

Disclaimer: All examples are preliminary

National Aeronautics and  
Space Administration



# On-demand TEMPO Data Transformations in the Cloud: Concatenating Concatenated Subsets

Daniel Kaufman

Andrey Kim, Christopher Durbin, James Norton, Owen Littlejohns,  
Georgina Hayes-Crepps, Jennifer Tindell, Hazem Mahmoud, Matthew  
Tisdale, John Kusterer

ESIP July 2024

Session: Transforming the NASA Earth Science Enterprise

A large, stylized graphic on the left side of the slide. It features a curved, semi-circular shape that frames a vibrant space scene. The scene includes a bright yellow sun in the lower-left, a blue and white Earth in the lower-right, a large grey moon in the center, a ringed planet (Saturn) in the upper-left, and a reddish planet (Mars) in the upper-middle. The background is a deep blue and green nebula with numerous white stars.

## *Outline*

- Motivating Mission and Data
- Tools at Hand
- Challenges to overcome
- Aggregating data in the cloud



## Tropospheric Emissions: Monitoring of Pollution

Selected as NASA's first Earth Venture  
Instrument in 2012

To measure Atmospheric Composition



# Data Structure



UV/Visible spectrometer view of North America, from a geostationary orbit

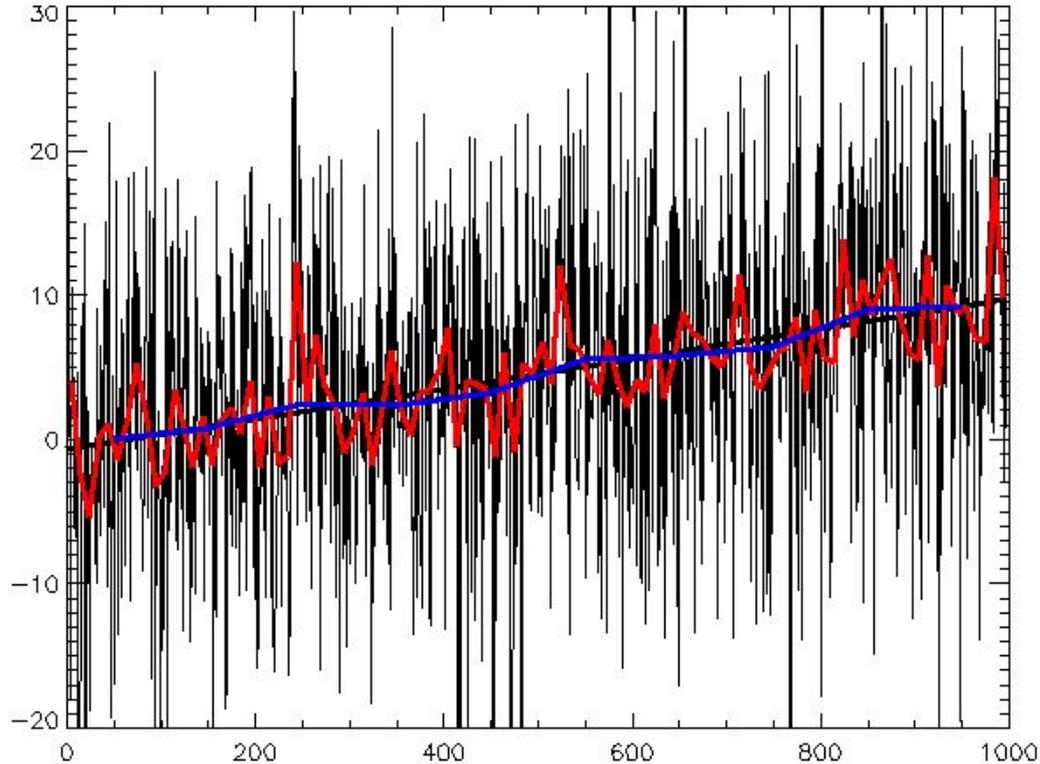
Measures ozone, nitrogen dioxide, formaldehyde, aerosols, and other pollutants

Resolution:

Hourly (~10 mins for special studies)

Sub-urban spatial scale

# Data Structure



To make a time series of a particular variable from a particular location...

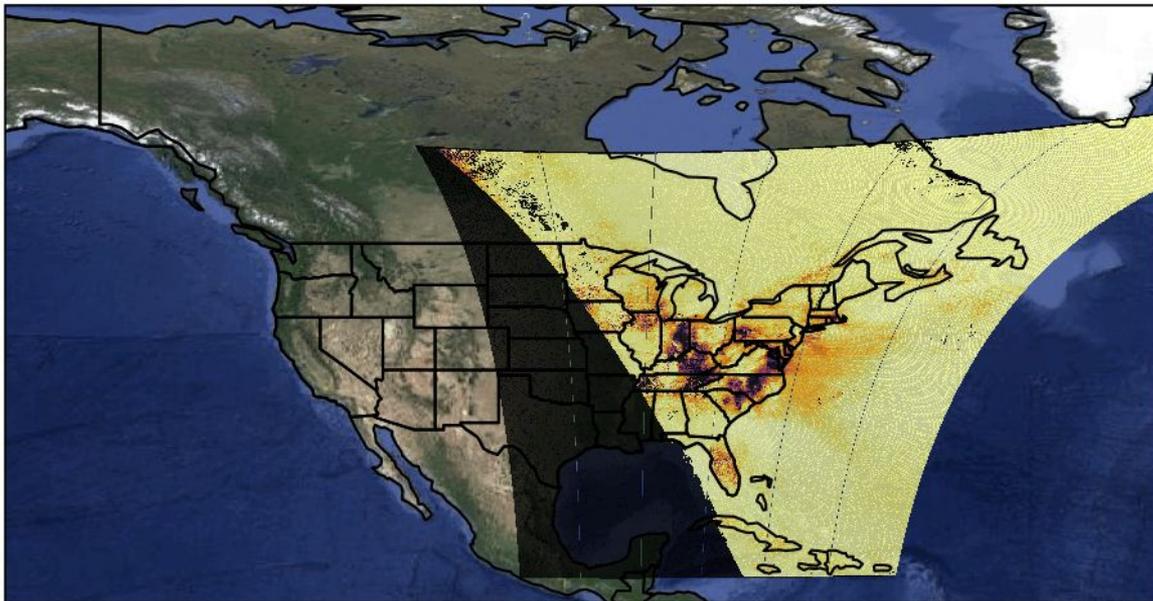
The data may need to be spatially subsetted, and possibly concatenated into workable object in memory.

## **Objective:**

Support the ability for users to download multiple files subsetted to an area of interest and aggregated into a single file.

