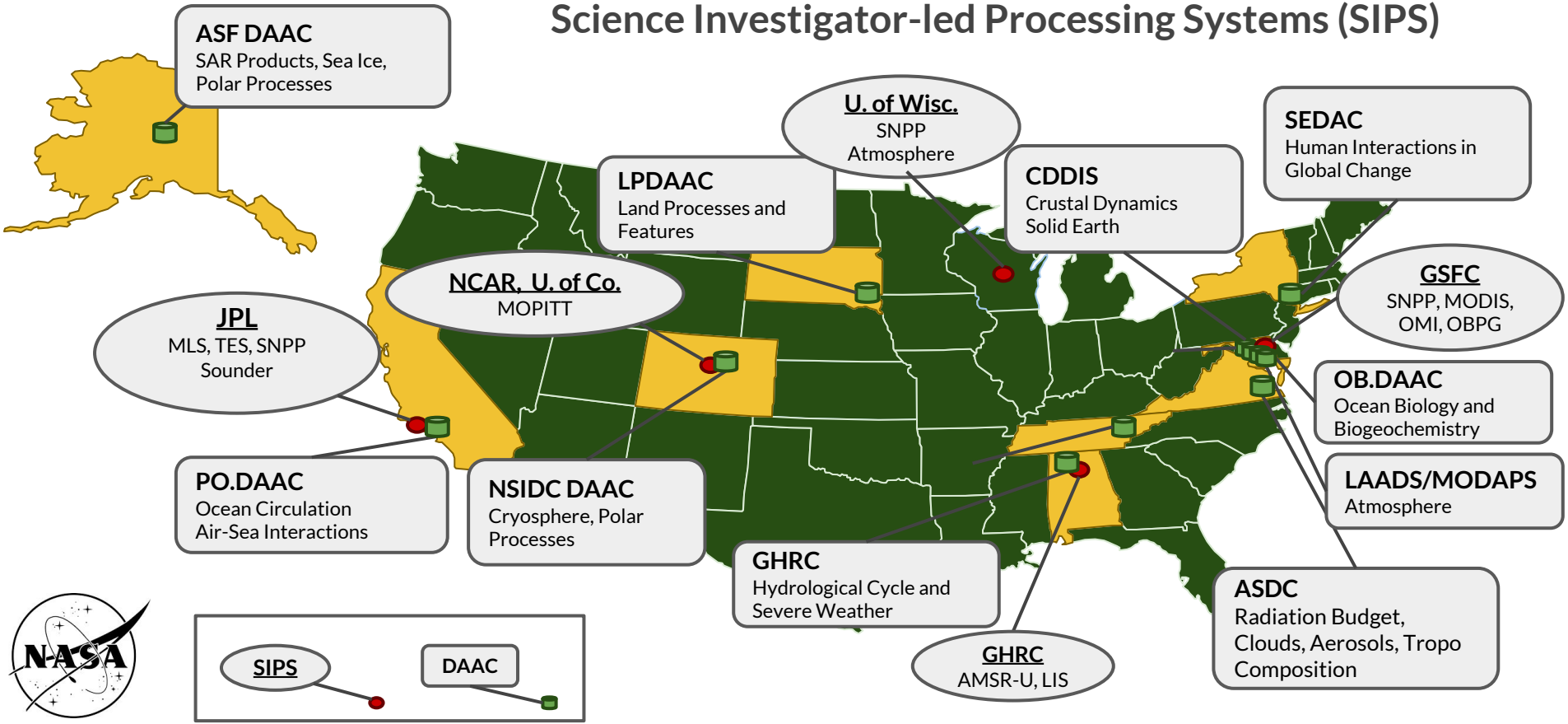
*Earth Observing System Data and Information System (i say “ee-oh-ess-diss”)

