Planes, Satellites, Trucks, Ships, and Fiber Optics - How to manage data coming from the bottom of the world

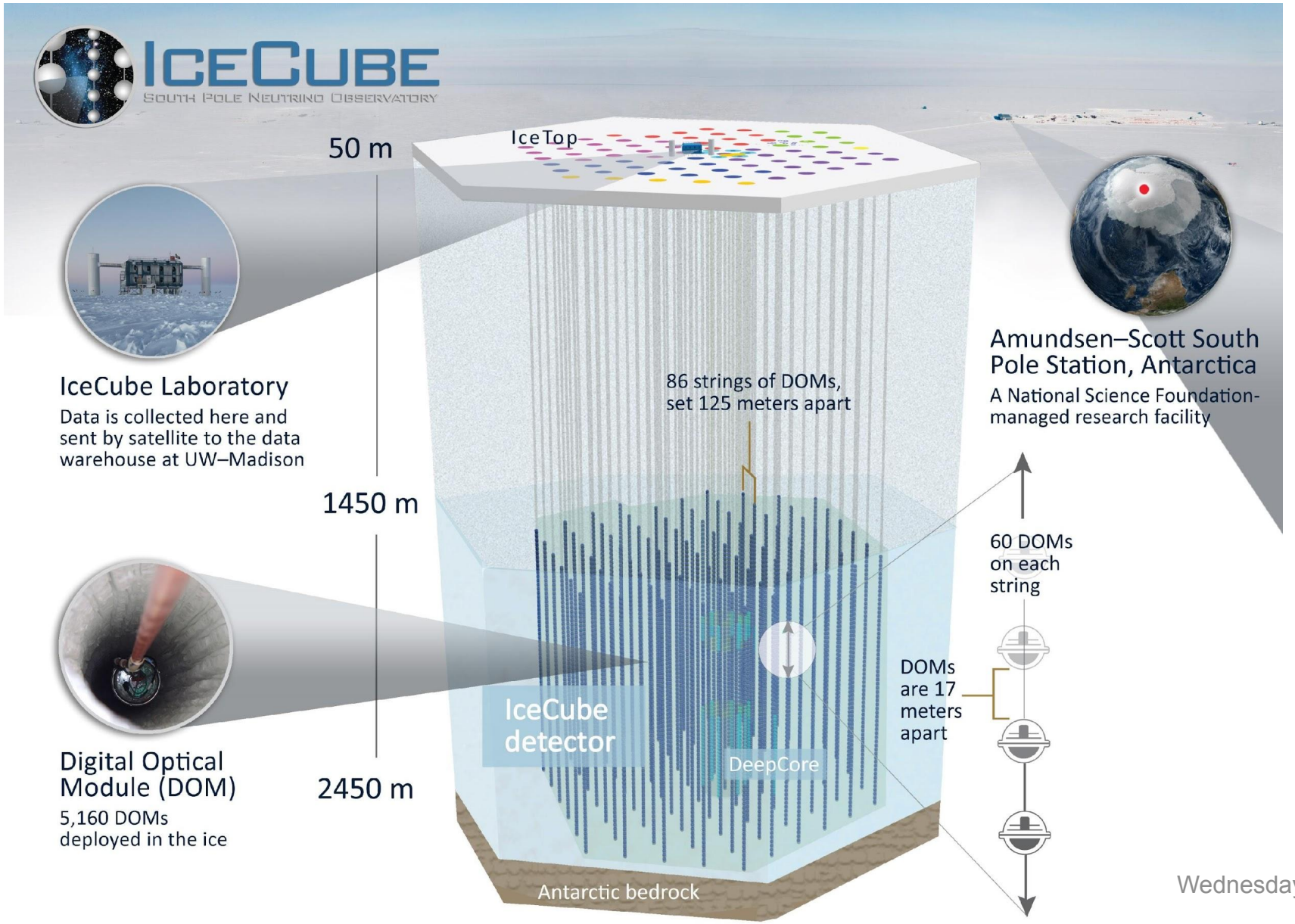
Benedikt Riedel
UW-Madison

CI Compass: Data Management: From Instrument to First Storage

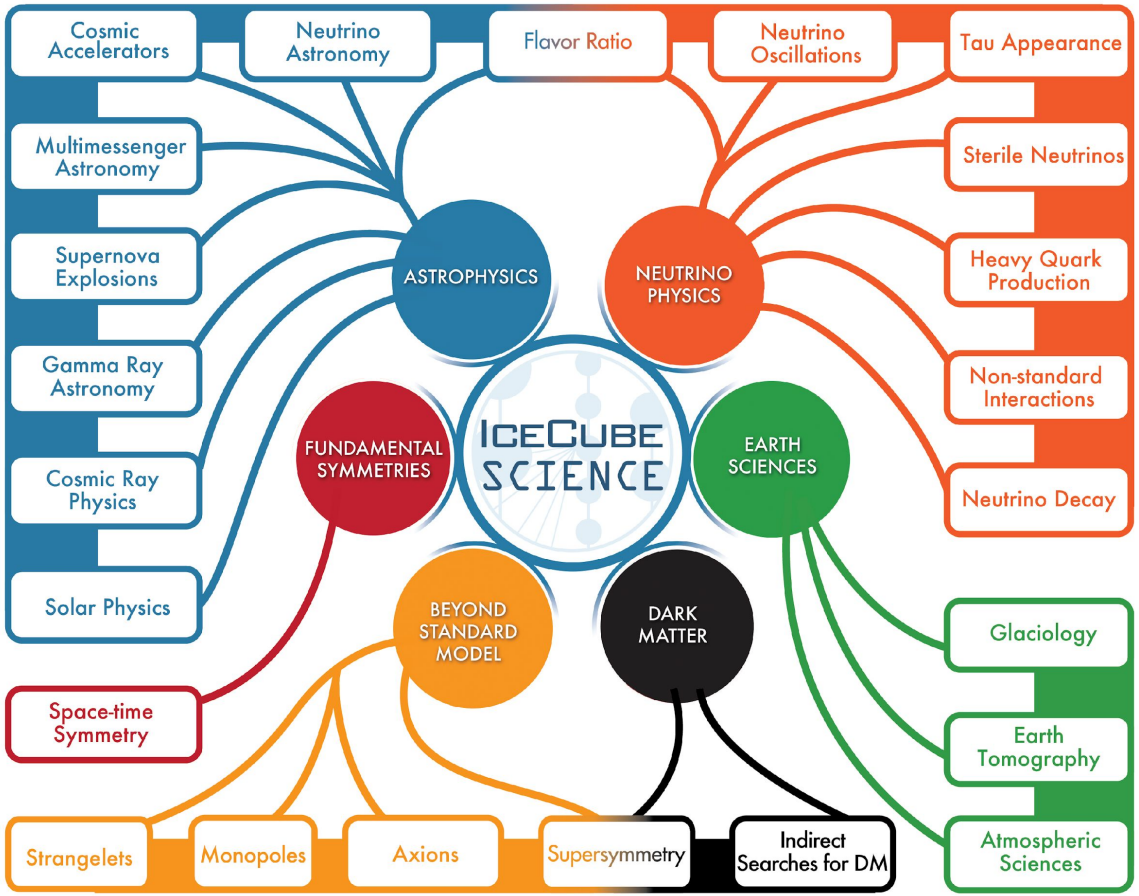
22 January 2025

IceCube
South Pole Neutrino Observatory

IceCube

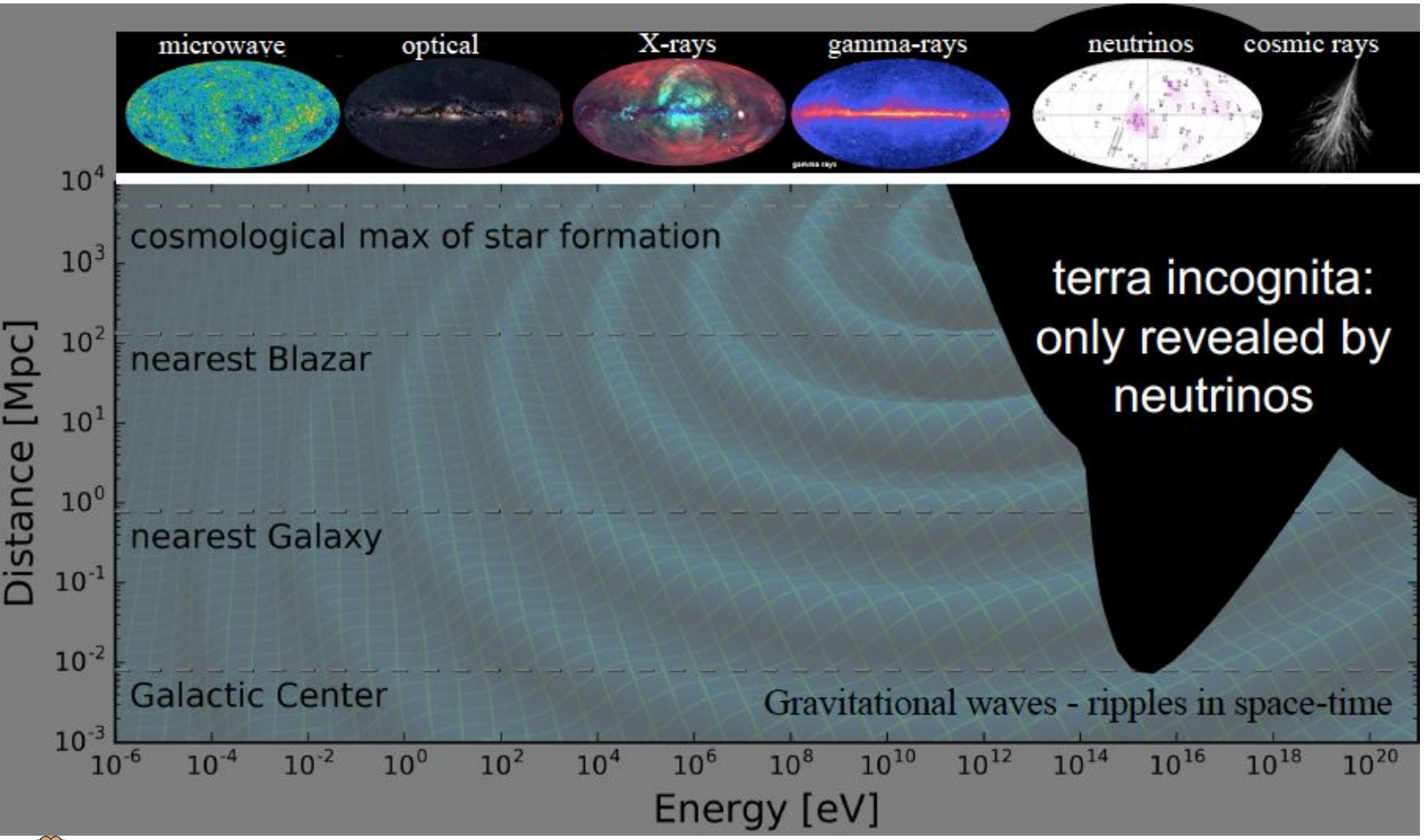

IceCube
South Pole Neutrino Observatory
50 m
IceTop
IceCube Laboratory
Data is collected here and sent by satellite to the data warehouse at UW–Madison
86 strings of DOMs, set 125 meters apart
Amundsen–Scott South Pole Station, Antarctica
A National Science Foundation-managed research facility
1450 m
60 DOMs on each string
DOMs are 17 meters apart
IceCube detector
DeepCore
Digital Optical Module (DOM)
5,160 DOMs deployed in the ice
2450 m
Antarctic bedrock

IceCube Science



- Novel instrument in multiple fields
- Broad science abilities, e.g. astrophysics, particle physics, and earth sciences
- Lots of data that needs to be processed in different ways
- Lots of simulation that needs to be generated

IceCube Science – Why neutrinos?



