



DARK ENERGY
SURVEY



Ciemat

Handling the Dark Energy Survey

Ignacio (Nacho) Sevilla

CIEMAT, Madrid -- UIUC Visiting Scholar
on behalf of the DES Collaboration



DARK ENERGY
SURVEY



Ciemat

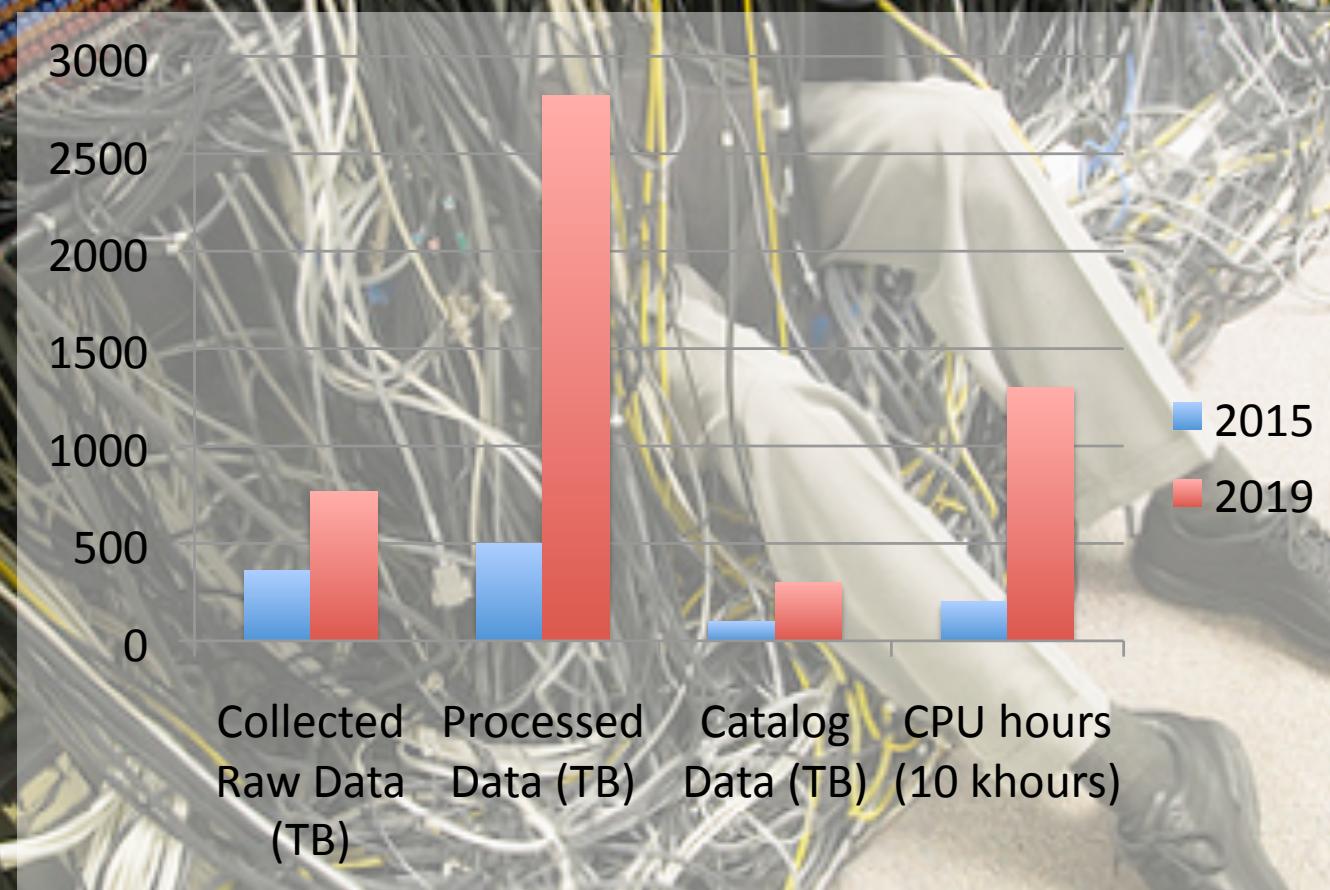
Handling the Dark Energy Survey

1. Processing
2. Visualization
3. Analysis
4. Machine Learning

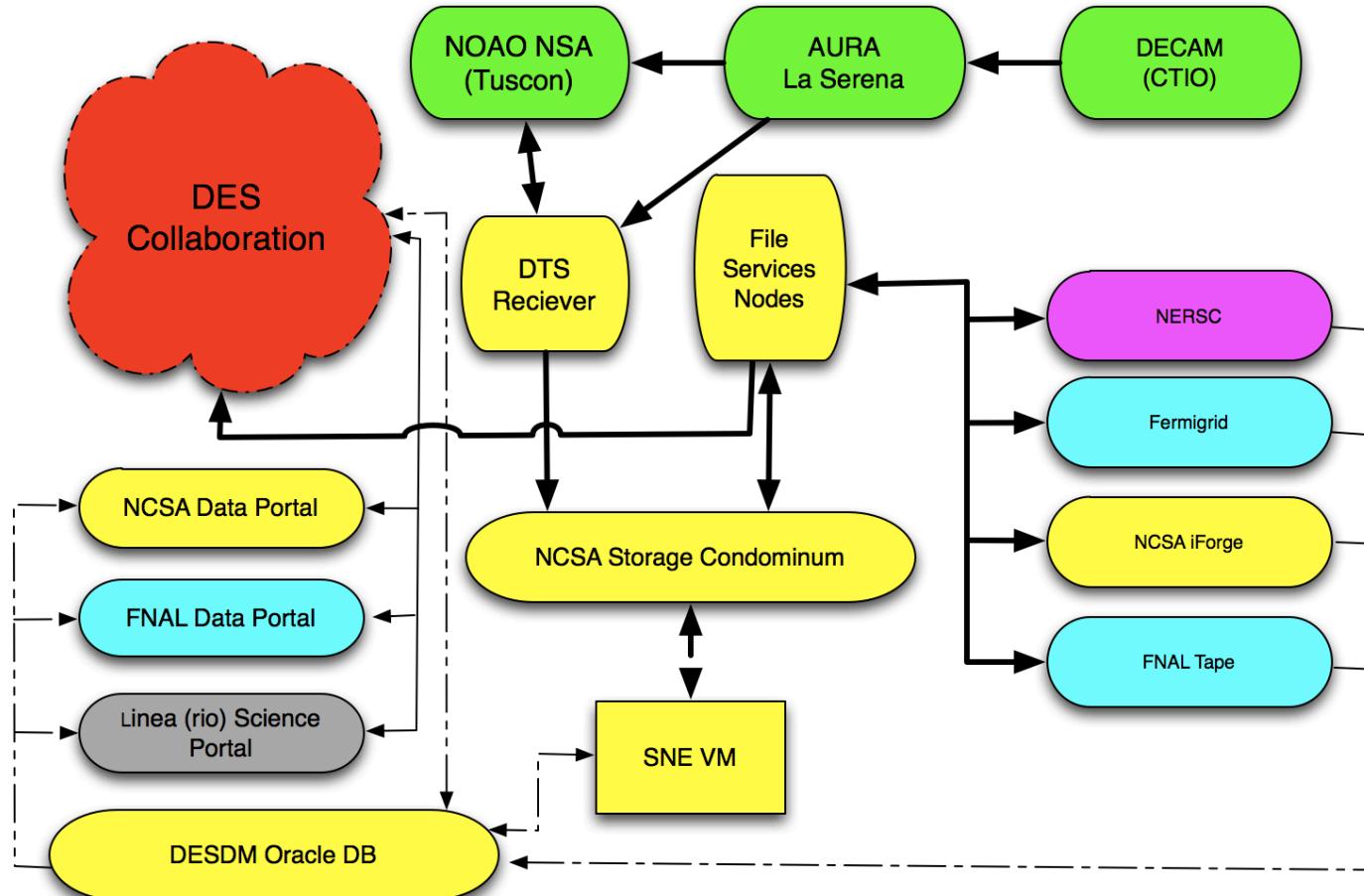
DES is a wide and deep photometric survey

- Survey project using **4 complementary techniques**:
 - I. Cluster Counts
 - II. Weak Lensing
 - III. Large-scale Structure
 - IV. Supernovae
- Two multiband (**photometric**) surveys using 4m Blanco Telescope:
5000 deg² grizY to 24th mag AB griz
30 deg² repeat (SNe)
- Just finished observing second year (out of five)
- **Data Management system (NCSA)** to handle 300 GB/night, 525 nights, 4 PB data
(raw+(re-)processed) by 2018, 100s TB catalogs (not including sims)

But DES is not Big Data!



DES is being processed using a variety of resources



D. Petravick, DESDM (NCSA) see R.Gruendl's talk at DECam workshop
I.S. et al. arXiv:1109.6741; J.Mohr et al. arXiv:1207.3189

Image verification is crowdsourced to Collaboration

The screenshot shows the homepage of the DES exposure checker. At the top, there is a navigation bar with links to 'DES exposure checker', 'Viewer', 'Tutorial', 'FAQ', 'Statistics', 'API', 'Gallery', 'Hodge-podge', 'SVA1', and 'Y1A1'. The 'Y1A1' tab is highlighted. Below the navigation bar, the title 'DES exposure checker' is displayed in large bold letters. A subtitle below it reads 'See how real DES images look. Discover flaws we would otherwise have missed.' and 'Make our data better!'. A blue 'Let's go' button is located at the bottom left of this section. To the right, there is a large number '9044' with the text 'images checked' underneath. Below that, a green '0' is followed by the word 'today'.

What's this good for?

We all want to do science with the DES data. However, no data set is perfect, so we need to identify and account for imaging artifacts in our data. This application helps in two ways:

- It gives you convenient access to the DES annual release finalcut images and shows the processing steps that are already in place to identify and correct for known problems.
- It enables everyone to search for undetected artifacts. We gather your submissions, analyze them, and feed them back to the science working groups and DESDM. This way, we can refine the processing pipelines to catch the flaws that we know are there.

How can I start?

Check out our [Tutorial](#) and you'll know what to do in no time.

How are we doing so far?

The basic [statistics](#) are computed in real time. More detailed analyses will become available once we have gathered enough submissions, but you can go to our [API page](#) and download the anonymized submissions right now.

How can I give feedback?

The code is on [github](#). Please open a new [Issue](#) to let us know what we could add, improve, ...

If you find particular failure cases and you think they may constitute a new class of artifacts, please go to the [DESDM users wiki](#), where we list and discuss such cases.