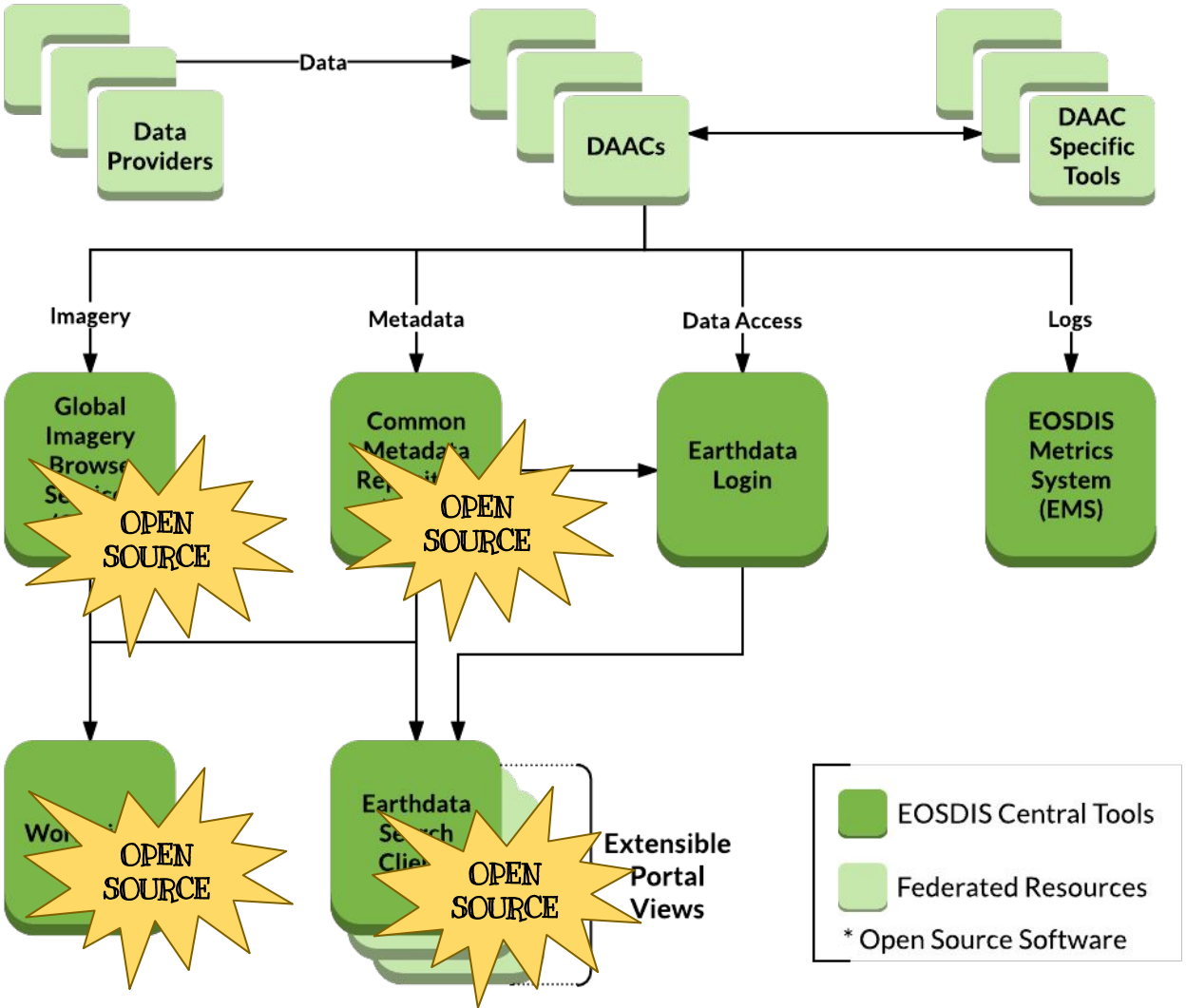
Putting EOSDIS in Context

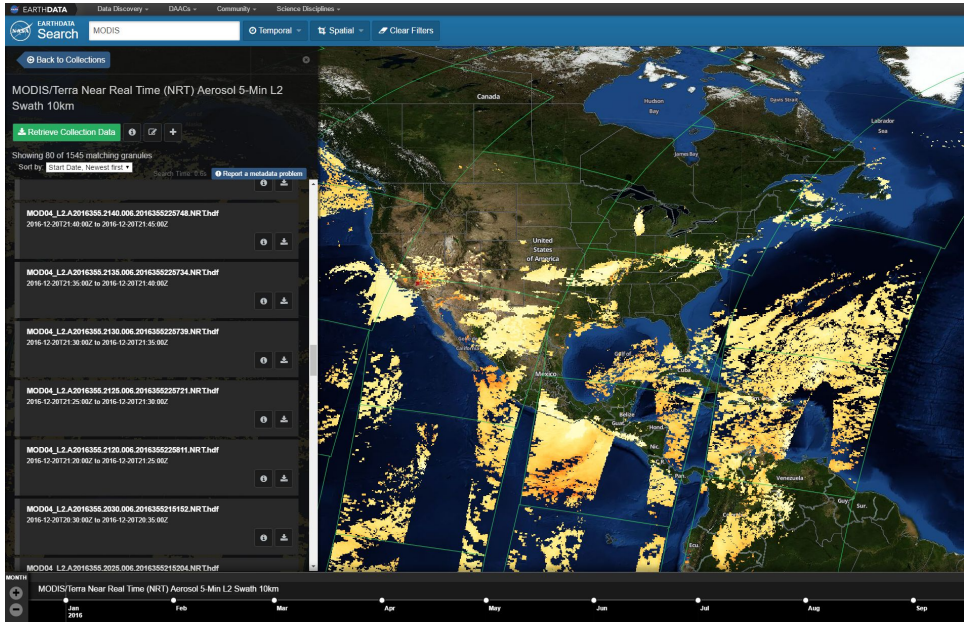


*Subset, reformat, reproject

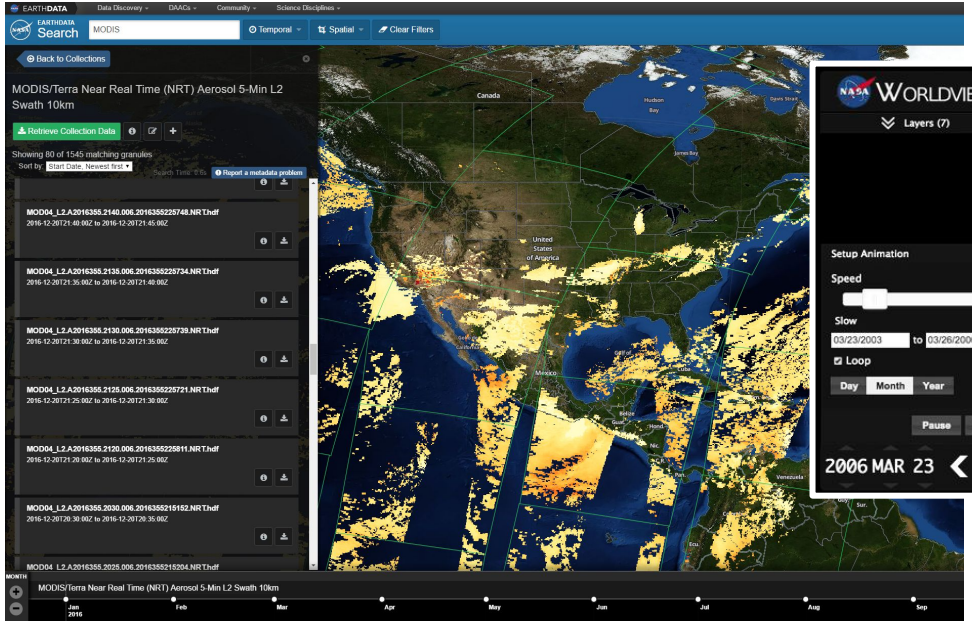
Distributed Active Archive Centers (DAACs), archive and distribute standard data products produced by Science Investigator-led Processing Systems (SIPS)



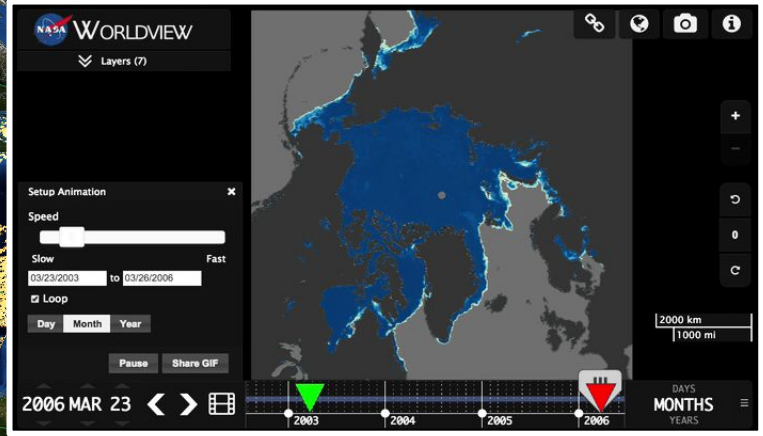

Data
Data Providers
DAACs
DAAC Specific Tools
Imagery
Metadata
Data Access
Logs
Global Imagery Browse
Common Metadata
Earthdata Login
EOSDIS Metrics System (EMS)
OPEN SOURCE
OPEN SOURCE
OPEN SOURCE
OPEN SOURCE
Earthdata
Extensible Portal Views
EOSDIS Central Tools
Federated Resources
* Open Source Software



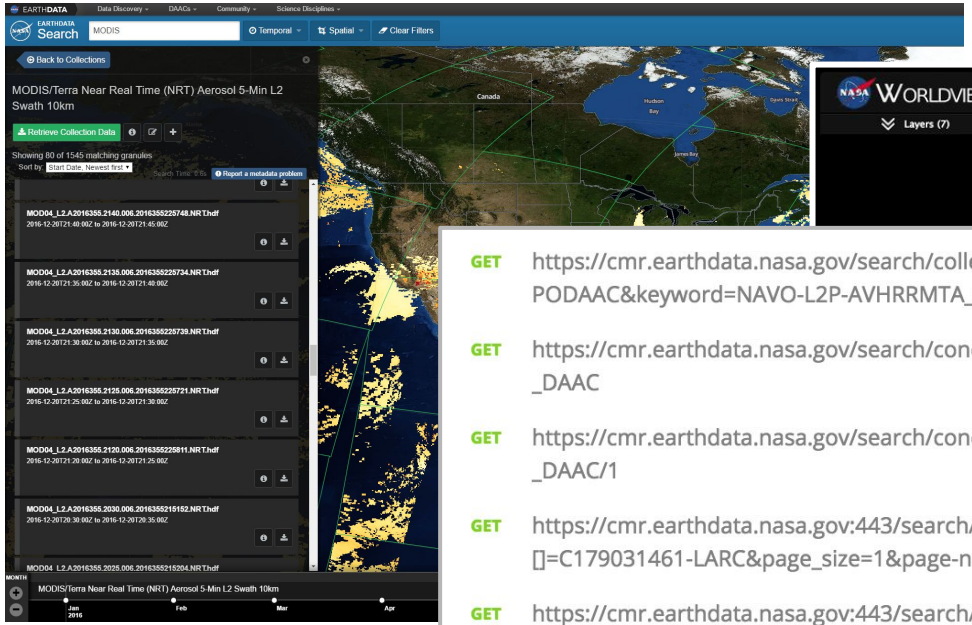
Data Centric End Users
https://search.earthdata.nasa.gov



