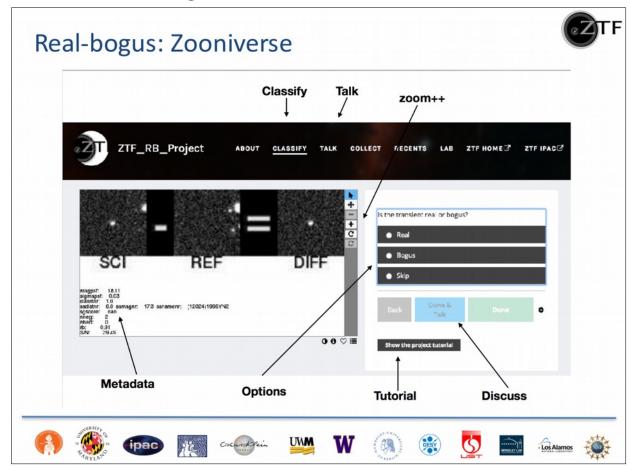# KDIG Thursday May 31

Matthew Graham: *ML aspects of ZTF*

- ML[classification] => deal with training data

- Storage

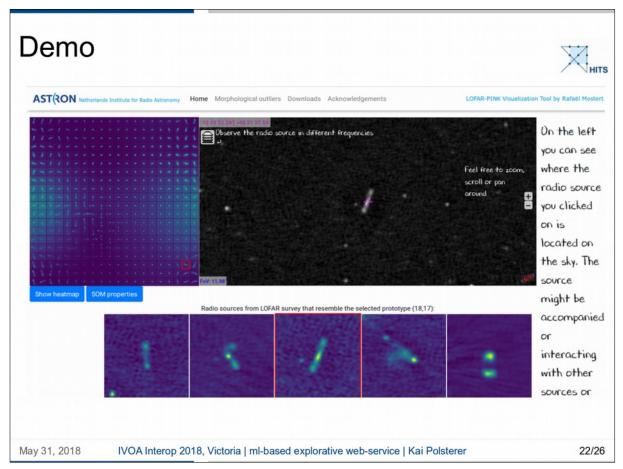- Provenance

- Formats

KDIG Thursday May 31

Kai Polsterer: *ML based, explorative web service*

- ML[dimensionality reduction] => deal with found relations
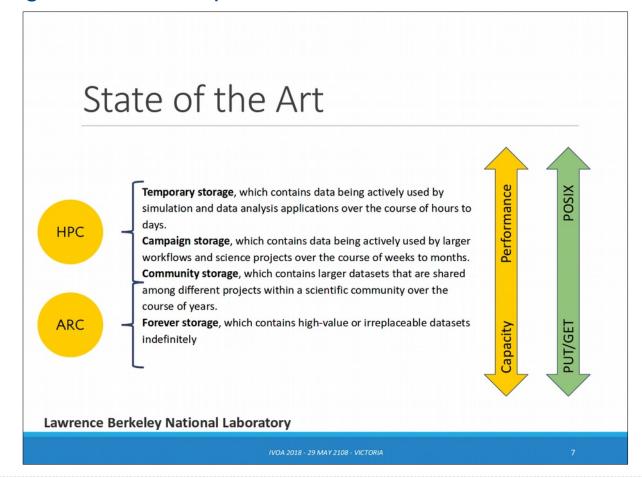
- Storage

- Provenance

- Formats

KDIG Thursday May 31

Giuliano Taffoni: *HPDA: a storage view on data analytics*

- Data analytics => storage/hardware requirements

- Change of paradigm

State of the Art

HPC
ARC

Temporary storage, which contains data being actively used by simulation and data analysis applications over the course of hours to days.
Campaign storage, which contains data being actively used by larger workflows and science projects over the course of weeks to months.
Community storage, which contains larger datasets that are shared among different projects within a scientific community over the course of years.
Forever storage, which contains high-value or irreplaceable datasets indefinitely

Performance — Capacity
POSIX — PUT/GET

Lawrence Berkeley National Laboratory

IVOA 2018 - 29 MAY 2108 - VICTORIA

7

KDIG Summary

Are we ready for machine learning?

- Storage of ML data

- Provenance in ML analysis

- Formats to store / exchange ML related data

- Change of storage paradigm

What are we missing?