# Data Structure

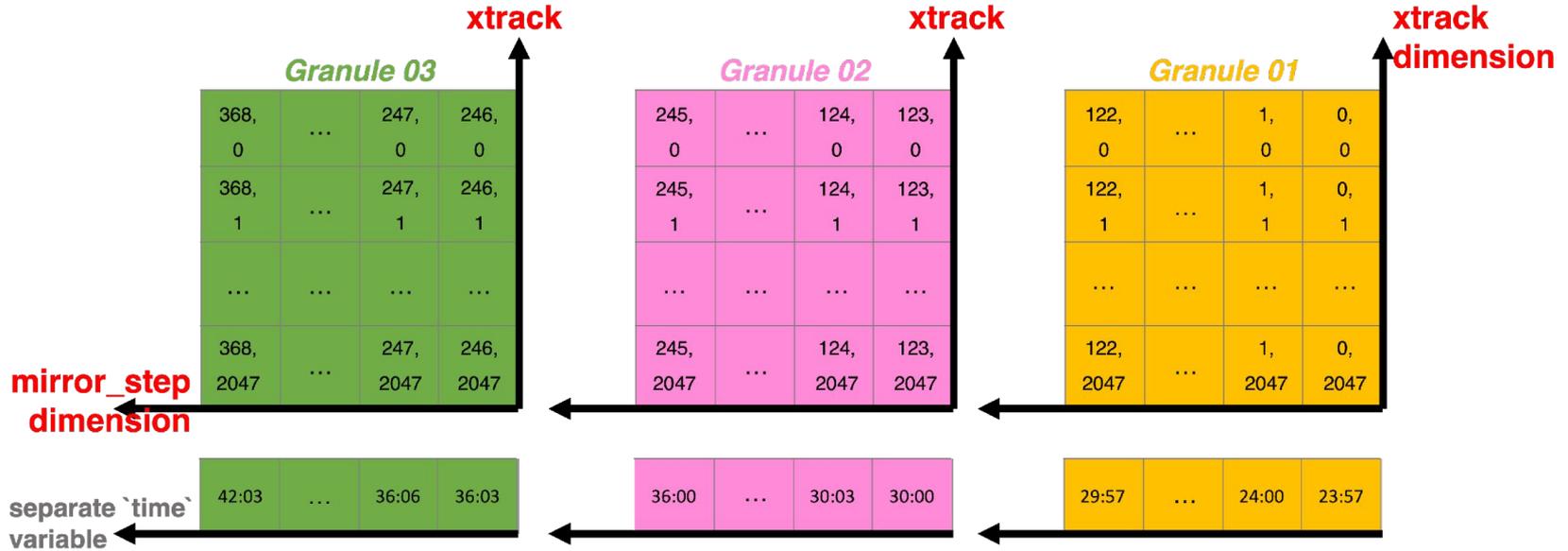
2024-05-09 10:41:07 to 2024-05-09 11:14:16; SCAN S001



Hourly scans are transferred and stored as six-minute “granules”

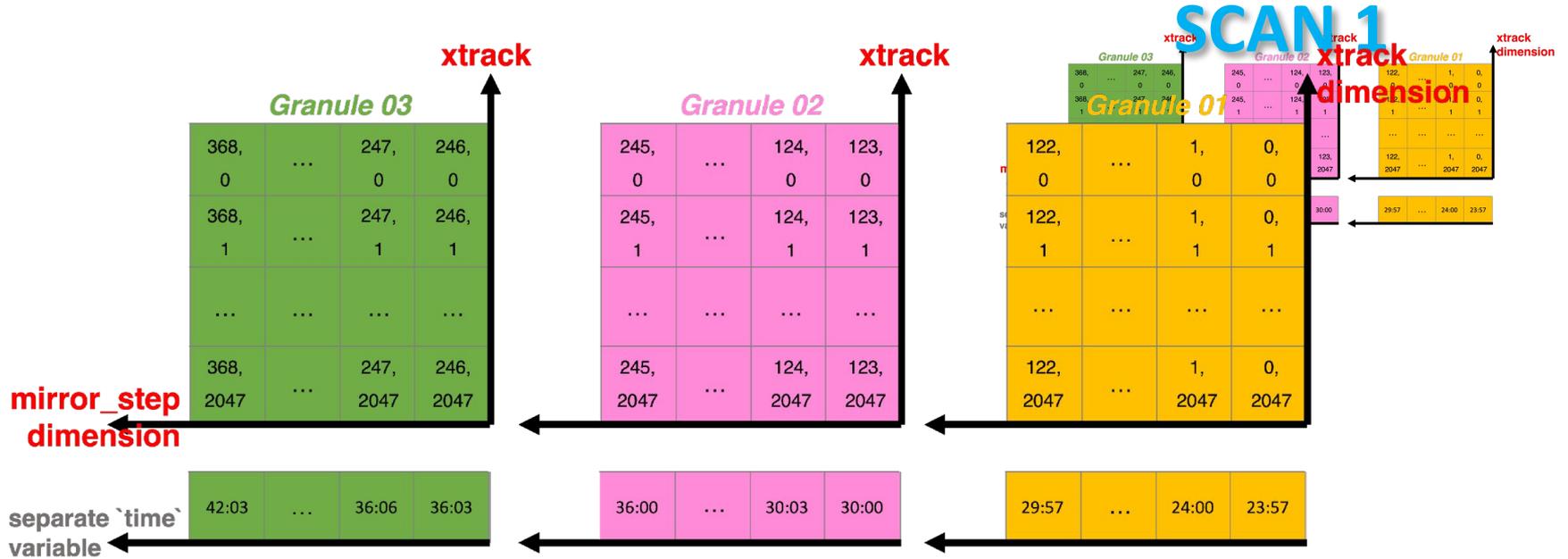
# Data Structure

(for Level-2, two dimensional variables)



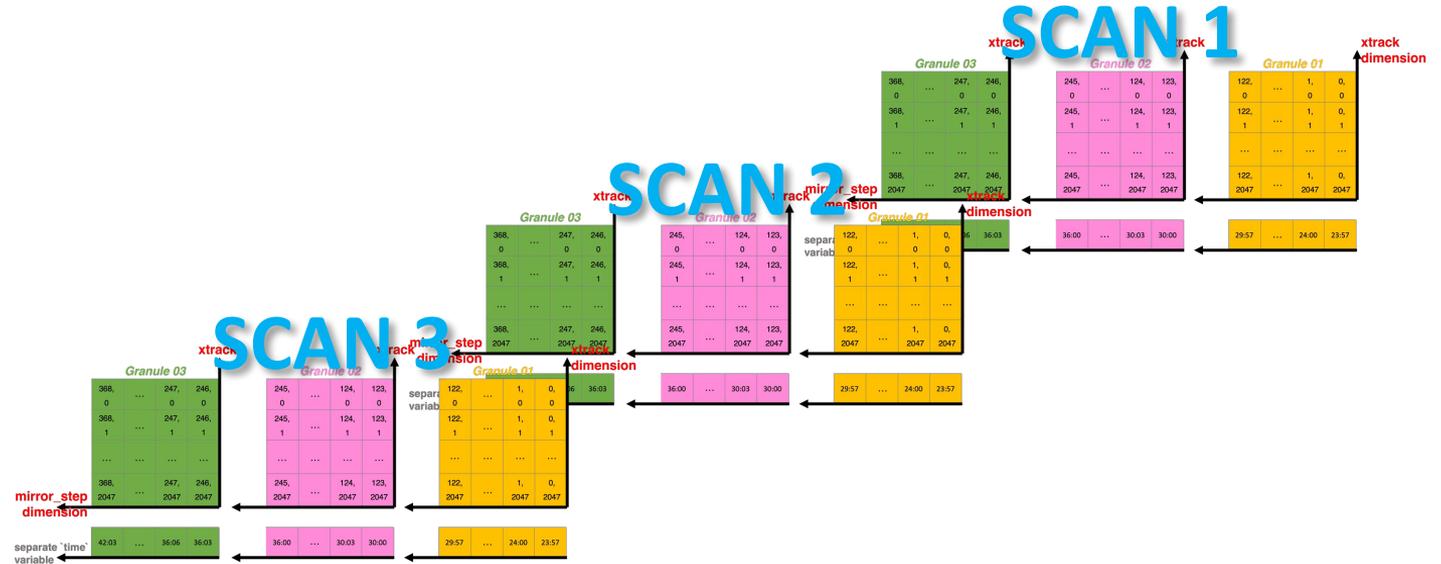
# Data Structure

(for Level-2, two dimensional variables)



# Data Structure

(for Level-2, two dimensional variables)





# Tools already at hand

- Harmony (<https://harmony.earthdata.nasa.gov/> and <https://github.com/nasa/harmony>)
- CONCISE (CONCatenation SErvice; <https://github.com/podaac/concise>)
- Level-2 Subsetter
- Other open-source libraries