Data Centric End Users
https://search.earthdata.nasa.gov



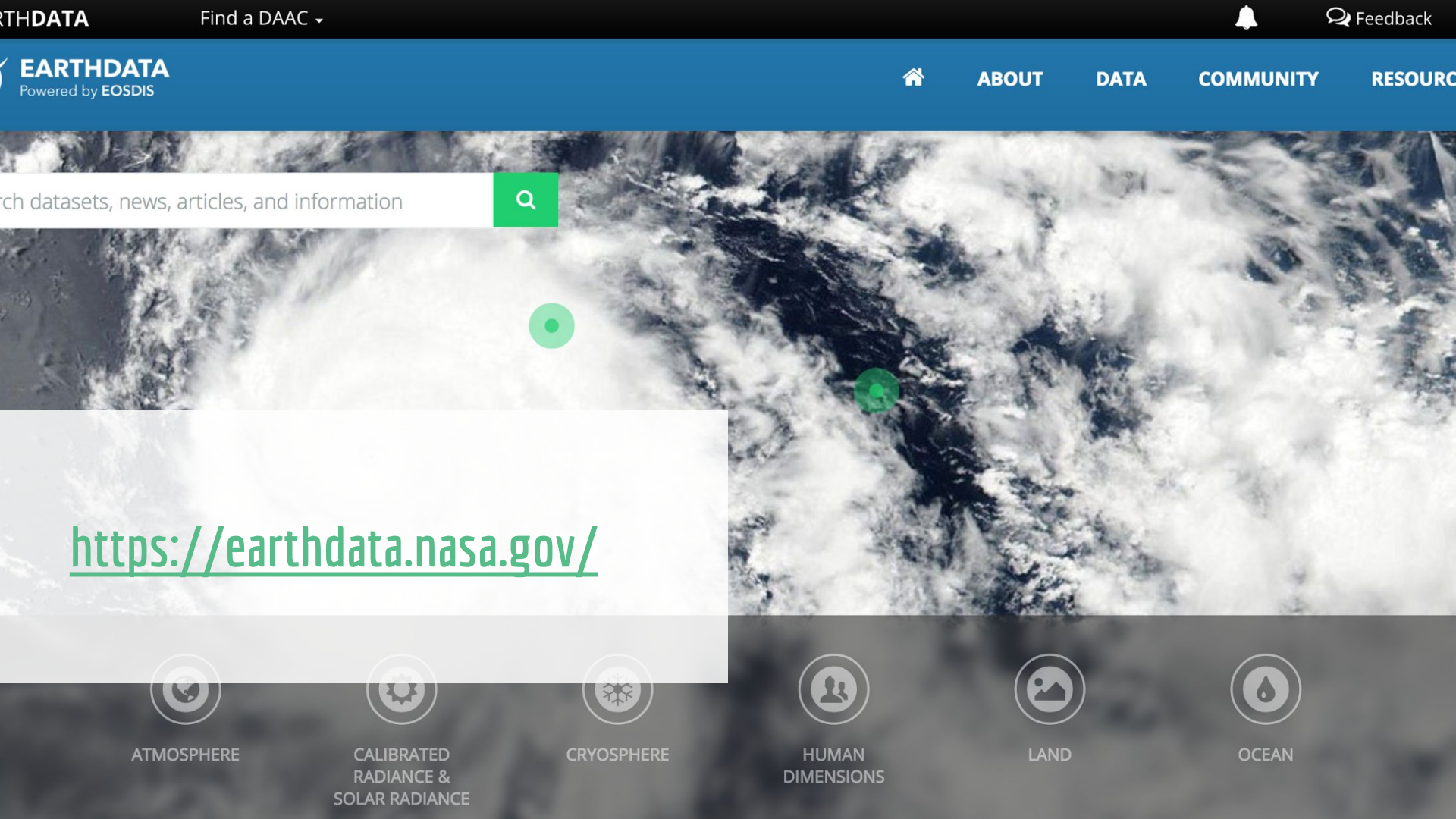
Imagery Centric End Users
https://worldview.earthdata.nasa.gov



Data Centric End Users
https://search.earthdata.n

Code Centric End Users
https://cmr.earthdata.nasa.gov/search

c End Users
ew.earthdata.nasa.gov


RTHDATA
Find a DAAC
Feedback
EARTHDATA
Powered by EOSDIS
ABOUT
DATA
COMMUNITY
RESOURC
rch datasets, news, articles, and information
https://earthdata.nasa.gov/
ATMOSPHERE
CALIBRATED RADIANCE & SOLAR RADIANCE
CRYOSPHERE
HUMAN DIMENSIONS
LAND
OCEAN

And we are poised for some **really big** missions in the early 2020s*, so we've got exciting things on the horizon!

* NISAR and SWOT are going to increase EOSDIS daily total ingest from about 6 TB/day to over 110 TB/day


Second, a Quick How To for Open Sourcing at NASA
(ok, how does this work?)

“

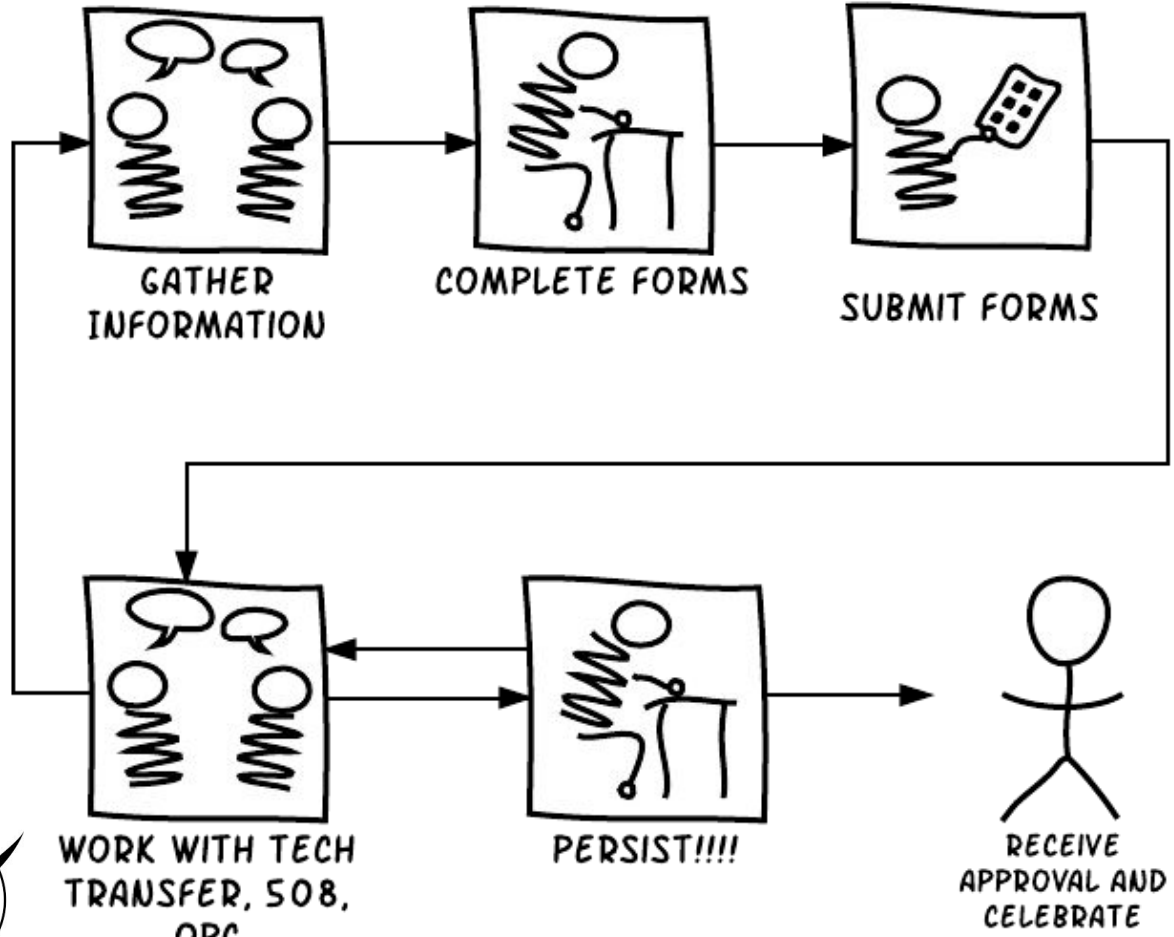
Enhanced reuse of custom-developed code across the Federal Government can have significant benefits for American taxpayers, including decreasing duplicative costs for the same code and reducing Federal vendor lock-in.[5]