- 20% of universe is dark to “traditional” astronomy, i.e. using electromagnetic waves/light
- Need a new set of “messengers” –Gravitational Waves and Neutrinos

A History of Neutrino Astronomy in Antarctica

1988 — Telescope in the Ice Envisioned

2000 — AMANDA Completed

2001 — Atmospheric Neutrinos Detected

2011 — IceCube Completed

2013 — Astrophysical Neutrinos Discovered

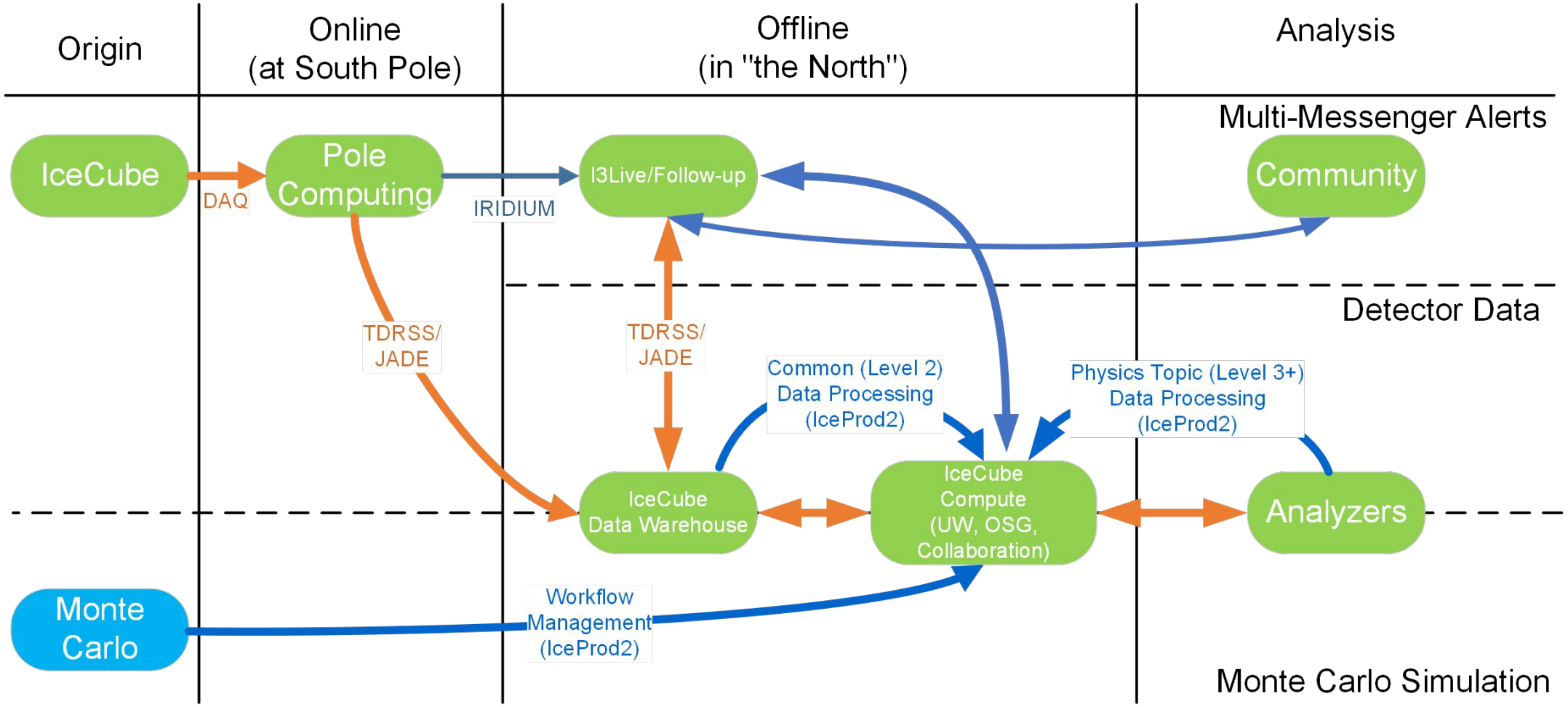
2018 — First Source TXS 0506+056 Identified