# NASA's Harmony project

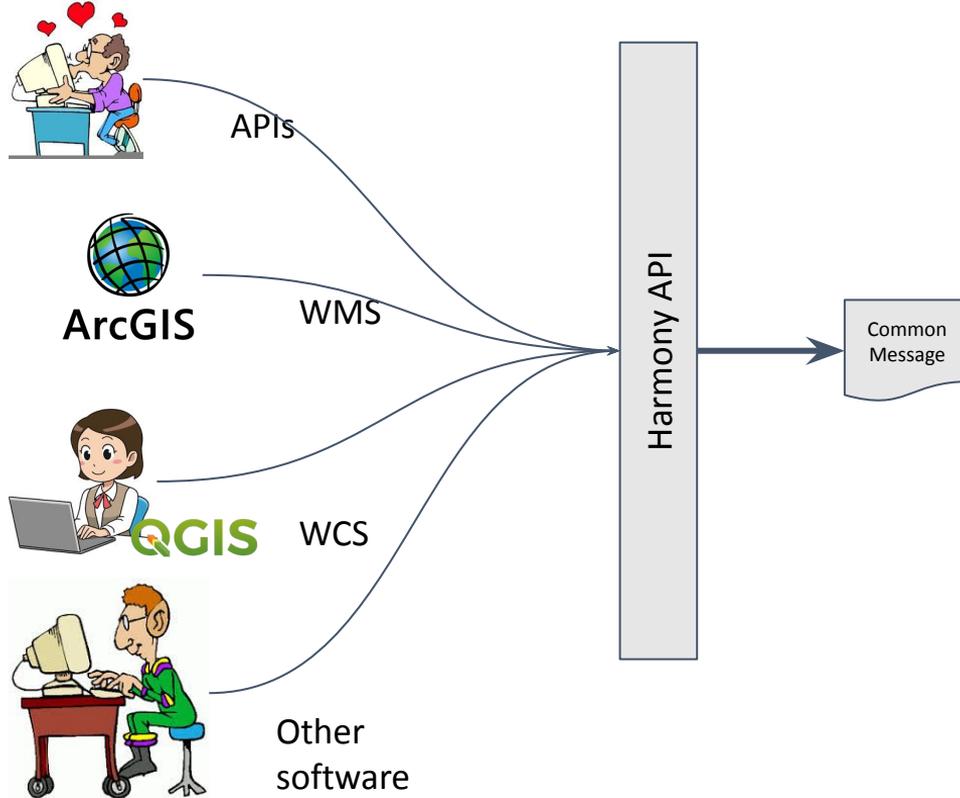
"Bringing Data Together by Bringing Services Together"

Regardless of science domain, Open Geospatial Consortium (OGC)-compatible tools and a customized Application Programming Interface (API) enable access cloud hosted data.

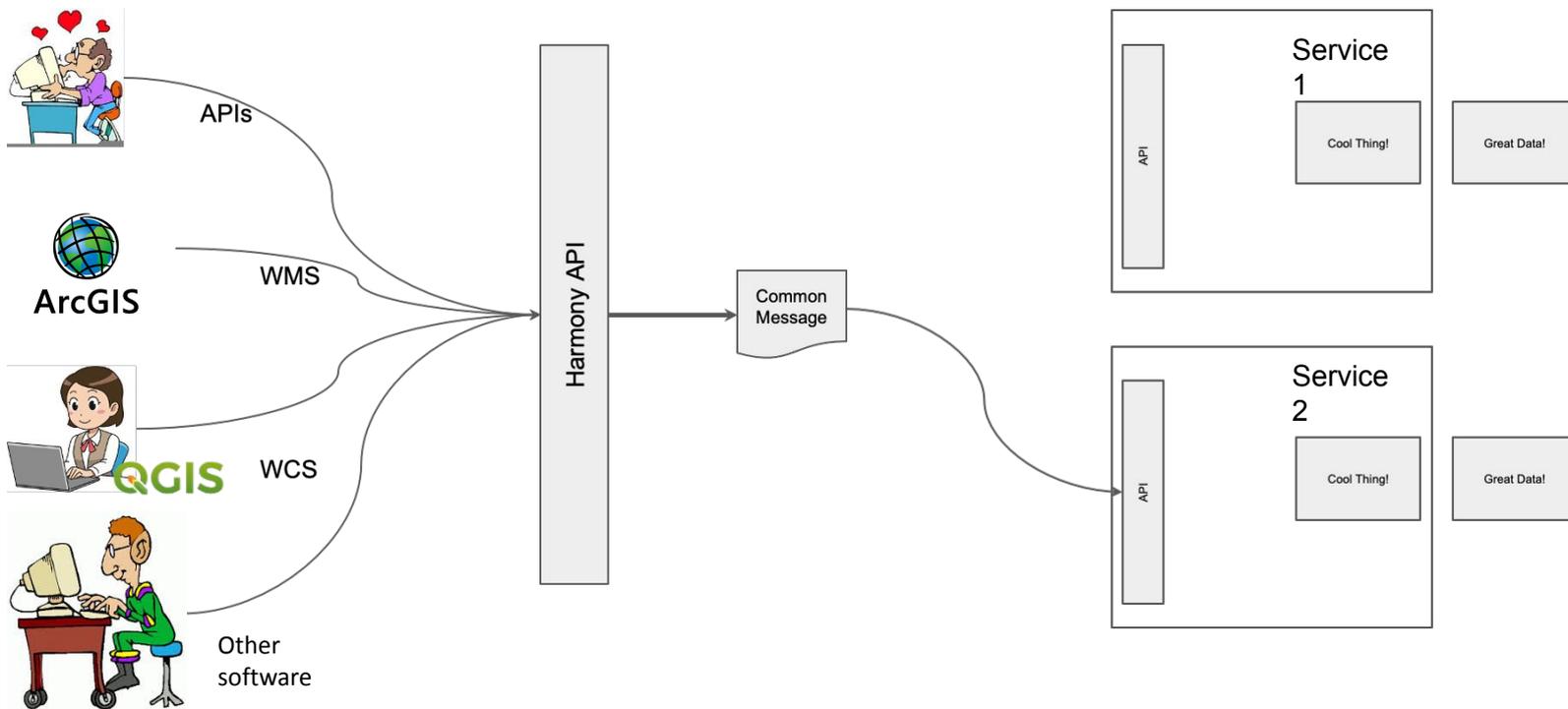
APIs allows users to request only the data they want, in the format and projection they want, allowing them to skip manual intermediary processing steps.



# NASA's Harmony project



# NASA's Harmony project

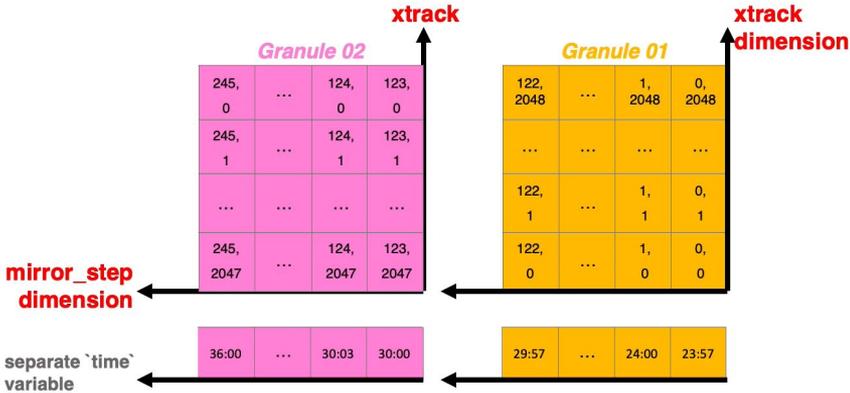


# Existing Harmony Concatenation Service

*combines data by creating a new dimension*

## CONCISE

("stacking")





# Challenges to overcome



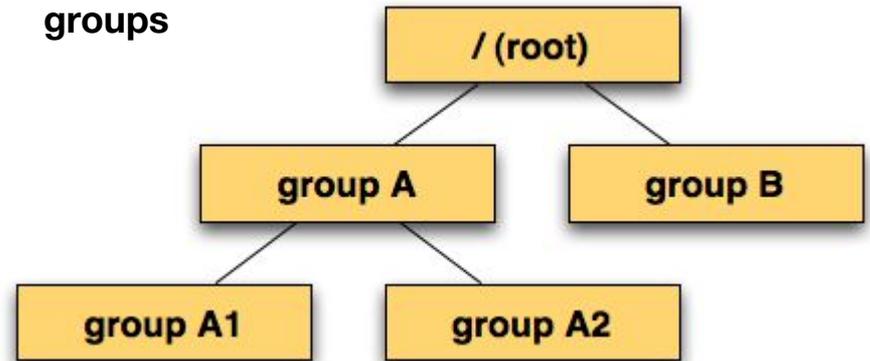
# Challenges to overcome

- Creating a new process for concatenation without “reinventing the wheel”