This policy also establishes a pilot program that requires agencies, when commissioning new custom software, to release at least 20 percent of new custom-developed code as Open Source Software (OSS) for three years, and collect additional data concerning new custom software to inform metrics to gauge the performance of this pilot.[6]

“

https://sourcecode.cio.gov/

Here is the high-level process



N.B. This process ensures we are inclusive in our practices and abide by federal law

built with http://www.its-all-design.com/ba-sketchbook-stencil-visio/

NASA's Open Sourcing Process in More Detail (a brief sub-presentation)

NASA releases tons of software!

TECHNOLOGY TRANSFER PROGRAM

https://software.nasa.gov/NASA_Software_Catalog_2017-18.pdf

What Forms Will I Need?

New Technology Report - NF1679 (online or via doc template)

(You will need the assigned NTR number (e.g. “GSC-17610”) for subsequent forms)

Export Control Form

Global Concerns Statement

508 Compliance Statement

GSFC Software Developer Form
(aka Software Release Request Authorization or SRRA)

Open Source Questionnaire

https://wiki.earthdata.nasa.gov/display/ESDSWG/Software+Release+Process+-+GSFC+Specific

Things to Be Prepared to Gather

List of prior publications related to software (conferences, etc)

Software Classification
http://nodis3.gsfc.nasa.gov/displayDir.cfm?Internal_ID=N_PR_7150_002B_&page_name=AppendixD

Distributed Dependencies (a bit more on that later)

508 Compliance Information (especially for Web Interfaces)

Contractors must be prepared to release copyright claims

(more information available at https://software.nasa.gov/)

NASA-Wide Software Classifications

Class A	**Human-Rated Space Software Systems**
Class B	**Non-Human Space-Rated Software Systems or Large-Scale Aeronautics Vehicles**
Class C	**Mission Support Software or Aeronautic Vehicles, or Major Engineering/Research Facility Software**

(e.g., Classes A through C are mostly software developed or acquired for Highly Specialized IT systems)

Class D	**Basic Science/Engineering Design and Research and Technology Software**
Class E	**Design Concept and Research and Technology Software**
Class F	**General Purpose Computing, Business and IT Software (Multi-Center or Multi-Program/Project)**
Class G	**General Purpose Computing, Business and IT Software (Single Center or Project)**
Class H	**General Purpose Desktop Software**

Notes: It is not uncommon for a project to contain multiple systems and subsystems having different software classes.

Source

Relevant NPRs

NPR 2210.1C

http://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=2210&s=1C

NPR 7150.2B

http://nodis3.gsfc.nasa.gov/displayDir.cfm?Internal_ID=N_PR_7150_002B_&page_name=main

(more information available at https://software.nasa.gov/).

Licensing

With Regards to Licensing

You will almost certainly need to provide a list of distributed dependencies.

https://en.wikipedia.org/wiki/Viral_license

Any GPL'd or "viral-licensed" software (or in distributed dependencies) will prevent the Office of Patent Counsel (OPC) from being able to release your work.

With Regards to Licensing

Releases are done under NASA Open Source Agreement (NOSA)

Apache 2.0 on a case-by-case basis*

With Regards to Licensing

NASA Open Source Agreement (#NASA)

The NASA Open Source Agreement, version 1.3, is not a free software license because it includes a provision requiring changes to be your "original creation". Free software development depends on combining code from third parties, and the NASA license doesn't permit this.

We urge you not to use this license. In addition, if you are a United States citizen, please write to NASA and call for the use of a truly free software license.

https://www.gnu.org/licenses/license-list.html#NASA

With Regards to Licensing

“

We are seeking the Apache 2.0 licenses based on the current statement from the gnu.org regarding its advice for adoption, explained below.

We feel that this would hinder contributors and give pause to anyone considering augmenting and extending our code with other existing code bases (e.g. ‘mash-ups’)

https://www.gnu.org/licenses/license-list.html#NASA

“

1022

number of projects released at software.nasa.gov

(End Sub-Presentation)


Recent Open Source Efforts
(only talking about EOSDIS, actively “working” repos)

Some of our more active code bases....

OnEarth	https://github.com/nasa-gibs/onearth
Worldview	https://github.com/nasa-gibs/worldview
Earthdata Search	https://github.com/nasa/earthdata-search
Common Metadata Repository	https://github.com/nasa/Common-Metadata-Repository
Metadata Management Tool	https://github.com/nasa/mmt
Cumulus	Approved for open source, not public yet
pyCMR	https://github.com/ghrcdaac/cmr

- more on these later

This is great!
Can’t wait to really start giving back!


Looking at Worldview:
How can we help?

University Students Working on NASA FOSS

University of Pennsylvania held a FOSS class in Fall of 2016.

A pair of students decided to tackle working on Worldview to get their feet wet in contributing.

The blogged about it!

https://www.cis.upenn.edu/~cdmurphy/foss/fall2016/
http://dylancodes.tumblr.com/tagged/cis399
https://leesaf.tumblr.com/

University Students Working on NASA FOSS Great Idea!

Off to a good, but rocky start!

10 months ago

#NASA Worldview

#Worldview

#Open Source

#Computer Science

#Software Development

Setting Up the Worldview Dev Environment on Linux

So! After quite an adventure trying to figure out how to set up the environment and fighting with vagrant, I've come to these concise instructions on getting this set up. *Please note*, I am using Ubuntu 14.04, so different Linux distros may have different steps for unpacking into the localhost.

The following are slight modifications on the Manual Setup instructions on the project github:

1. Download and install Node.js using the instructions on their site.

2. Clone the repo:

```
git clone https://github.com/nasa-gibs/worldview.git
cd worldview
```

Oh No!

9 months ago

#Open Source

#Software Development

#Cesium

#JavaScript

Getting Started with Cesium!

Unfortunately, a number of issues (mostly related to my lack of experience with web development) came up, and I will be unable to work with NASA's Worldview. Their team was very accommodating and did their best, but ultimately the project was too large and didn't have a large base of contributors that were able to help one another with working on issues.

However, I am very excited to begin contributing to Cesium, an open-source library for JavaScript for 3D globes and maps! Installing the program was as easy as typing 'npm install cesium', and forking/cloning the Github repo was also straightforward. What's more exciting is that other Penn students have already been contributing to this project (so I will have more face-to-face resources to rely on), and to boot, the founder of the project is actually a faculty member in the computer science department.

http://dylancodes.tumblr.com/post/151762667869/getting-started-with-cesium

We've made great strides since then!