2021 — Glashow Resonance Neutrino Identified

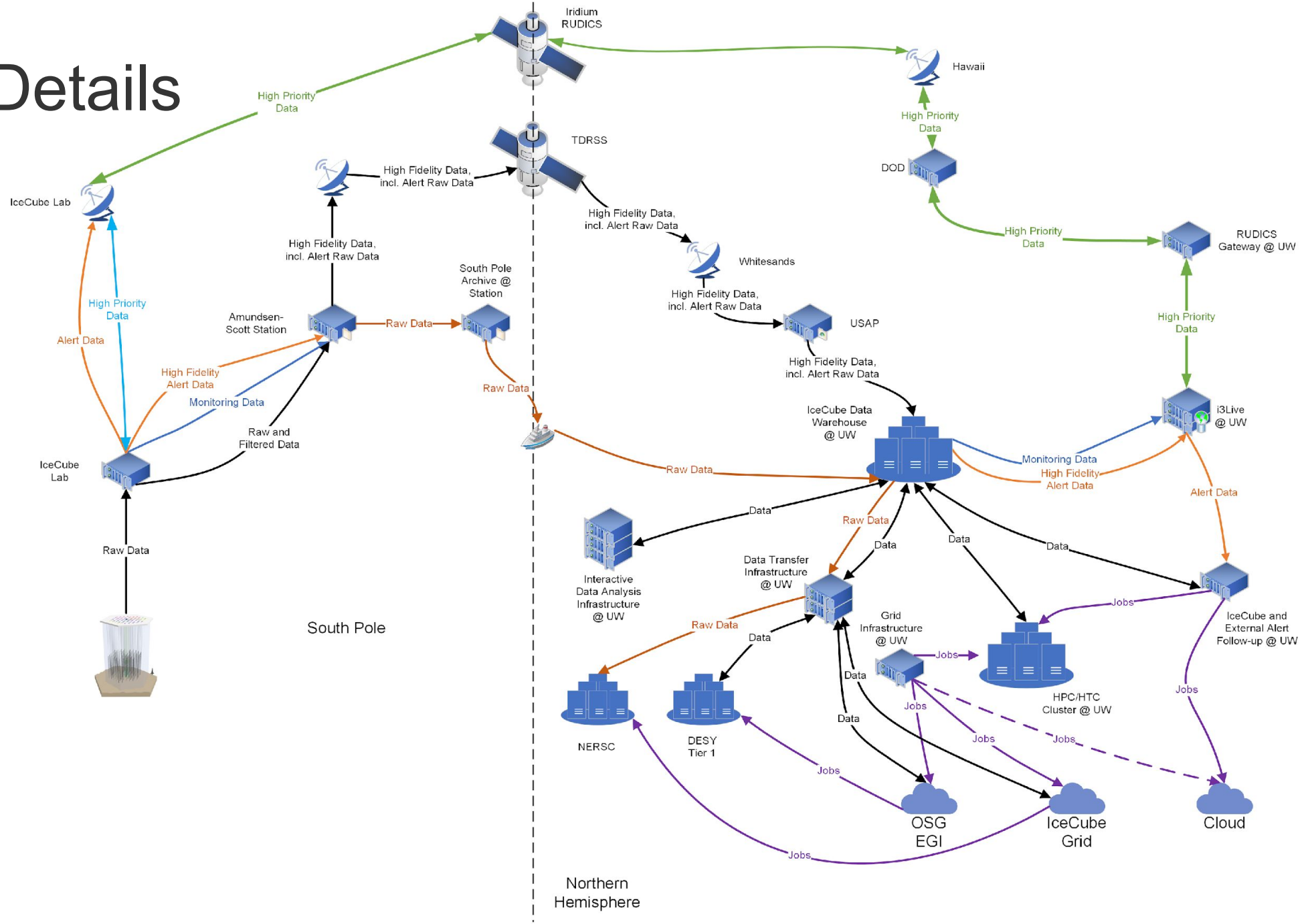
2022 — Second Source NGC 1068 Identified

2023 — Third Source Milky Way Identified

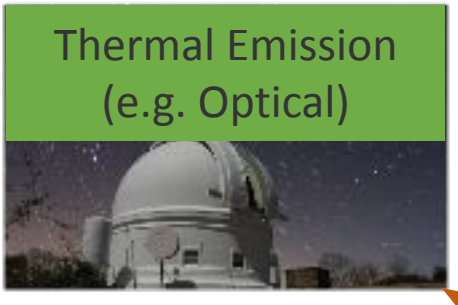
Summary

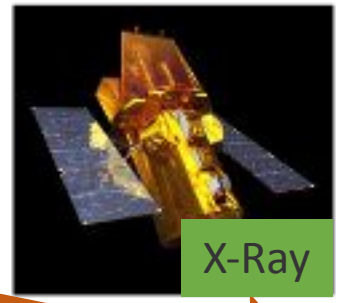

Origin
Online
(at South Pole)
Offline
(in "the North")
Analysis
Multi-Messenger Alerts
IceCube
DAQ
Pole
Computing
IRIDIUM
I3Live/Follow-up
Community
TDRSS/
JADE
TDRSS/
JADE
Detector Data
Common (Level 2)
Data Processing
(IceProd2)
Physics Topic (Level 3+)
Data Processing
(IceProd2)
IceCube
Data Warehouse
IceCube
Compute
(UW, OSG,
Collaboration)
Analyzers
Monte
Carlo
Workflow
Management
(IceProd2)
Monte Carlo Simulation

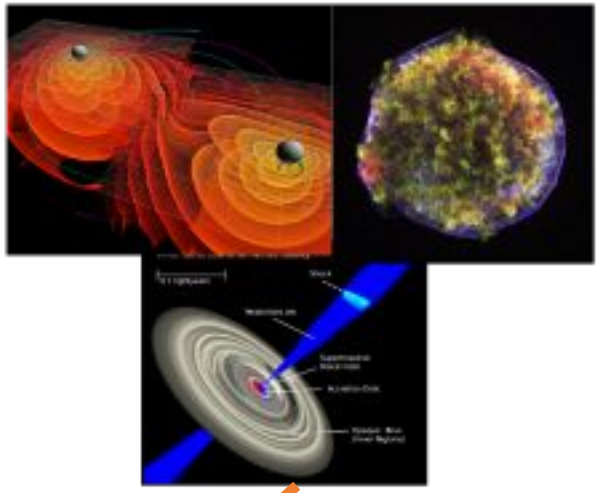
Gory Details



Data Routes – Data







NASA GCN

UW-Madison

NERSC

TACC
TEXAS ADVANCED COMPUTING CENTER

ESnet
ENERGY SCIENCES NETWORK

INTERNET2

GÉANT
Networks • Services • People

DESY.