# Challenges to overcome

- Creating a new process for concatenation without “reinventing the wheel”
- Handling group structure

netCDF-4 (Network Common Data Form)  
**groups**





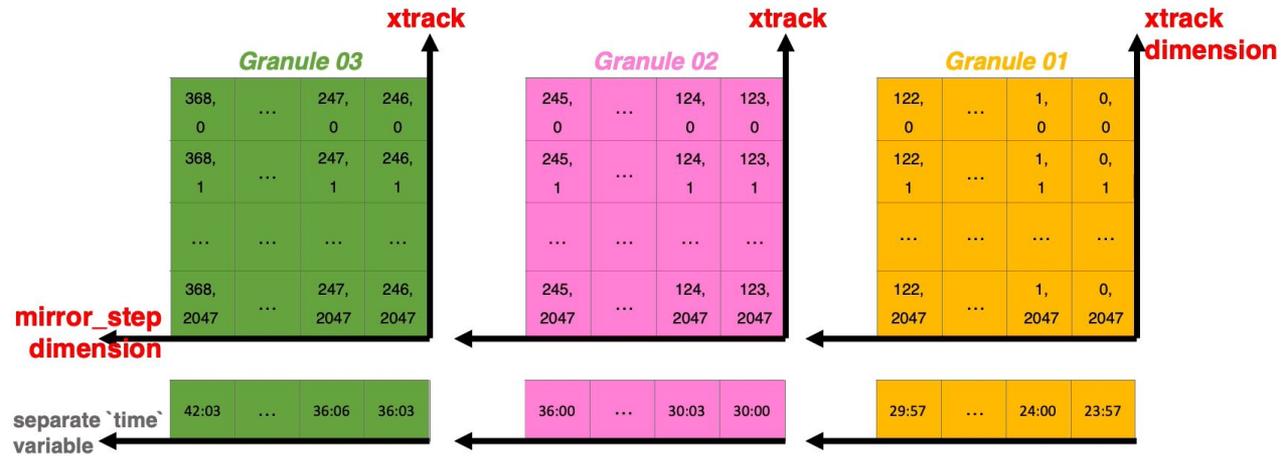
# Challenges to overcome

- Creating a new process for concatenation without “reinventing the wheel”
- Handling group structure
- Hooking up the new process to Harmony system



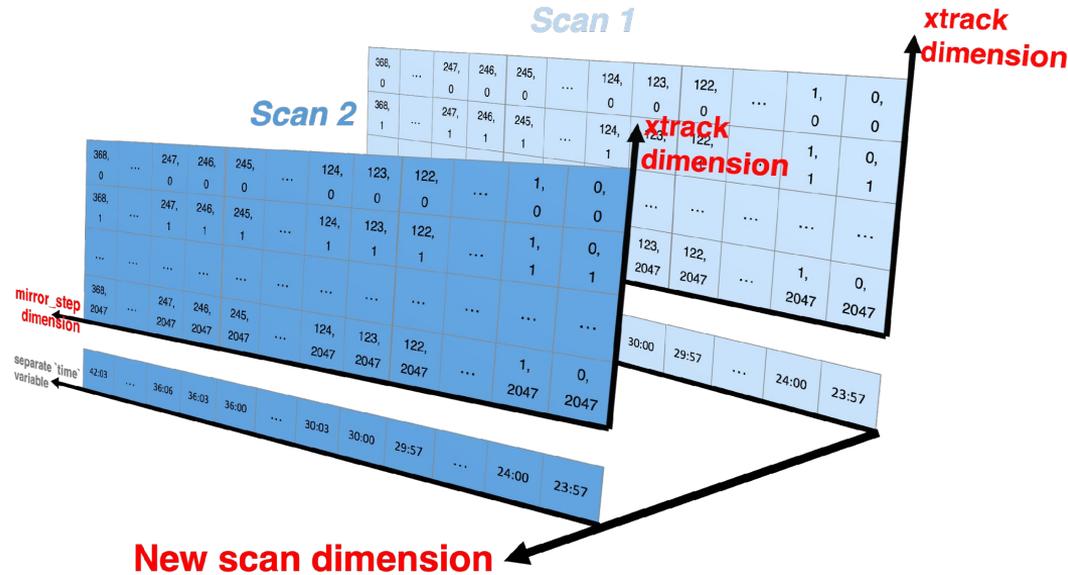
# “Stacking” and “Extending”