Project Roadmap

https://github.com/nasa-gibs/worldview/wiki/Worldview-Roadmap

Specific README Sections on Installation and Contribution

More transparent testing:

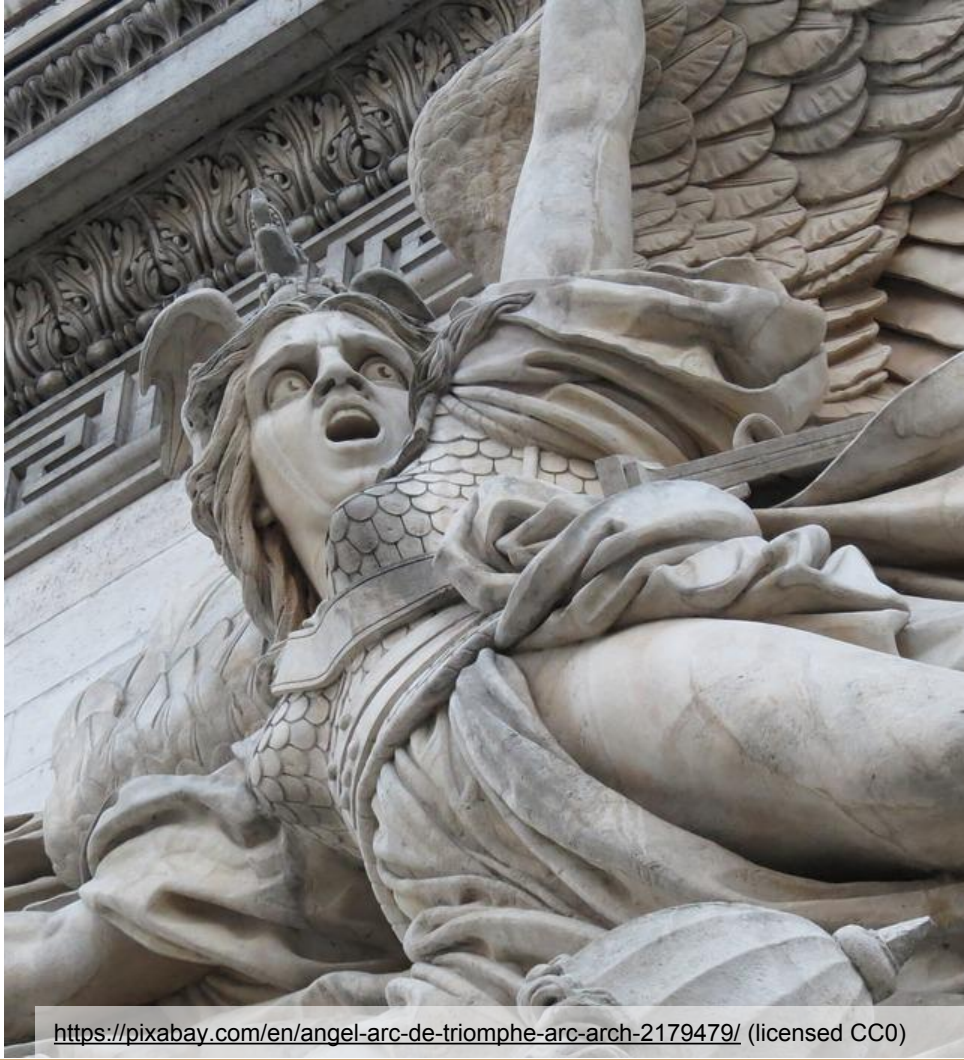
https://travis-ci.org/nasa-gibs/worldview

More transparent issue tracking:

https://waffle.io/nasa-gibs/worldview


What else could go wrong?

- Sometimes people fork projects and never attempt to remerge
- Sometimes code history gets deleted
- Sometimes people leave projects without identifying successors
- Sometimes the direction of the project can be unclear
- Sometimes code just gets thrown over the wall

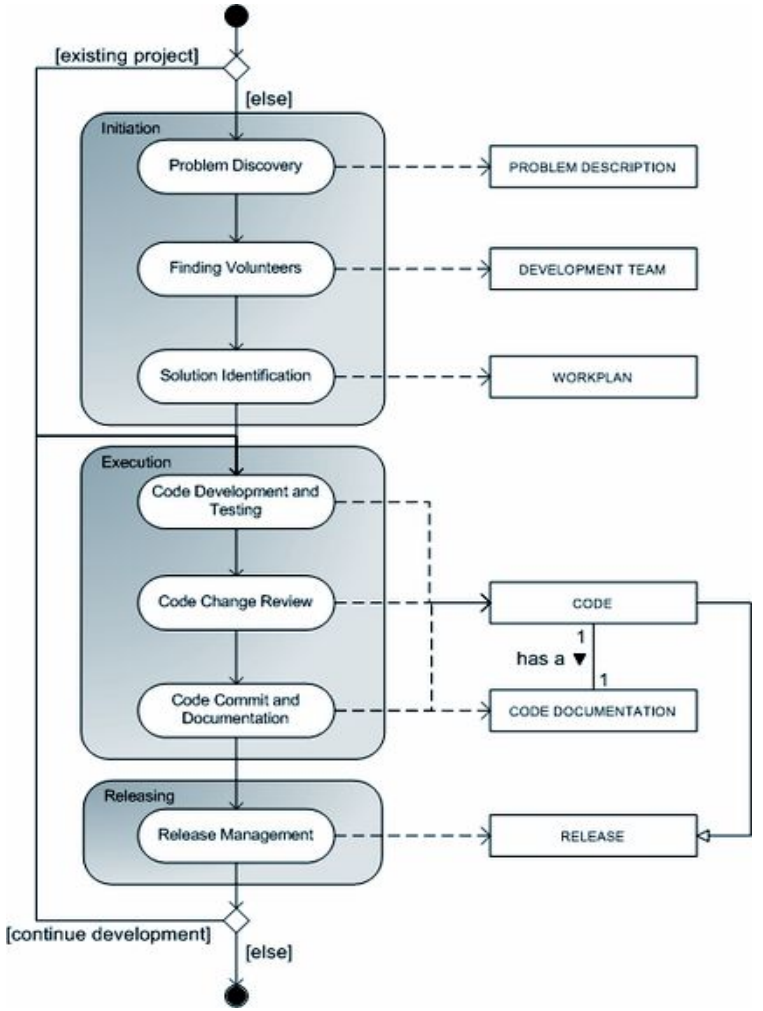


https://pixabay.com/en/angel-arc-de-triomphe-arc-arch-2179479/ (licensed CC0)

NASA is full of process and procedure.
How can we improve upon this one?

Starting with a clear plan.