Top contributors for Y1A1

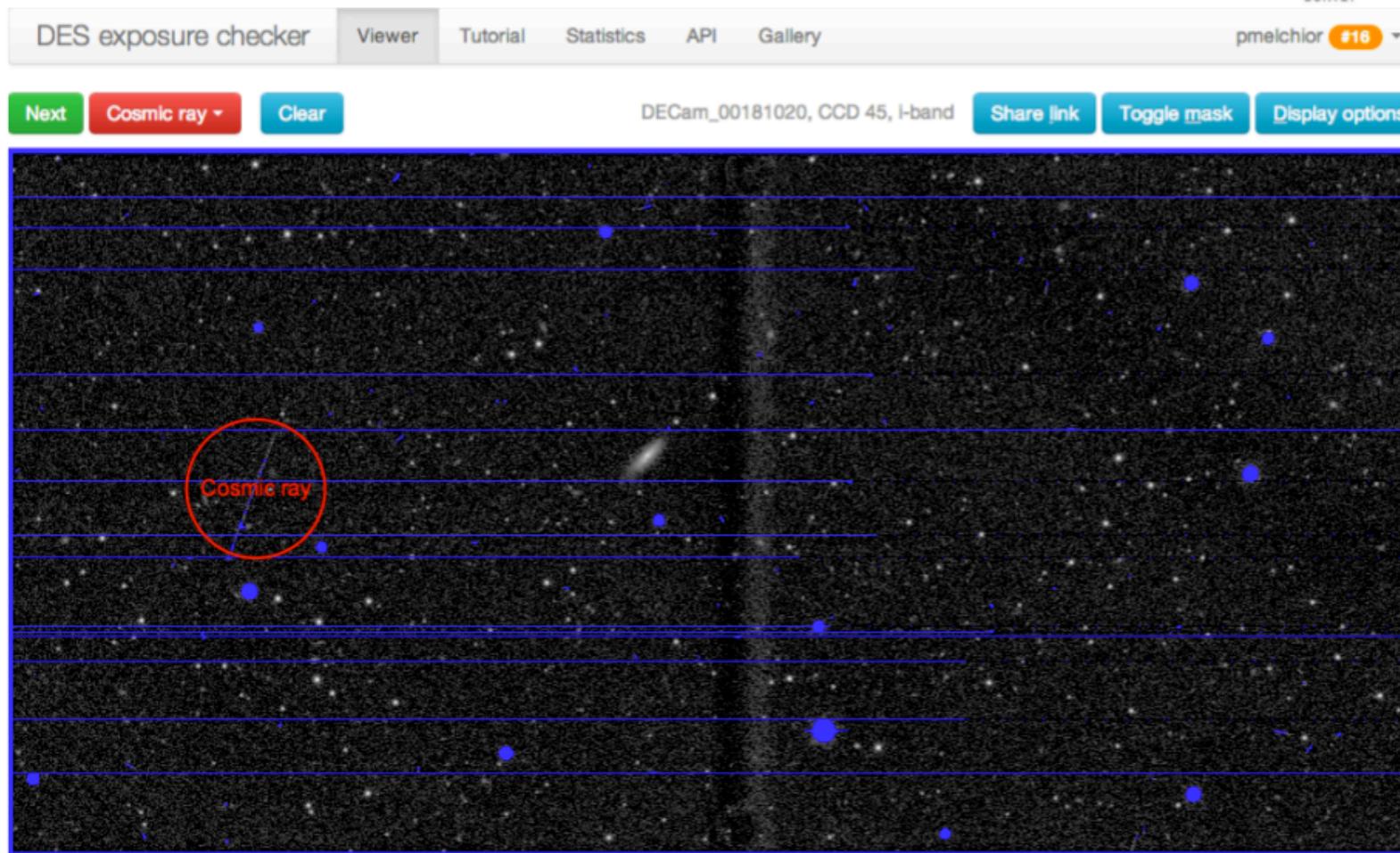
# 1	rpoulton		2018
# 2	yzhang		1471
# 3	kadrlica		1190
# 4	padsmith		1046
# 5	bellido		1001
# 6	aascott		976
# 7	pmelchior		260
# 8	esheldon		235
# 9	jhebsy		233
# 10	erykoff		137
# 11	pmartini		102
# 12	apapadopoulos		83
# 13	rgm		66
# 14	nsevilla		57
# 15	jdietrich		53

P. Melchior (OSU) (peter.m.melchior@gmail.com)
E. Sheldon (BNL), A. Drlica-Wagner (Fermilab)

- 1. Additional QA of single-epoch images**
- 2. Scientists get acquainted with characteristics of instrument.**
- 3. Possible science cases could benefit directly**

VISUALIZING

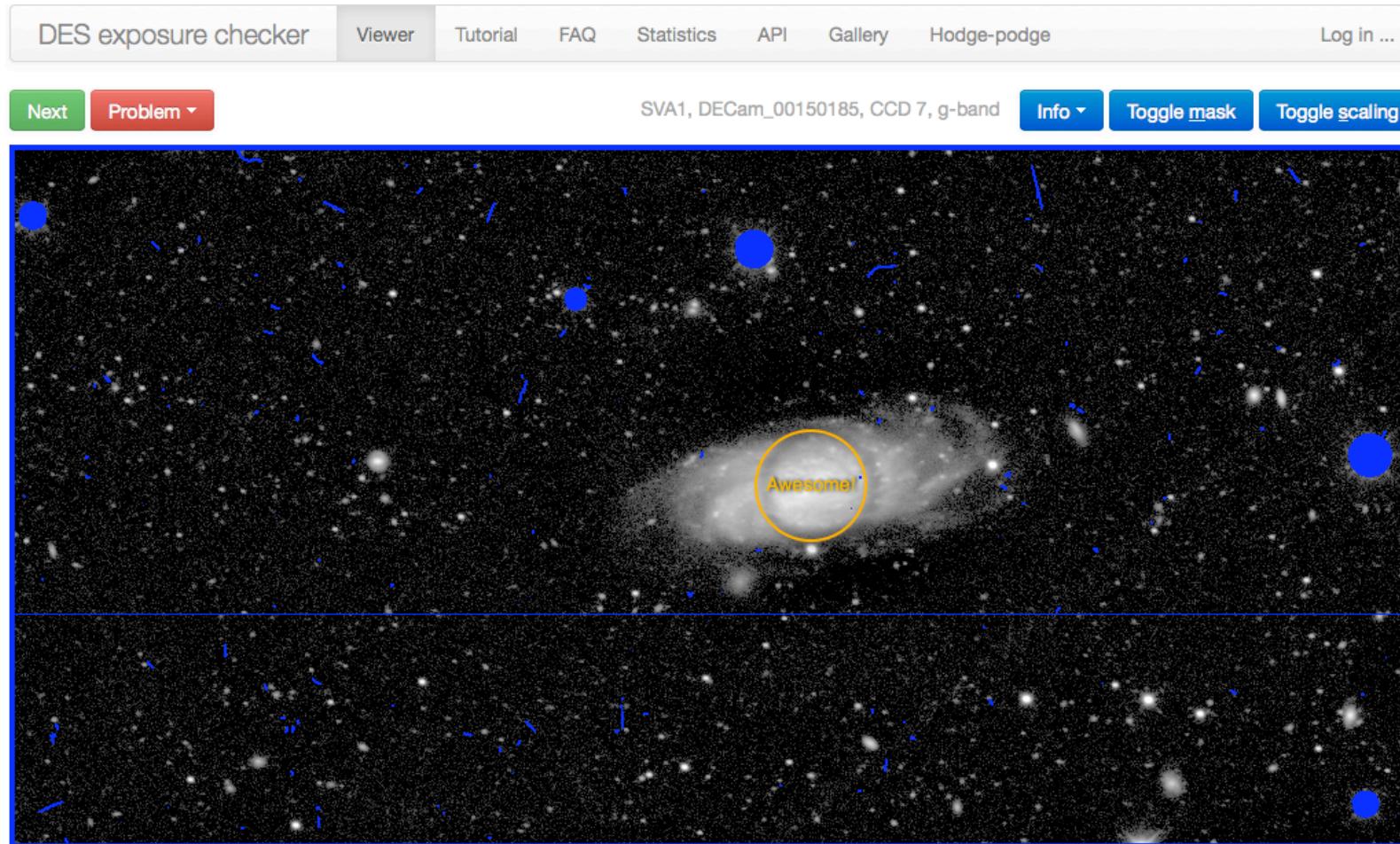
Scientists classify artifacts in pre-defined categories