## Structure of TEMPO Granules' 2D data



# An intuitive structure for further exploration and analysis

## Resulting Concatenated Data



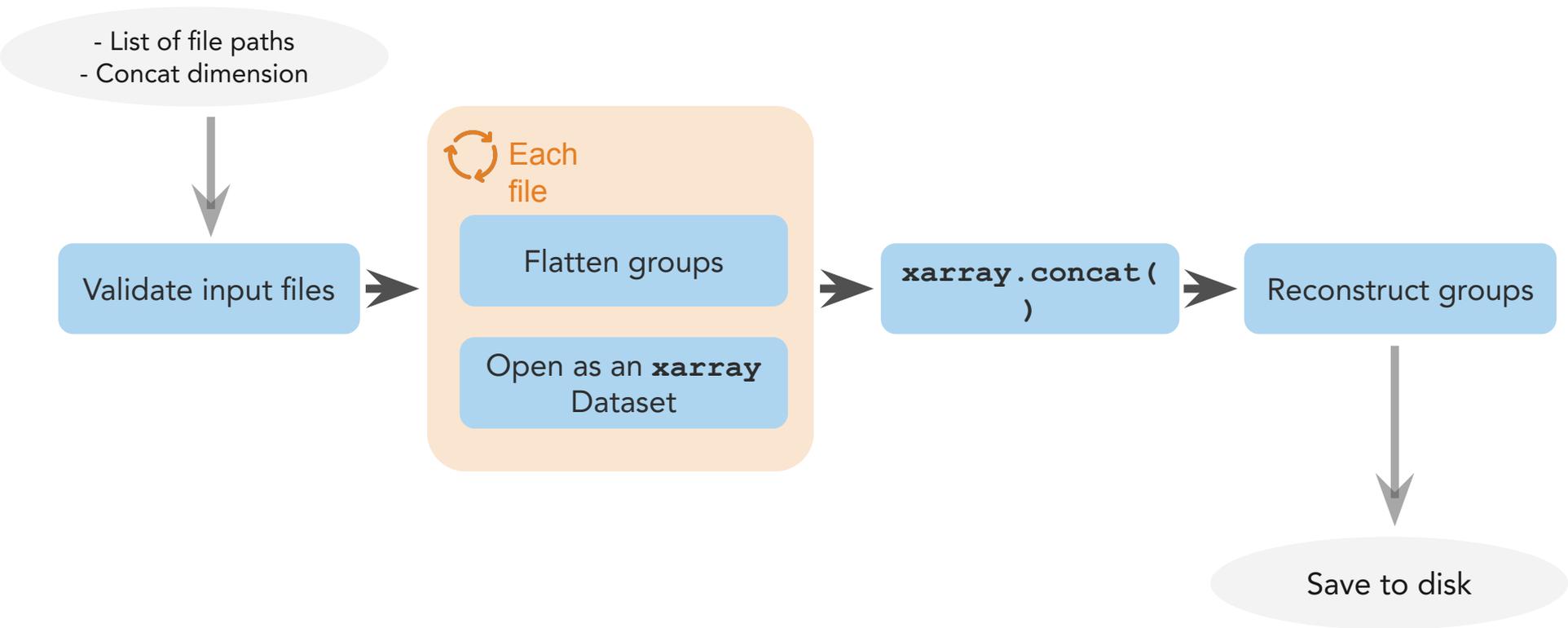
# Functionality Overview



Concatenate along dimension



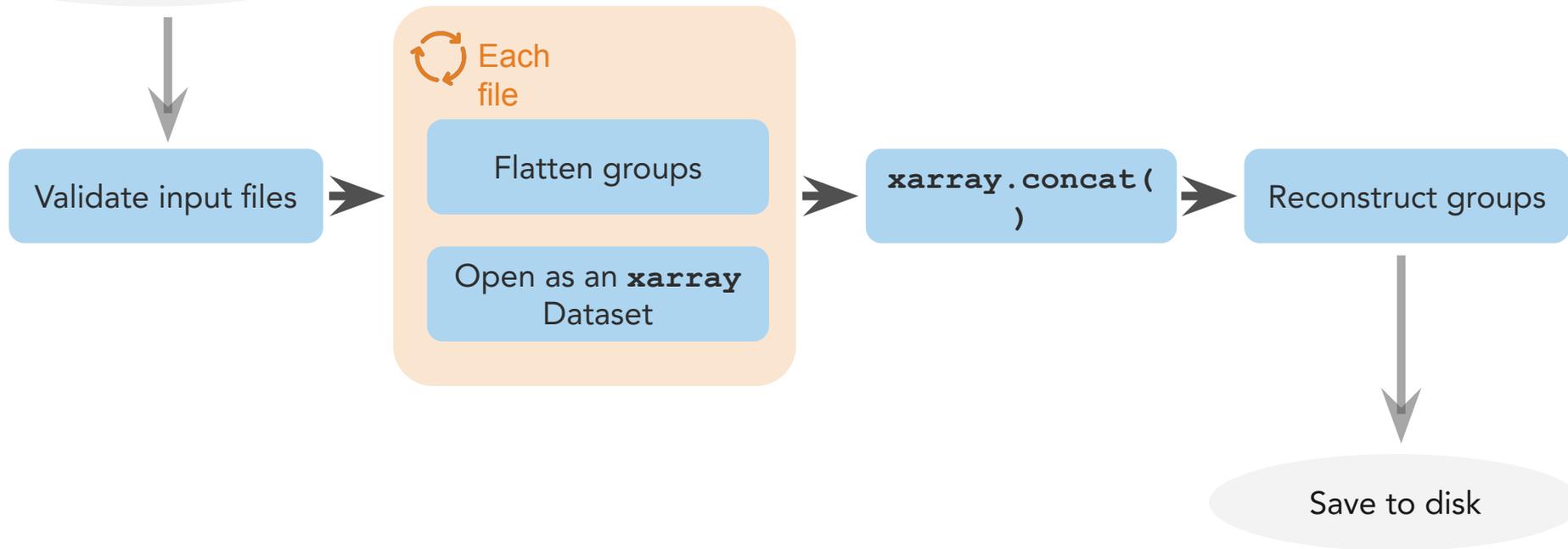
# Functionality Overview



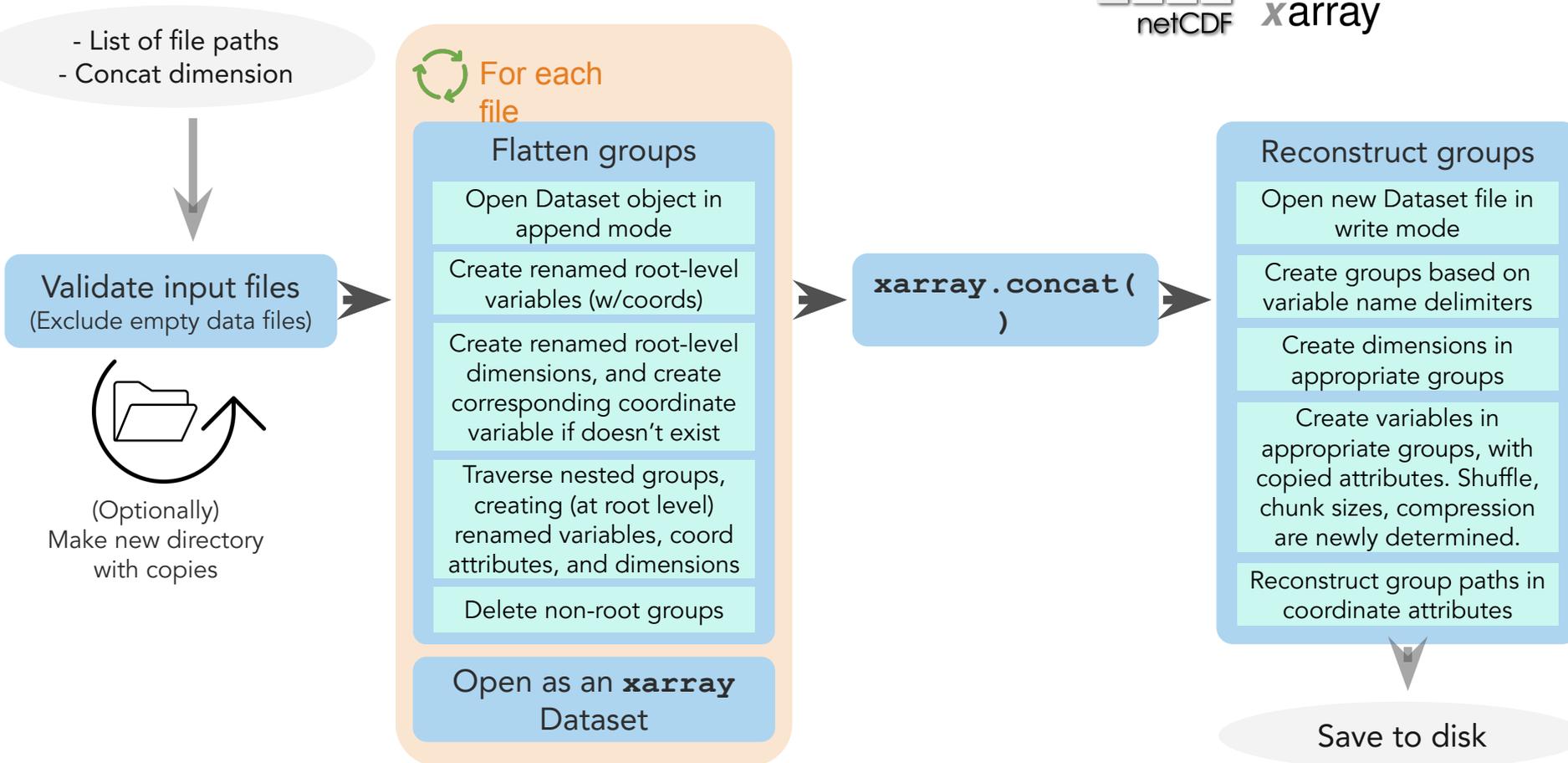
# Functionality Overview



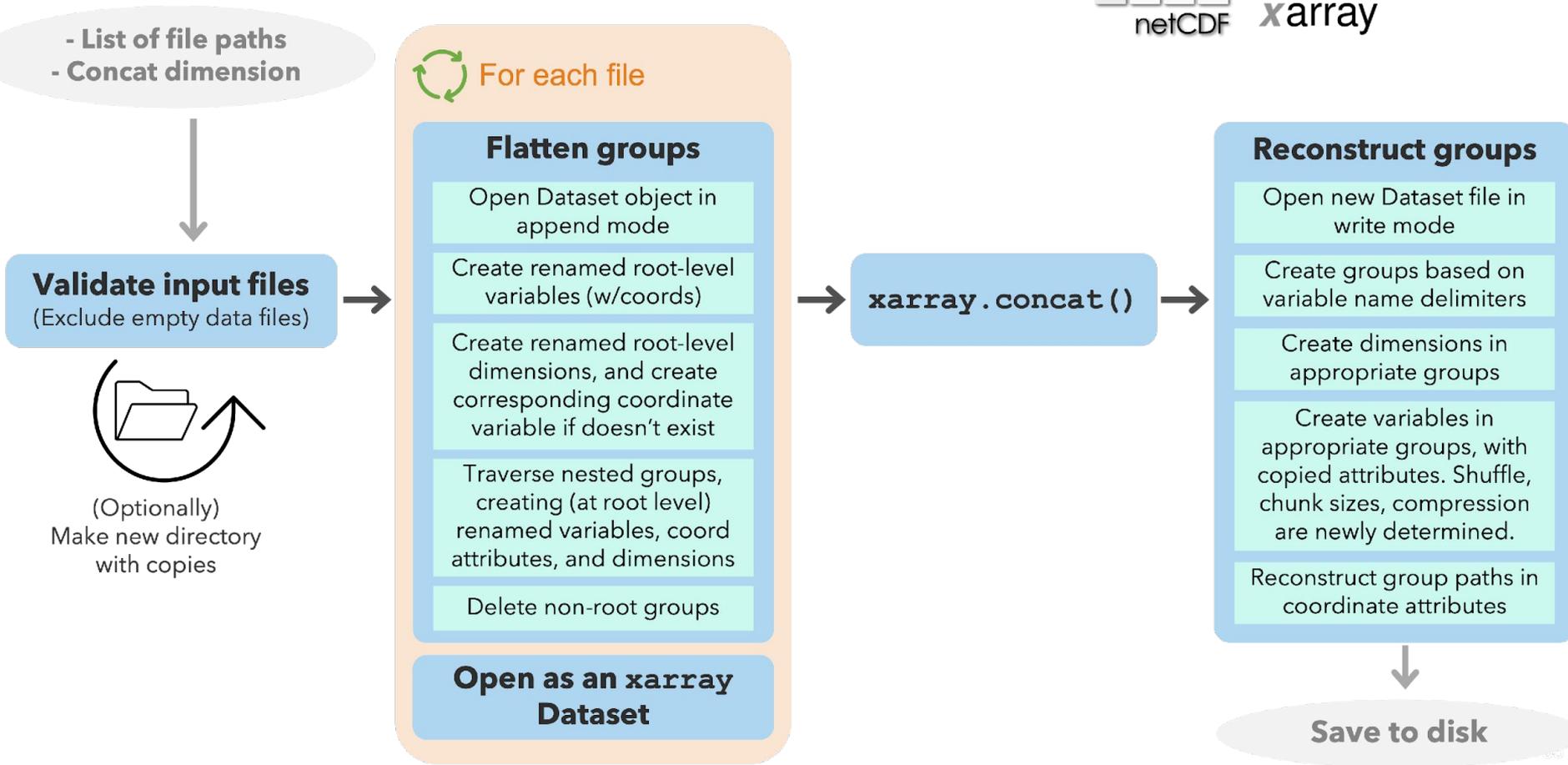
- List of file paths
- Concat dimension



# Functionality Overview



# Functionality Overview



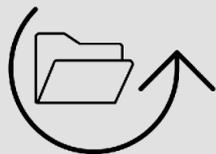
# Functionality Overview



- List of file paths
- Concat dimension

For each file