1. We want to use
   and reuse
   and reuse our software.
2. We have a willing group of volunteers
3. We have vision of how we want to evolve.
4. We have an opportunity to improve on
   “Execution” and “Releasing” in the open.



https://commons.wikimedia.org/wiki/File:OSSD_process_data_diagram.png

Embracing Open Source

(another case study in which we are starting on the ground floor)

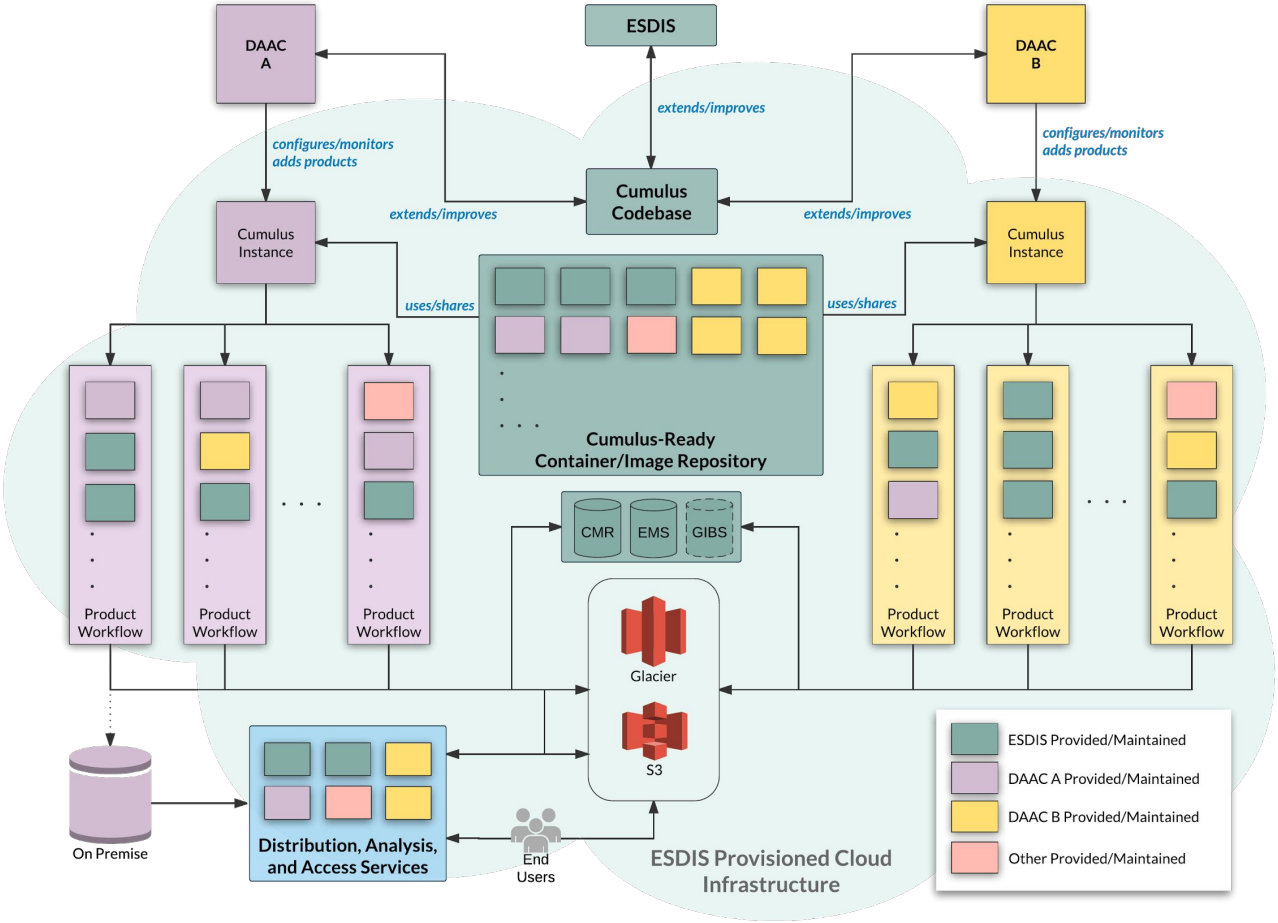
Prototyping DAACs in the Cloud

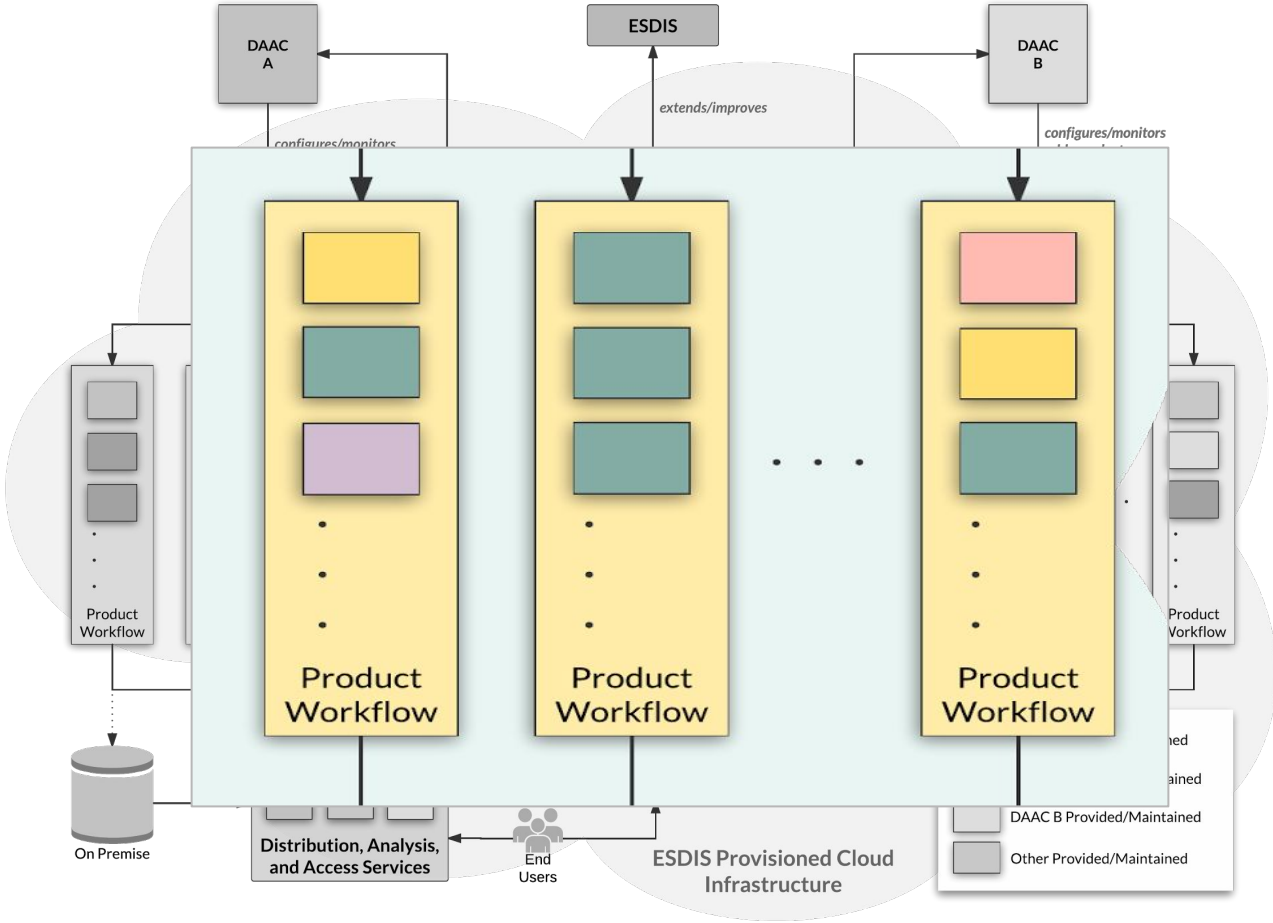
What is Cumulus?