P. Melchior (OSU) (peter.m.melchior@gmail.com)
E. Sheldon (BNL), A. Drlica-Wagner (Fermilab)

VISUALIZING

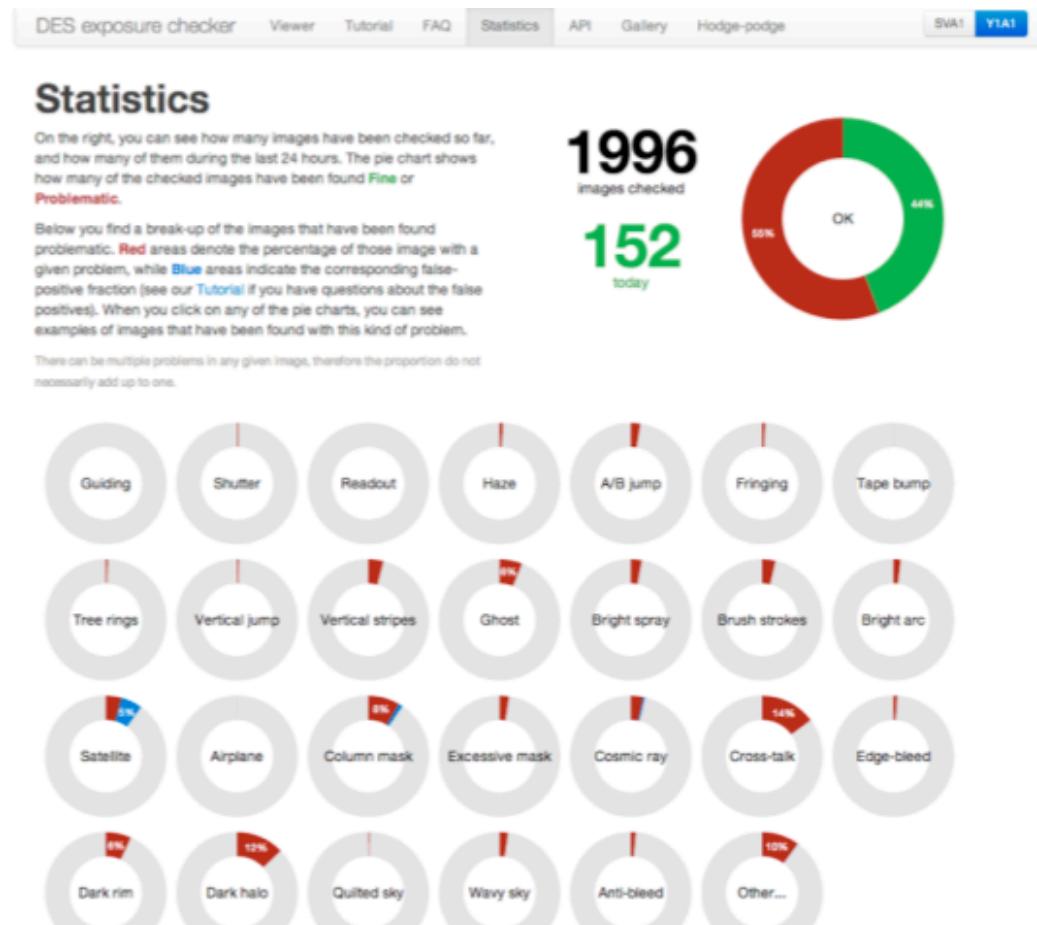
Scientists classify artifacts in pre-defined categories



P. Melchior (OSU) (peter.m.melchior@gmail.com)
E. Sheldon (BNL), A. Drlica-Wagner (Fermilab)

VISUALIZING

Statistics are compiled, data fed back to DESDM



- API for power-users to interact with artifact DB
- FAQ and wiki/forum
- Cross-validation between users