**Validate input files**  
(Exclude empty data files)



(Optionally)  
Make new directory  
with copies

## STITCHEE



()

### Reconstruct groups

Open new Dataset file in write mode

Create groups based on variable name delimiters

Create dimensions in appropriate groups

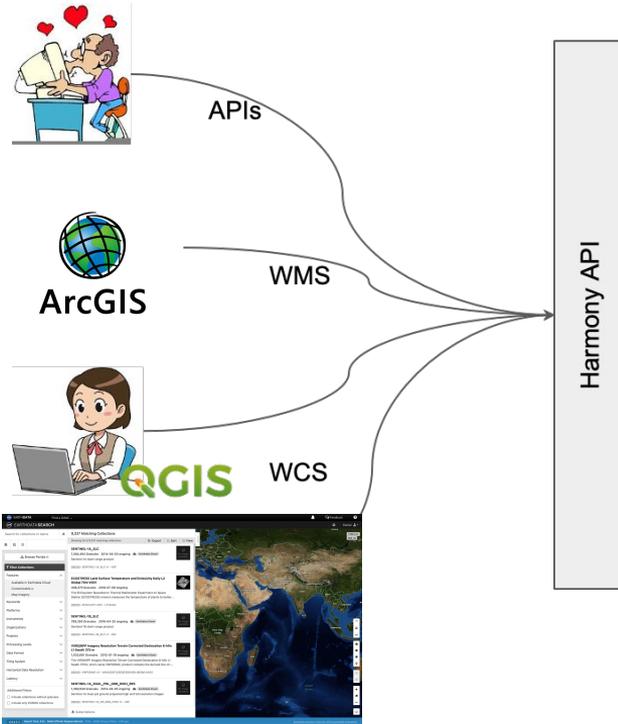
Create variables in appropriate groups, with copied attributes. Shuffle, chunk sizes, compression are newly determined.

Reconstruct group paths in coordinate attributes

**Open as an xarray Dataset**

**Save to disk**

# Functionality Overview



Earthdata Search

[search.earthdata.nasa.gov/search](https://search.earthdata.nasa.gov/search)