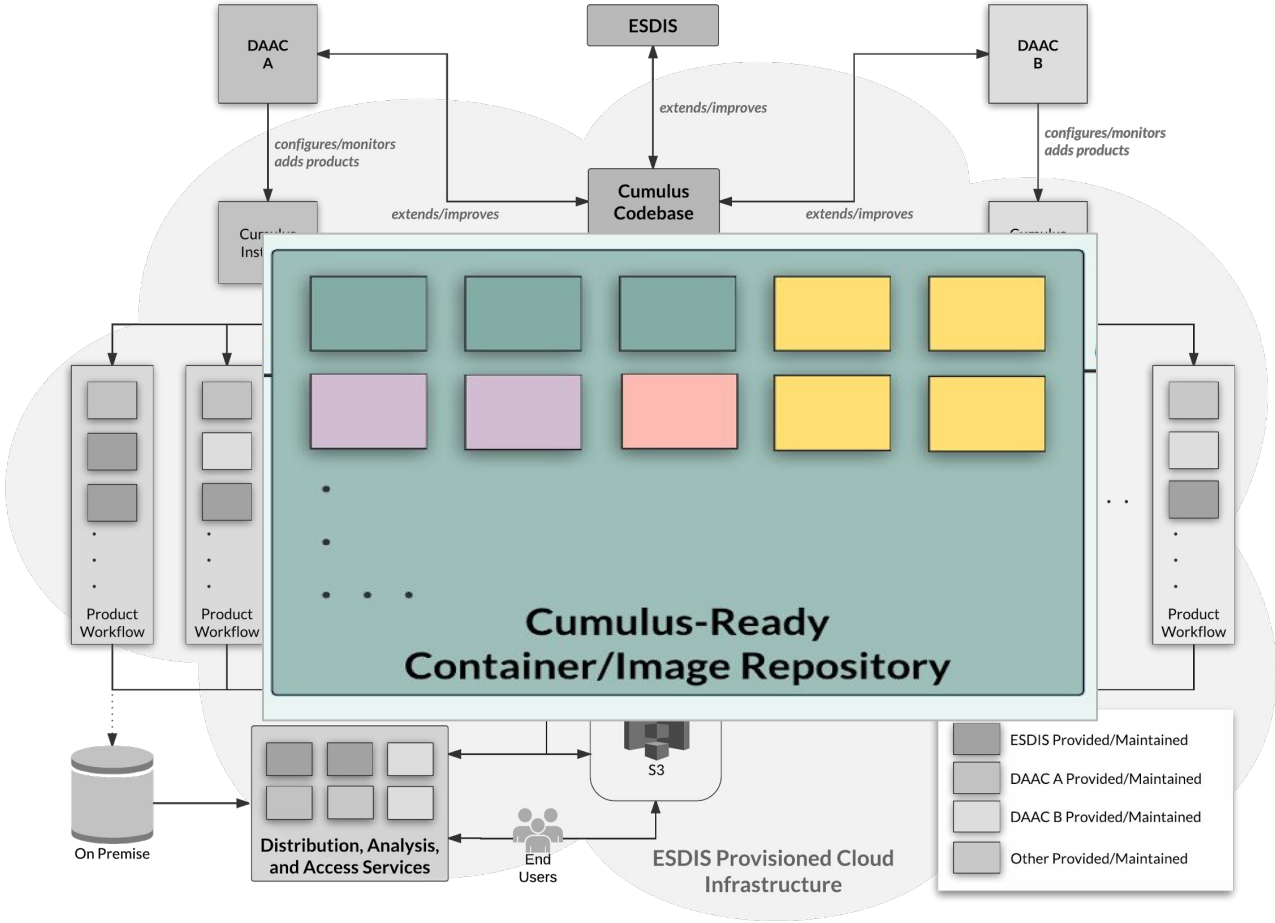
Lightweight cloud-native framework for data ingest, archive, distribution and management

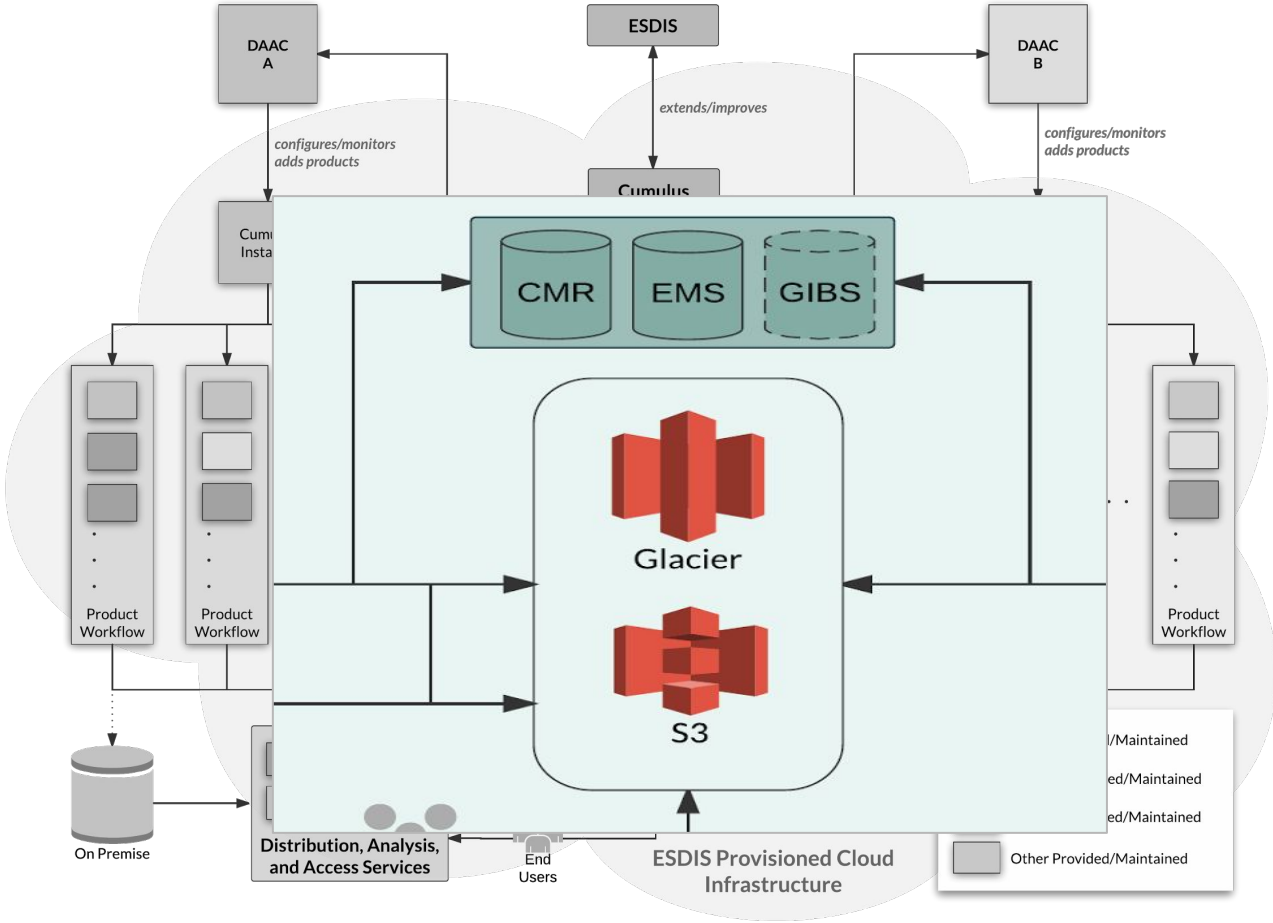
Goals

- Provide core DAAC functionality in a configurable manner
- **Enable DAACs to help each other with re-usable, compatible containers (e.g. widely applicable GIS components or sub-setters)**
- **Enable DAAC-specific customizations**


DAAC A
ESDIS
DAAC B
extends/improves
configures/monitors adds products
configures/monitors adds products
Cumulus Codebase
extends/improves
extends/improves
Cumulus Instance
Cumulus Instance
uses/shares
uses/shares
Cumulus-Ready Container/Image Repository
CMR
EMS
GIBS
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Glacier
S3
On Premise
Distribution, Analysis, and Access Services
End Users
ESDIS Provisioned Cloud Infrastructure
ESDIS Provided/Maintained
DAAC A Provided/Maintained
DAAC B Provided/Maintained
Other Provided/Maintained