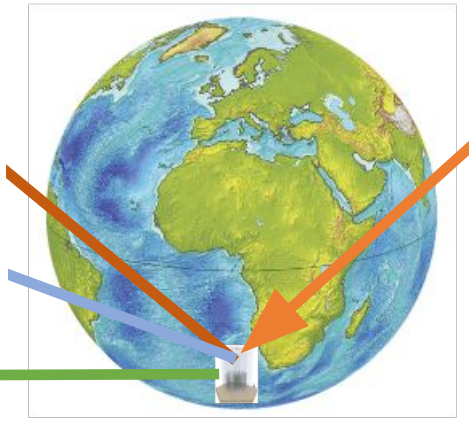
IRIDIUM

TDRSS

Prio. Data

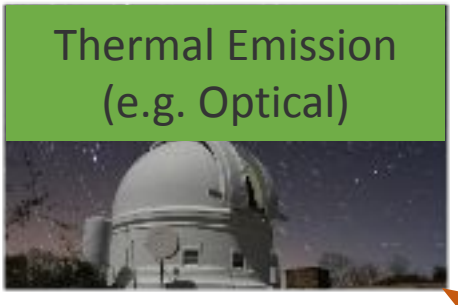
Filtered Data

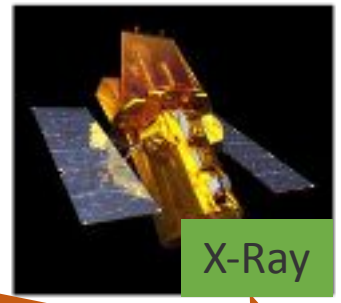
Archival Data



V

Data Routes – MMA



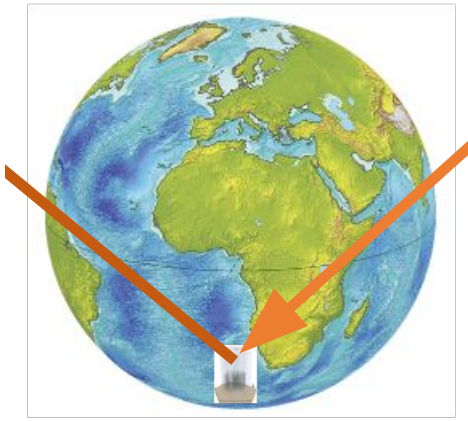


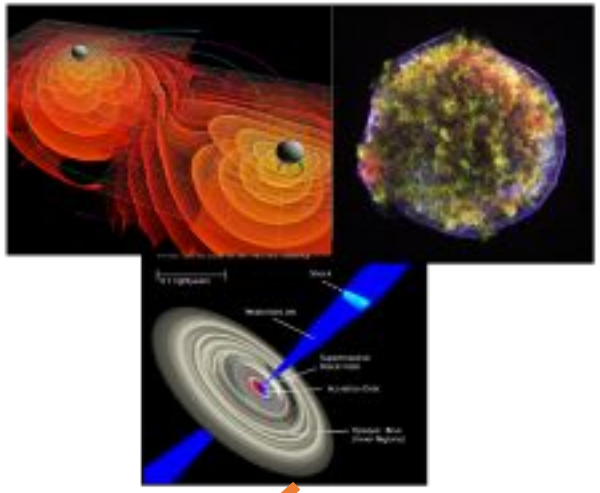
NASA GCN

UW-Madison

IRIDIUM

Prio. Data

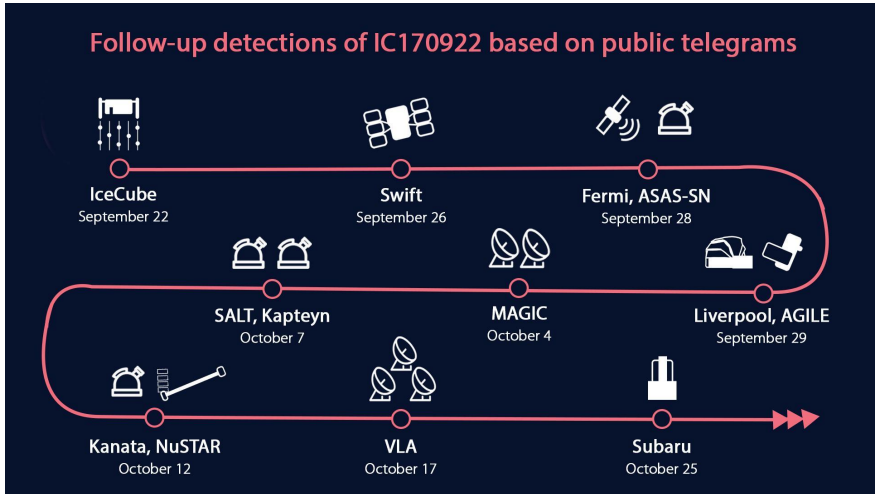


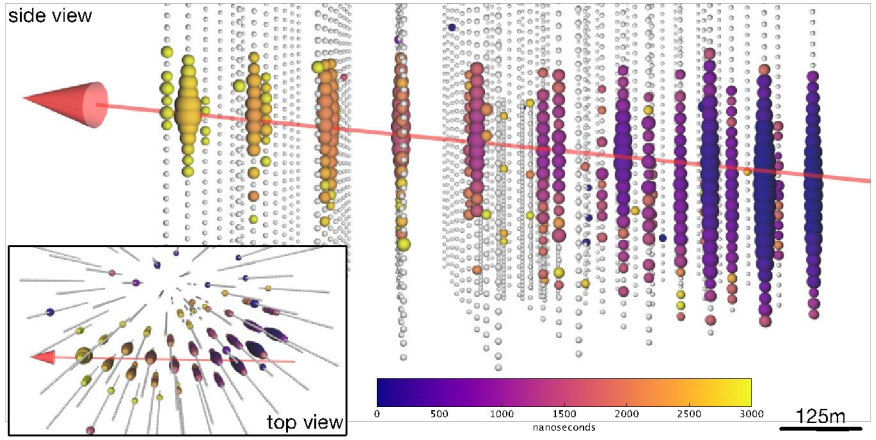


ν

- Alert Stream or Multi-Messenger Astrophysics (MMA)
- 24/7 connection at limited bandwidth (email with small attachment)
- 24/7 monitoring and communication with experiment operators