# A Chain of Services



Earthdata Harmony

Request



# A Chain of Services

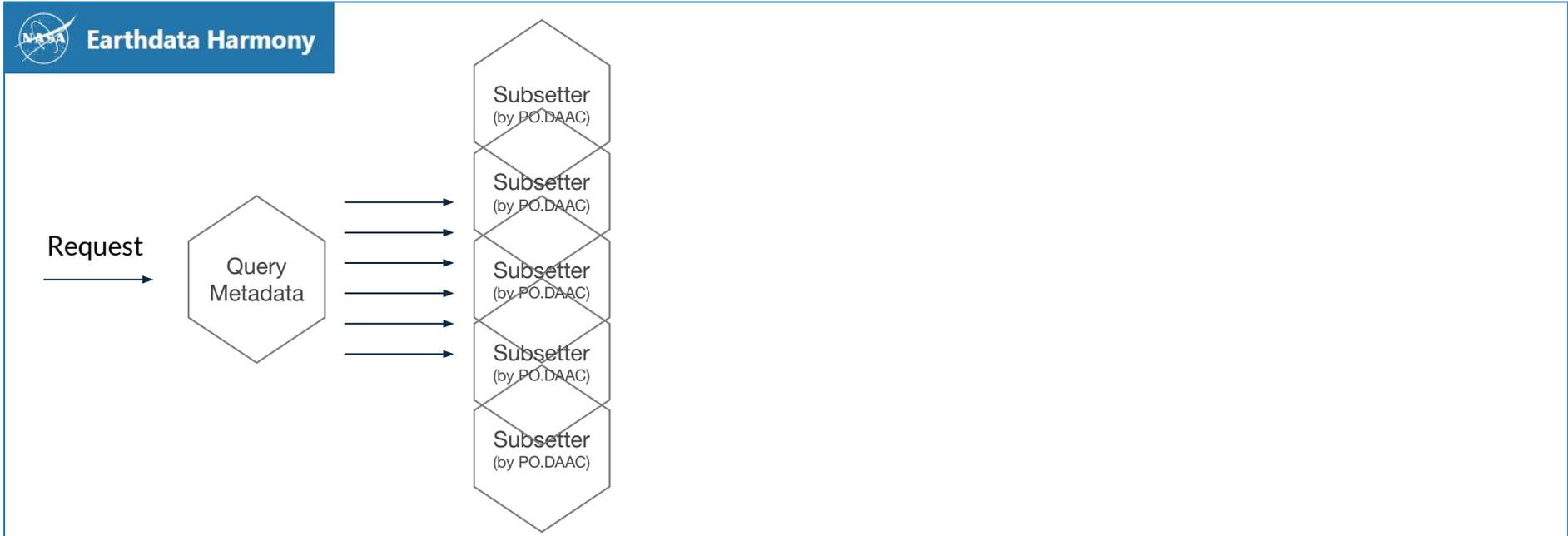


Earthdata Harmony

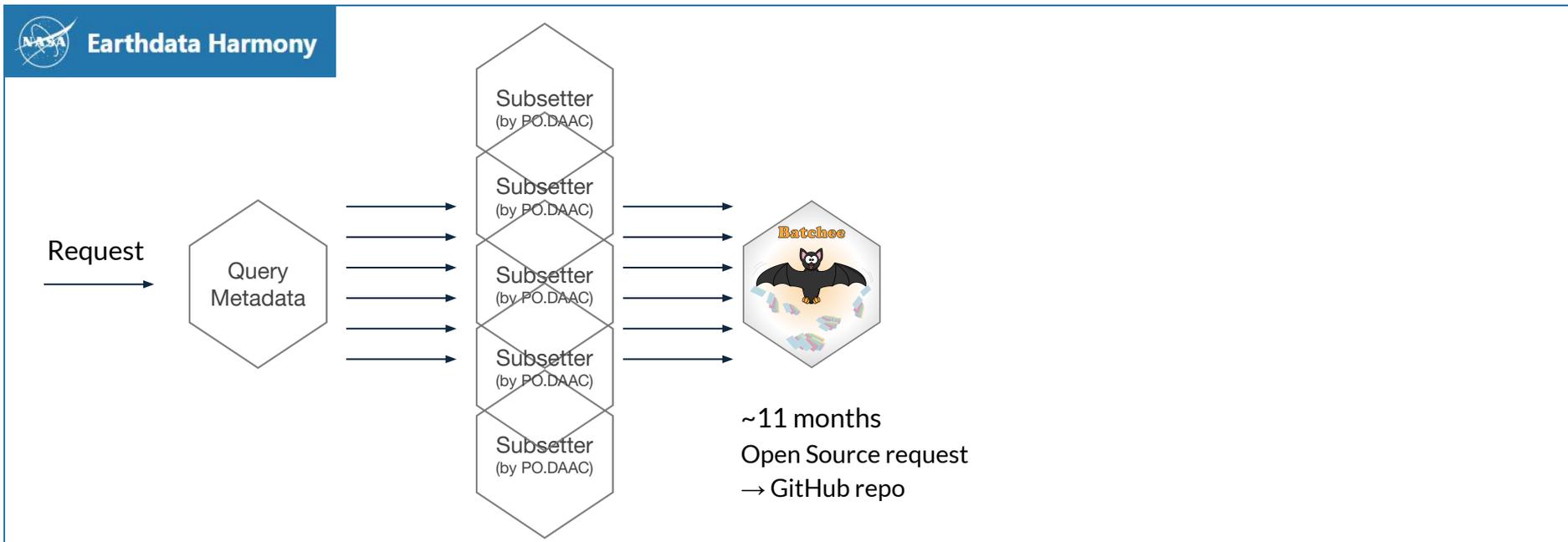
Request



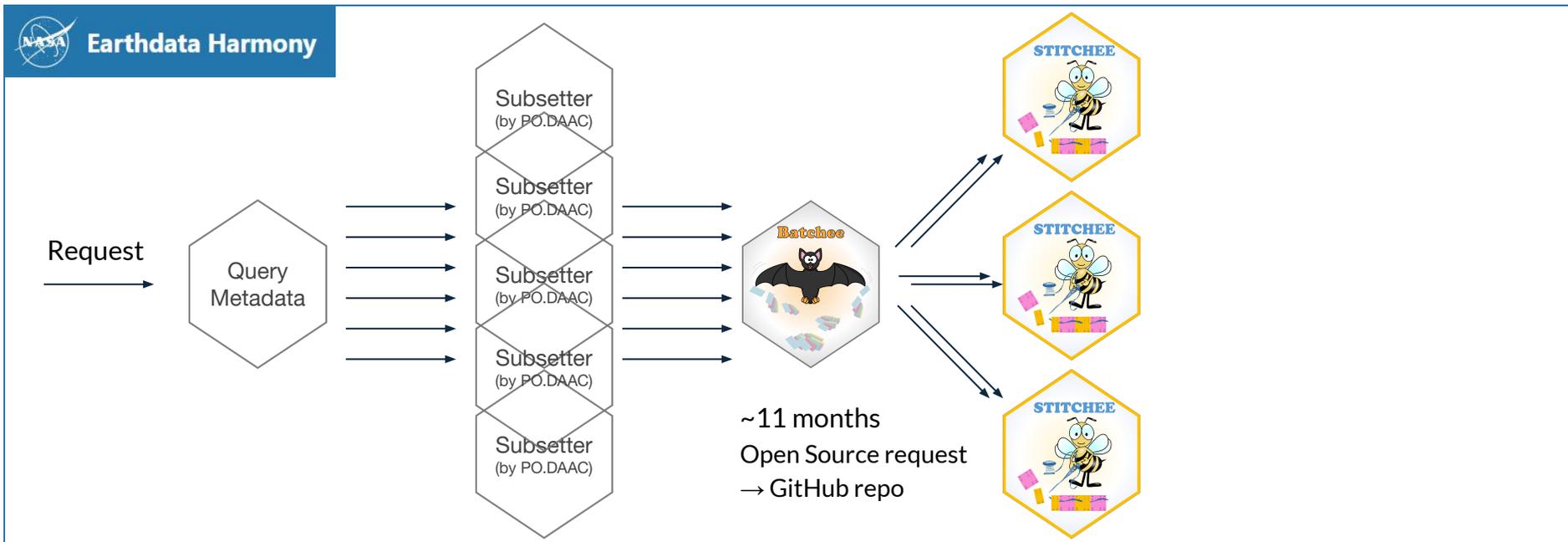
# A Chain of Services



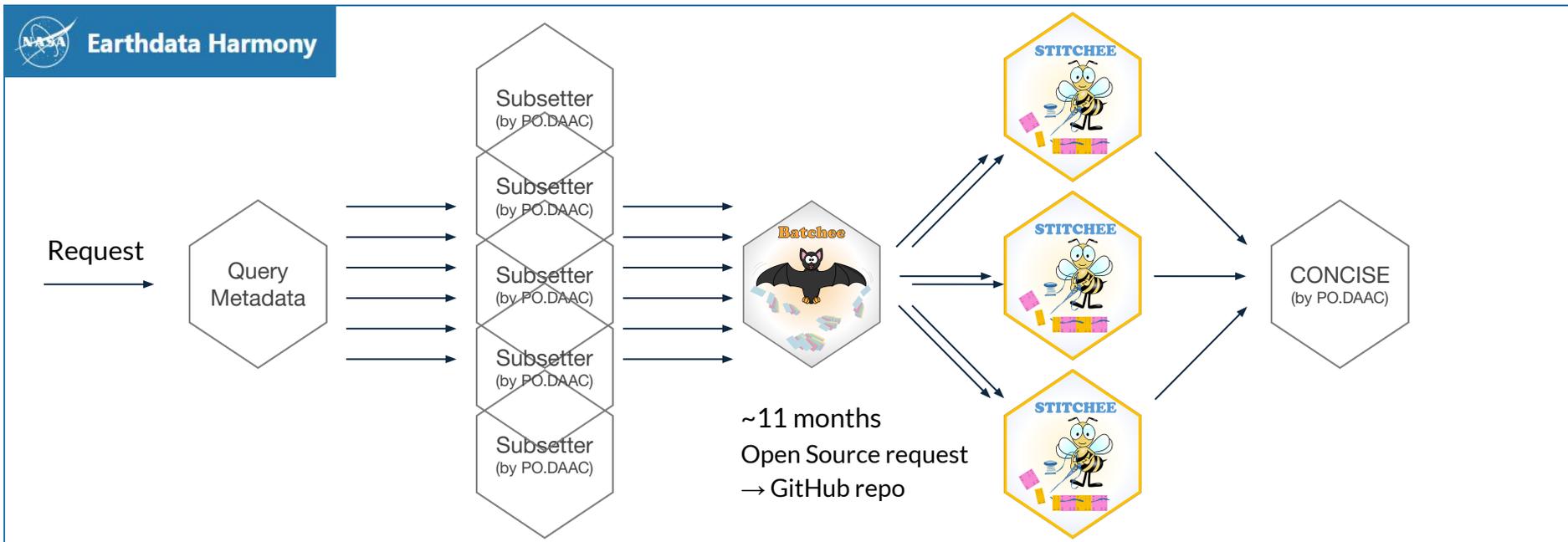
# A Chain of Services



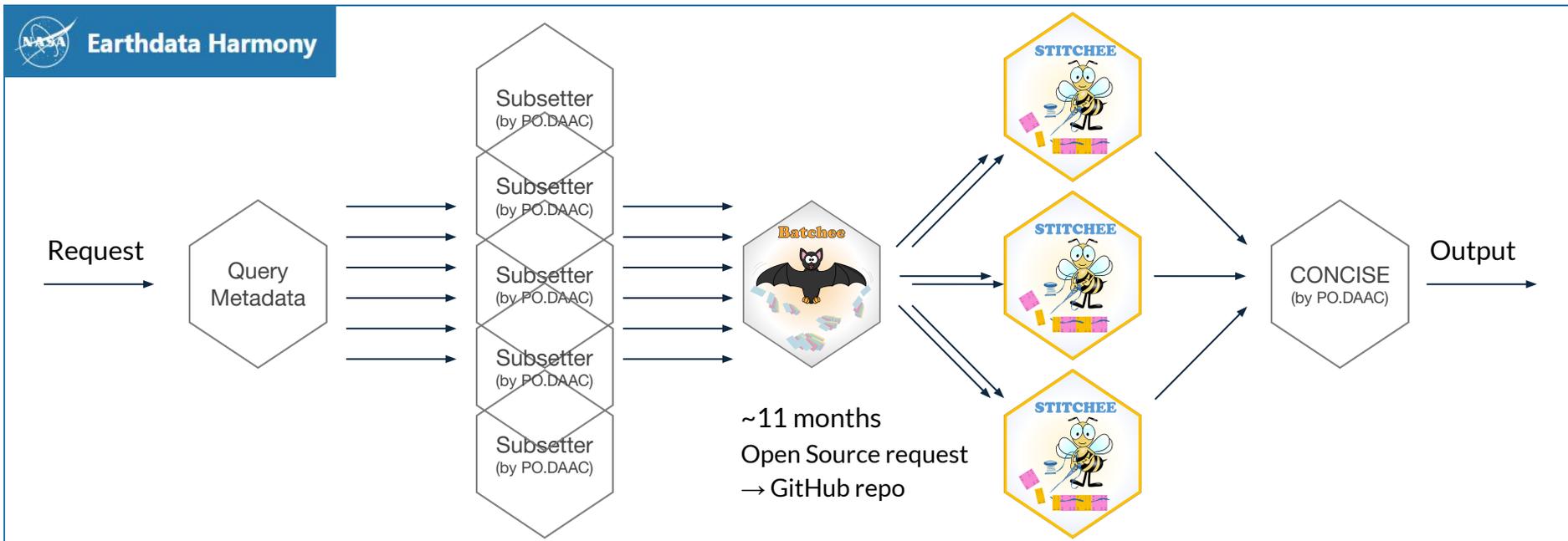
# A Chain of Services



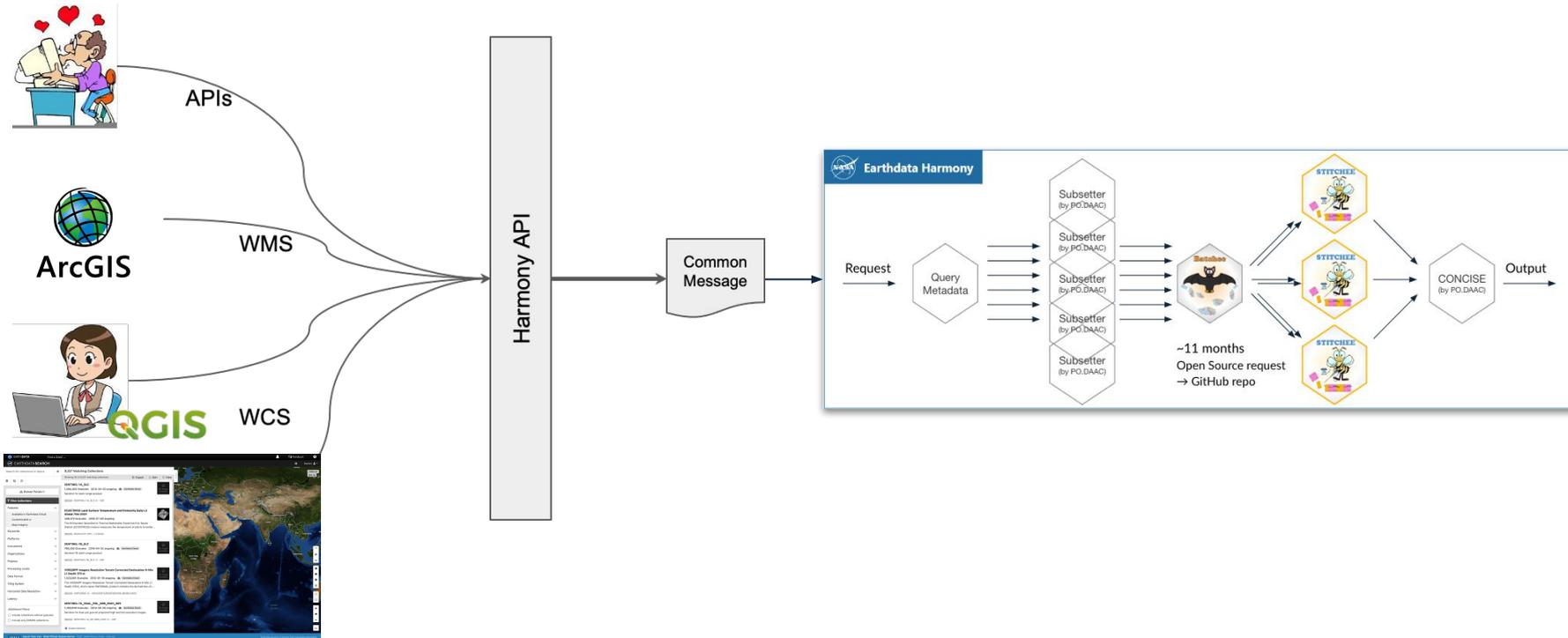
# A Chain of Services



# A Chain of Services



# Functionality Overview



Earthdata Search

[search.earthdata.nasa.gov/search](https://search.earthdata.nasa.gov/search)