DAAC A
ESDIS
DAAC B
extends/improves
configures/monitors
configures/monitors
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Product Workflow
On Premise
Distribution, Analysis, and Access Services
End Users
ESDIS Provisioned Cloud Infrastructure
DAAC B Provided/Maintained
Other Provided/Maintained


DAAC A
ESDIS
DAAC B
extends/improves
configures/monitors adds products
configures/monitors adds products
Cumulus Codebase
extends/improves
extends/improves
Cumulus-Ready Container/Image Repository
Product Workflow
Product Workflow
Product Workflow
S3
On Premise
Distribution, Analysis, and Access Services
End Users
ESDIS Provisioned Cloud Infrastructure
ESDIS Provided/Maintained
DAAC A Provided/Maintained
DAAC B Provided/Maintained
Other Provided/Maintained


DAAC A
ESDIS
DAAC B
extends/improves
configures/monitors adds products
configures/monitors adds products
Cumulus
CMR
EMS
GIBS
Glacier
S3
Product Workflow
Product Workflow
Product Workflow
On Premise
Distribution, Analysis, and Access Services
End Users
ESDIS Provisioned Cloud Infrastructure
Other Provided/Maintained

We have long been a system of stovepipes and “not invented here” types. And that is changing as we evolve

We need to streamline, and create unified, interoperable system that can grow with us. Something we can claim group ownership of.

So … how do we do this “right”

Drafting a contribution guide

1. Provide working definitions of the high-level components of the Cumulus system, including specifying which of those components are governed by this document.
2. Establish roles and responsibilities for contributions to Cumulus NASA's EOSDIS
3. Identify key communication flows, as well as information on documentation, testing and deployment paradigms
4. Outline high-level process expectations for Cumulus contributions and provide example process flows for these contributions

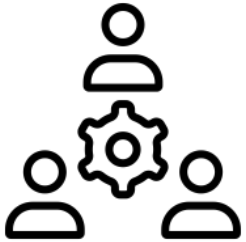
Mimic the Apache Way, but for a simpler use case and narrower audience

Establishing Project Roles and Responsibilities

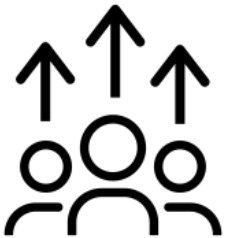
Project Management Board

Project Leadership Team

Technical Leadership Team



Committers



Contributors

As we scale (up or down) we can adjust/combine/tailor these roles.
We can adapt to other projects/systems.

Form a Working Group!

give them homework

set a release timeline

Group Membership

@Kathleen Baynes	@Christine Smit	@Jason Werpy
@Christopher Lynnes	@Darla Werner	@Manil Maskey
@Mark McInerney	@Taylor Wright	Rustem Albayrak
@Chris Stoner	@Patrick Quinn	@Jason Duley
@Ian Schuler	@Christopher Torbert	@Lewis McGibbney
@Michael Gangl	@Wayne Burke	@Ajinkya Kulkarni

Individual Assignments

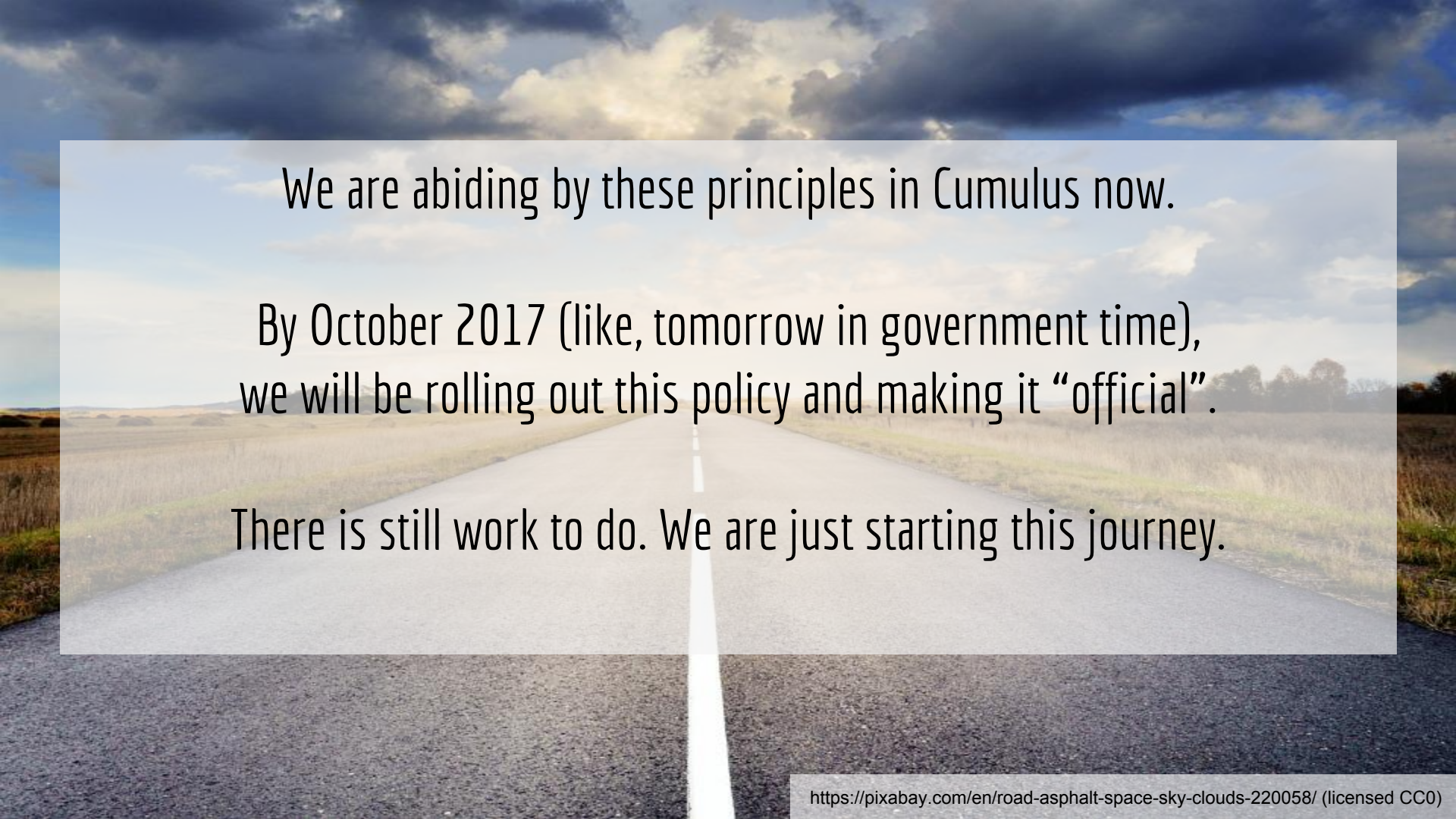
Volunteer For An Assignment / Add A Resource

Understand your Obligations and Limitations

- Schedule regular meetings with the NASA Office of Patent Counsel
  - Do we need contributor licenses agreements?
  - Can we use docker hub?
  - What system evolutions require new a release process?
  - How do we keep NASA informed of how we are proceeding?
  - How can we help in guiding NASA policy at large?

Denouement

(so, what's next?)



We are abiding by these principles in Cumulus now.

By October 2017 (like, tomorrow in government time),
we will be rolling out this policy and making it “official”.

There is still work to do. We are just starting this journey.

Questions?

katie.baynes@nasa.gov