P. Melchior (OSU) (peter.m.melchior@gmail.com)
E. Sheldon (BNL), A. Drlica-Wagner (Fermilab)

VISUALIZING

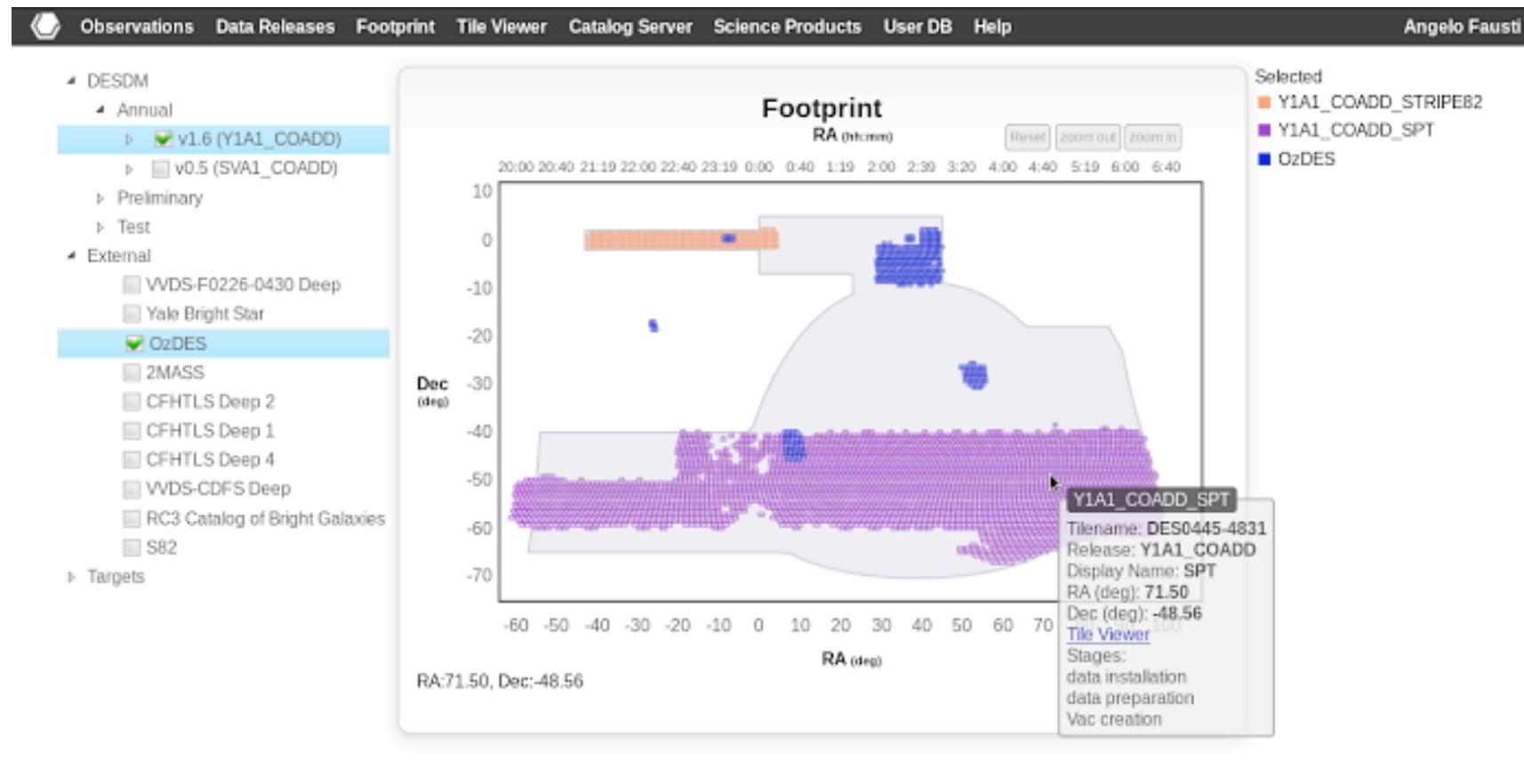
The project's Science Portal provides higher-level QA, value added catalogs

The screenshot shows the DES Science Portal Data Server page. At the top, there is a navigation bar with links for Observations, Data Releases, Footprint, Tile Viewer, Catalog Server, Science Products, User DB, and Help. To the right of the navigation bar is the name "Ignacio N. Sevilla". Below the navigation bar, the main content area has a title "DES Science Portal: Data Server". A section titled "From the Data Server instance @ FNAL you have access to following tools:" lists several tools: Observations, Data Releases, Footprint, Tile Viewer, Catalog Server, and Science Products. Below this, a note states: "The system is designed to be self-evident, use the help icon (?) available on each page." Another note says: "The Science Portal is a facility developed by [LineA](#). If you have any question please contact us through the helpdesk@linea.gov.br". On the right side of the page, there is a sidebar titled "Tweets" which displays four tweets from the account @des_portal. The tweets are as follows:

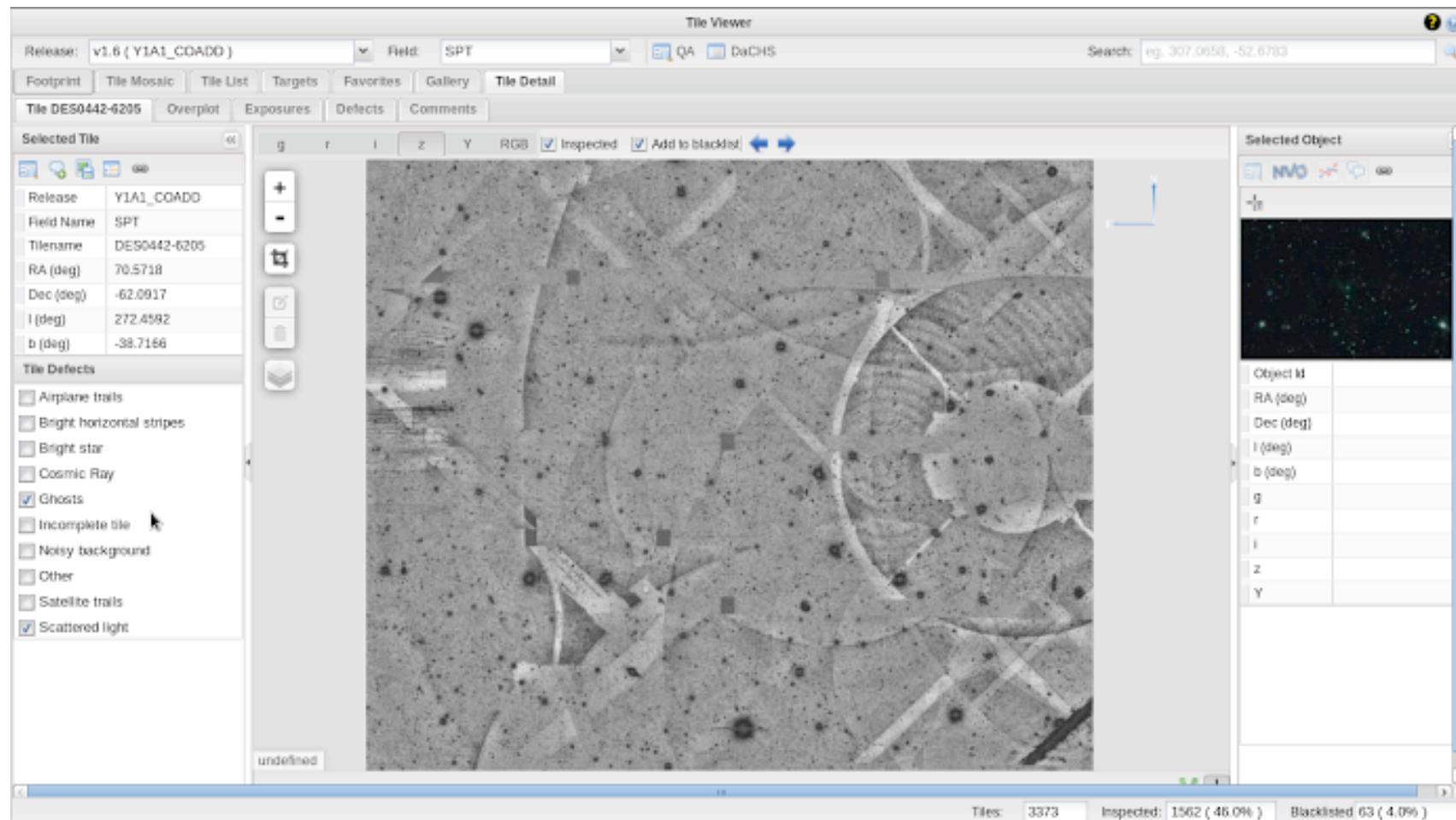
- QR results for night 2015-02-18 are available (10h ago)
- QR results for night 2015-02-17 are available (17 Feb)
- QR results for night 2015-02-16 are available (16 Feb)
- QR results for night 2015-02-15 are available (15 Feb)