# Summary



**TEMPO: unprecedented views of air quality over North America**

# Summary



**TEMPO: unprecedented views of air quality over North America**



**Transforming data on-demand** requires sophisticated data management & operations

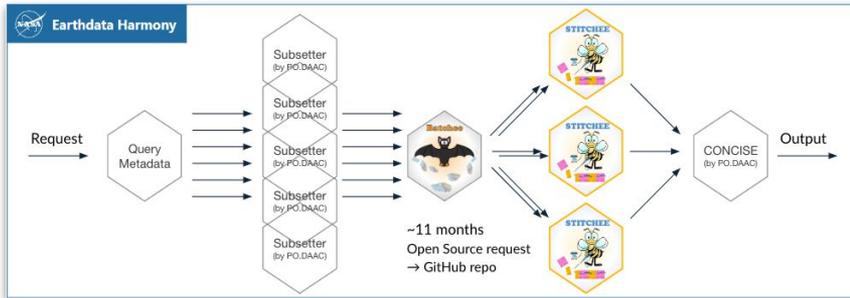
# Summary



TEMPO: unprecedented views of air quality over North America



Transforming data on-demand requires sophisticated data management & operations



Array stacking was insufficient for TEMPO's high temporal resolution, geostationary data

New approach developed and integrating into Harmony to concatenate data **along existing dimension**