IceCube Science – Multi-Messenger Astrophysics

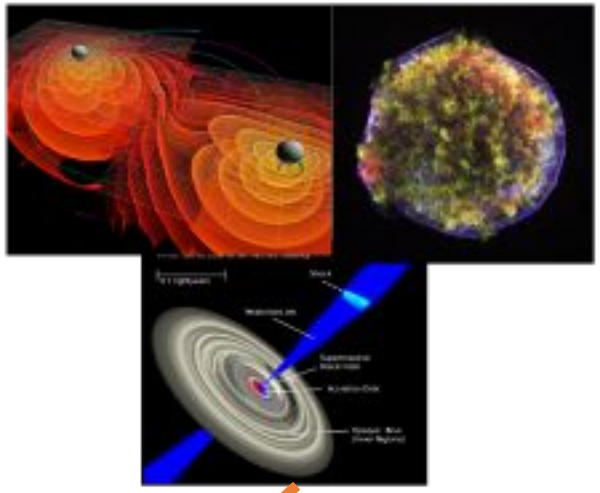


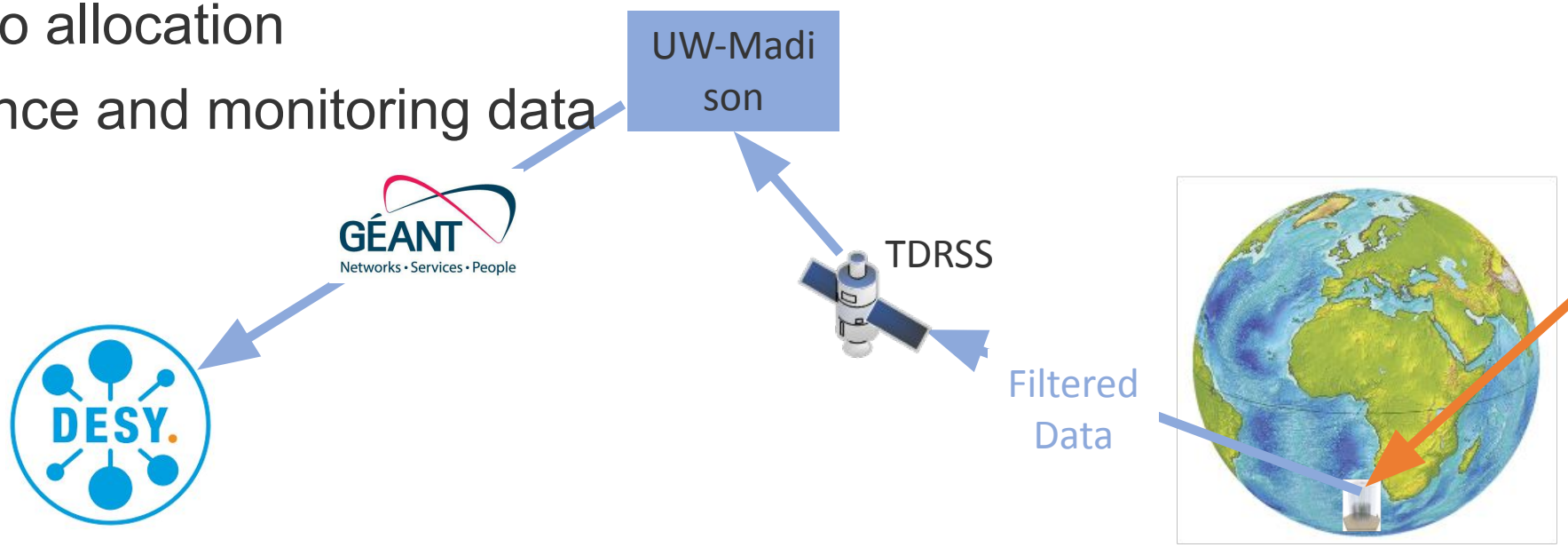


- Multi-Messenger Astrophysics (MMA)
  - Observing astrophysical phenomena with more than one “messenger” (gravitational waves, neutrinos, EM)
  - One of NSF’s Big 10 Ideas
- IceCube detected an event that came from Blazar TXS 0506+056
  - Follow up observations by several observatories/telescopes showed signal
  - Back catalog showed access for this source
- Fast response to alerts requires significant computing resources