At the bottom of the sidebar, there is a button labeled "Tweet to @des_portal".

Different datasets can be overlaid and linked



Coadds can be inspected and combined

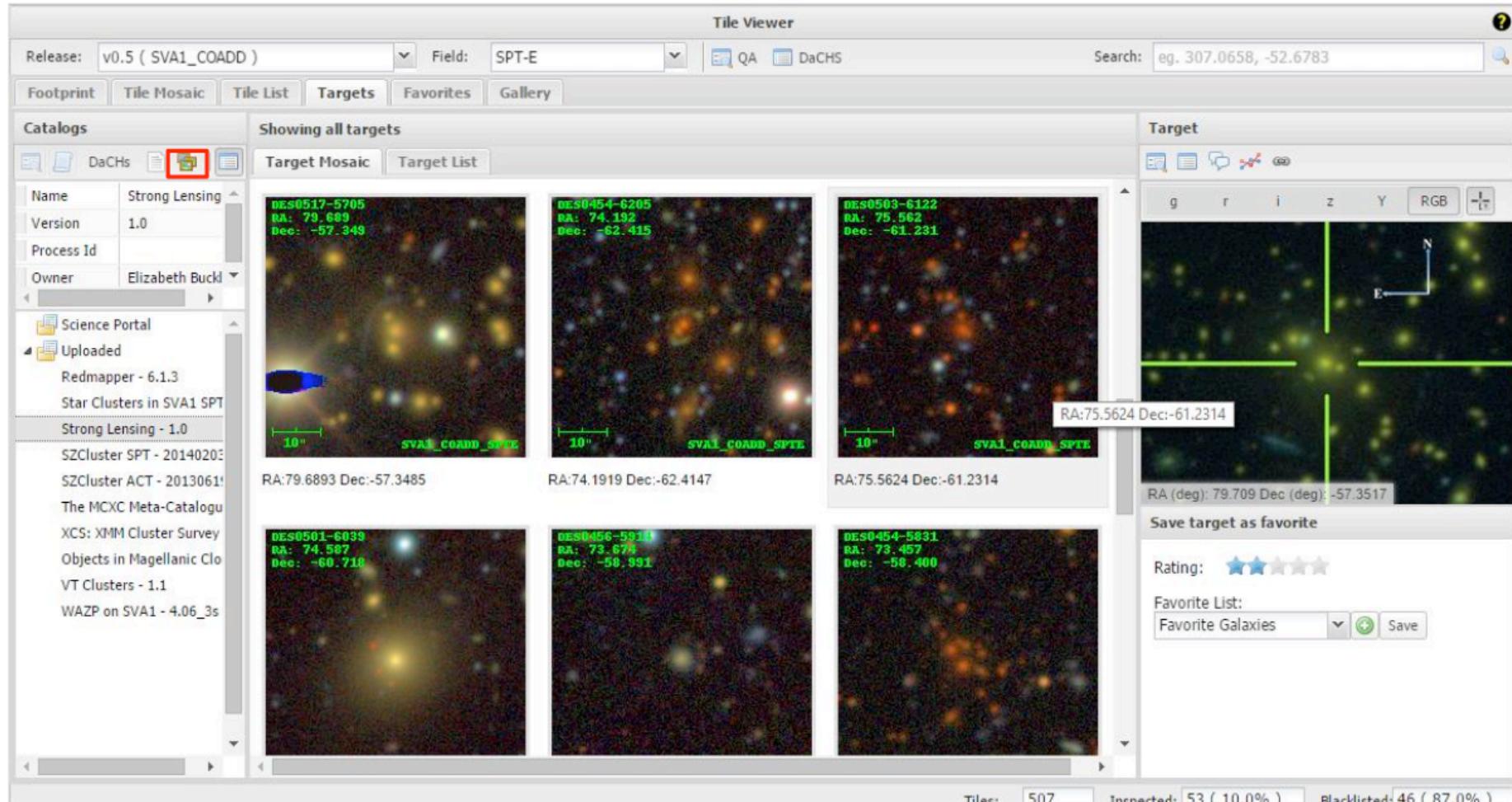


L. da Costa (ldacosta@on.br)

A.Fausti, R.Ogando++ Observatório Nacional (Rio, Brazil)

VISUALIZING

Cutouts can be created with an uploaded catalog



L. da Costa (ldacosta@on.br)

A.Fausti, R.Ogando++ Observatório Nacional (Rio, Brazil)

VISUALIZING

Scientists can query the database, store the user-defined catalogs

The screenshot shows a web-based user interface for querying a database. At the top, there is a navigation bar with links: Observations, Data Releases, Footprint, Tile Viewer, Catalog Server, Science Products, User Query, Upload, and Help. On the right side of the navigation bar, it says "Angelo Fausti". Below the navigation bar, there is a title "User Query" with a yellow warning icon. On the left, there is a sidebar menu under "Database" which includes FNAL (selected) and NCSA. The sidebar has several categories: Release, Input Tables, External Catalogs, My Library, and Sample Queries. Under "My Library", there are entries: High redshift QSOs, Blue galaxies (selected), Bright Color Outliers, g and r dropouts, Brown Dwarfs, and Sample Queries. The main area contains fields for "Name" (Blue galaxies) and "Description" (Query proposed by R. Kron to select blue galaxies). Below these fields is a code editor containing a SQL query:

```
SELECT coadd_objects.id, tilename, ra, dec FROM yla1.coadd_stripe82.coadd_objects WHERE (mag_model_g between 16.5 and 22.5) AND ((mag_psf_g - mag_model_g) > 0.05) AND ((mag_model_g - mag_model_r) BETWEEN -0.6 and 0.2) AND flags_i<=3 LIMIT 1000
```

At the bottom of the code editor, there are checkboxes for "Create Cutouts" and "# objects:1000" with a checked checkbox. Below the code editor are four buttons: Edit Query, Clear Form, Show Sample, and Submit.

Update Time: Thu Feb 12 19:57:01 2015

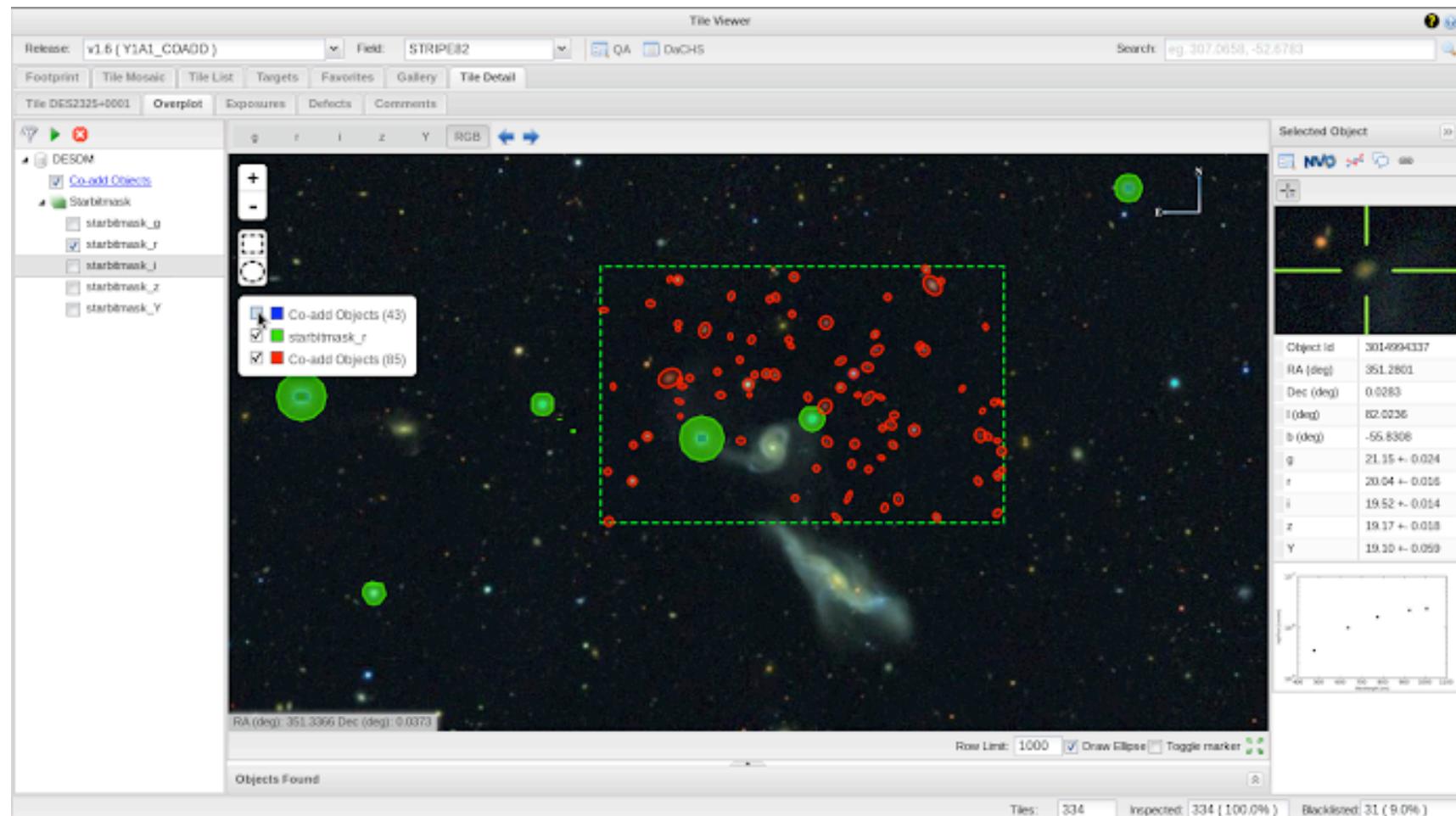
Powered by

L. da Costa (ldacosta@on.br)

A.Fausti, R.Ogando++ Observatório Nacional (Rio, Brazil)