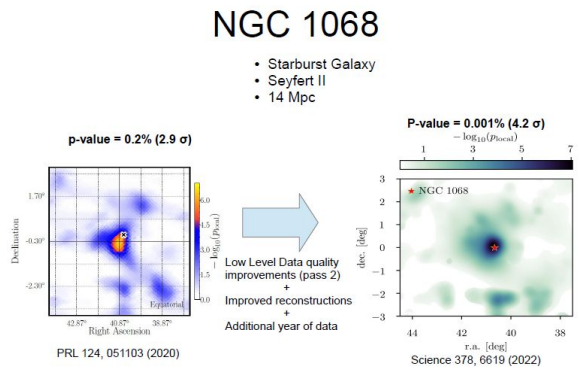
Data Routes – Filtered Data

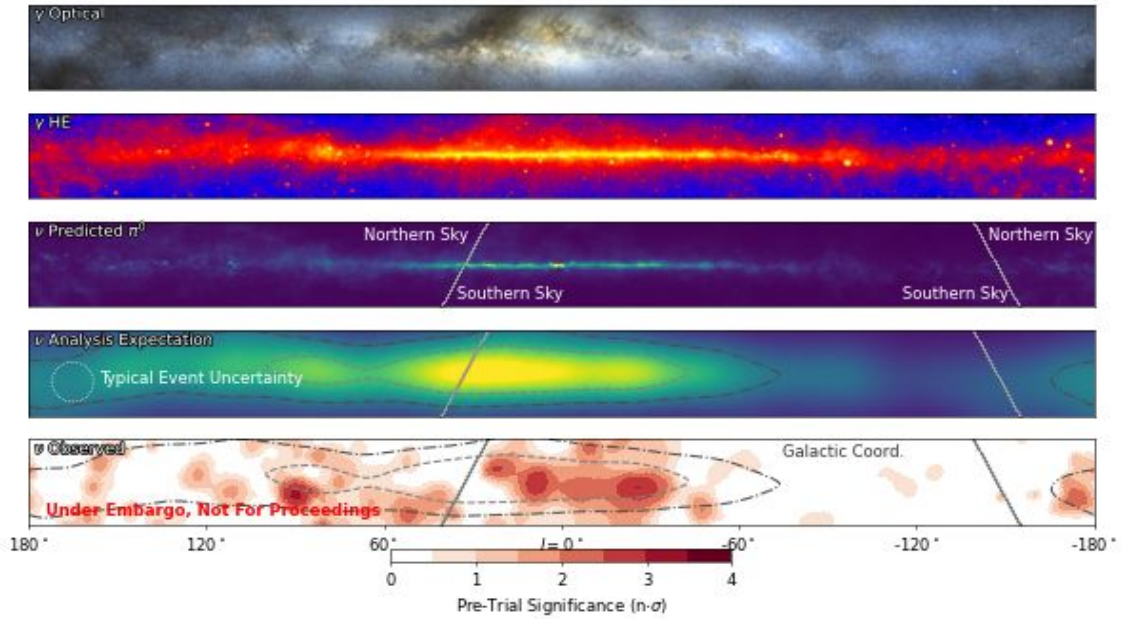
- O(daily) connection for with shared bandwidth (500 GB/day) with 110 GB/day allocated
- On-site lossy compression to fit into allocation
- Science and monitoring data





A new view of the Cosmos



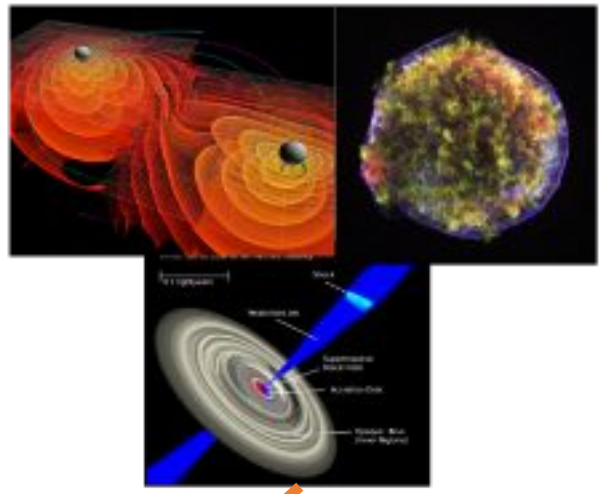


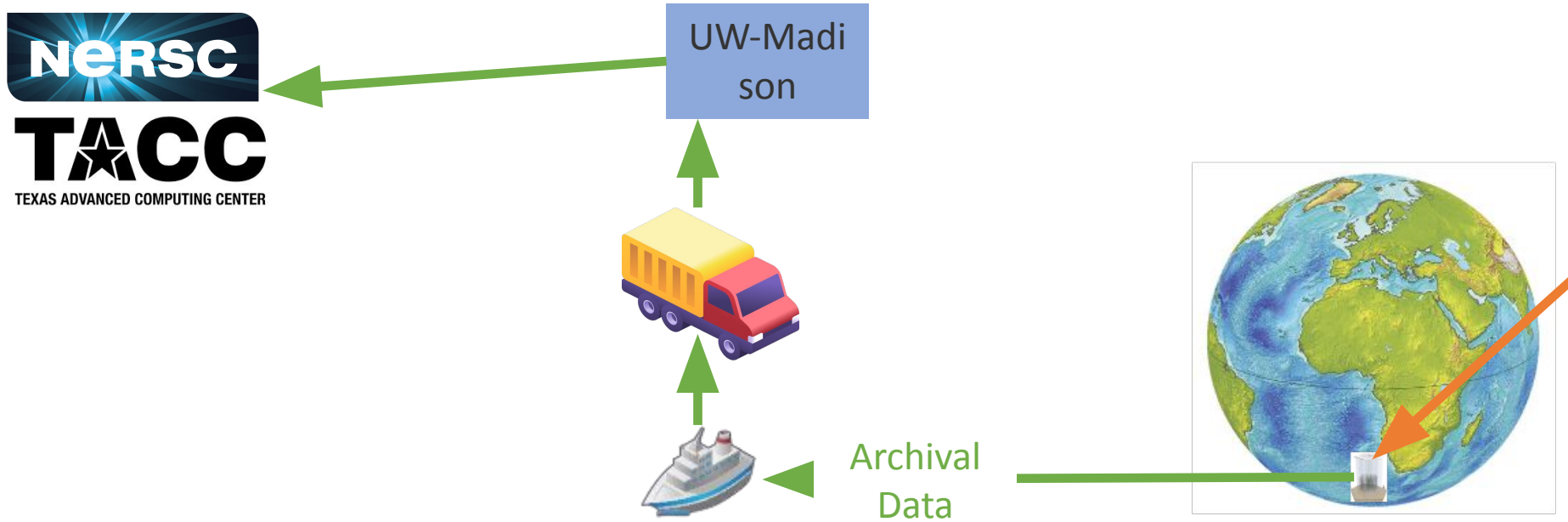
- Using “line-shaped” events to check locations in the sky for “above threshold”
- Purely statistical method
- Second High Energy Neutrino Point Source

- Using “ball-shaped” events for “large” (e.g. Milky Way) or isolated sources a good option for Southern Sky point source search – Clear differentiation from background
- Maximum-Likelihood method for cascade pointing insufficient to find a source – Using BDT and CNN to find and reconstruct cascade events
- Third High Energy Neutrino Point Source! – Galactic Plane

Data Routes – Archives

- Yearly shipping of “all” data from South Pole to “North” for archival
- Allow future processing with updated detector understanding
- Potential to find “new” results

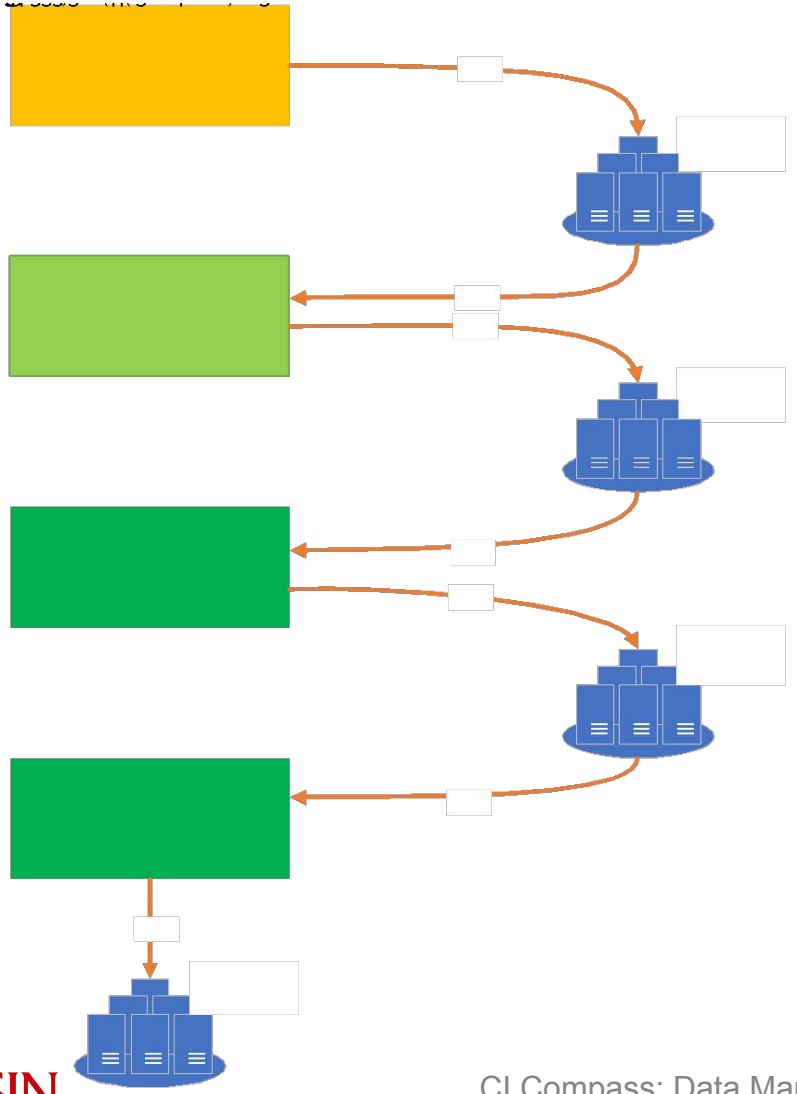




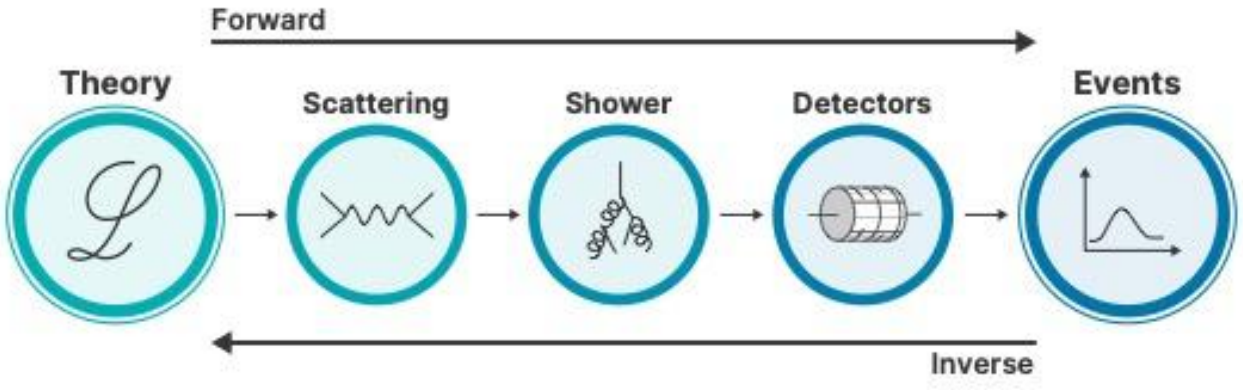
?

- Beyond the Standard Model Physics
  - Can we improve our sensitivity?
- Explore new data analysis methods – Machine Learning
  - etc.

Monte Carlo Simulation



Signal and Background Simulation
(Neutrinos and Cosmic Ray Muons)



Comparing Data to Simulation

- Monte Carlo simulation is required to test theory to data
- Need to produce and store simulation data

Thank you!

Questions?