VISUALIZING

The catalogs are overlaid on the images

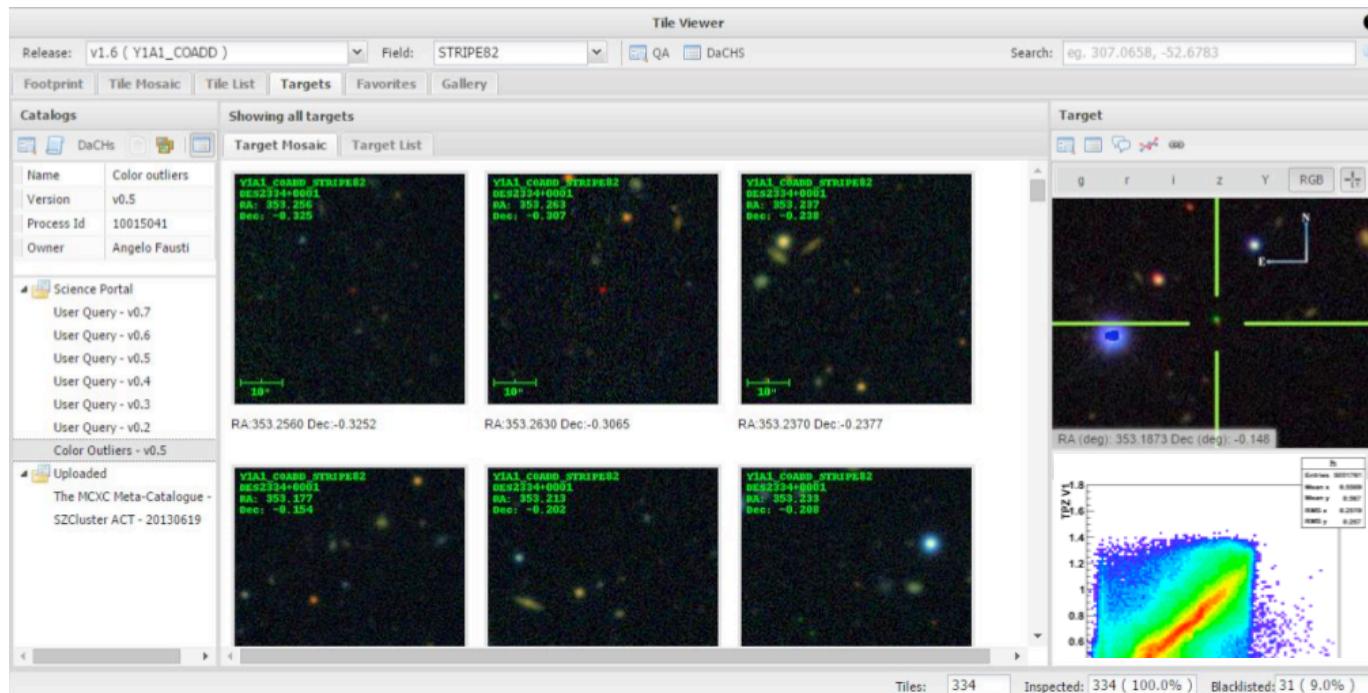


L. da Costa (ldacosta@on.br)

A.Fausti, R.Ogando++ Observatório Nacional (Rio, Brazil)

VISUALIZING

Enabling usage of distributed resources



Enable machine-independent, easy-to-use and robust analysis:

- Ease of deployment
- Making possible the use of other resources
- Facilitating scalability

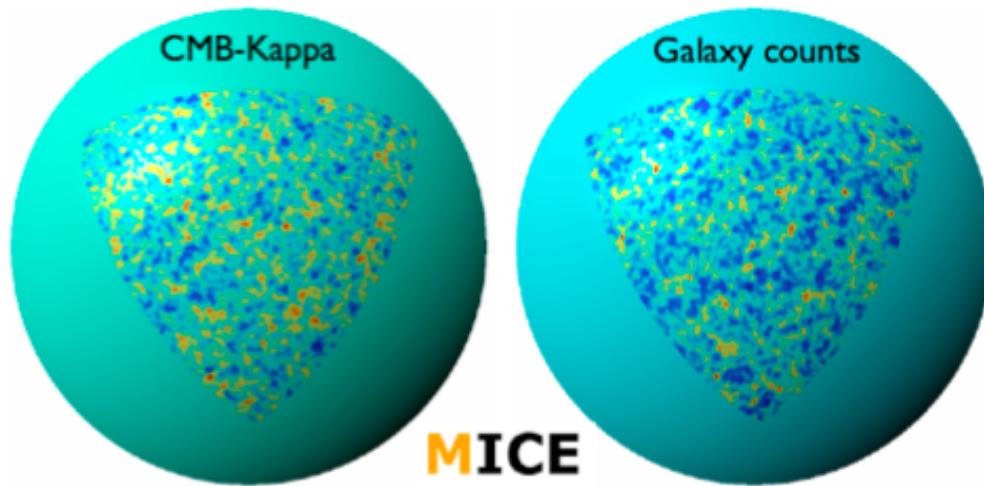
R. Brunner (professor.brunner@gmail.com)

M. Turk, D.Petravick (UIUC/NCSA)



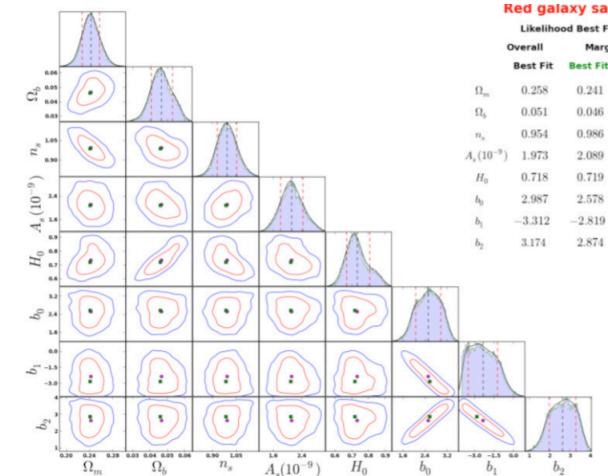
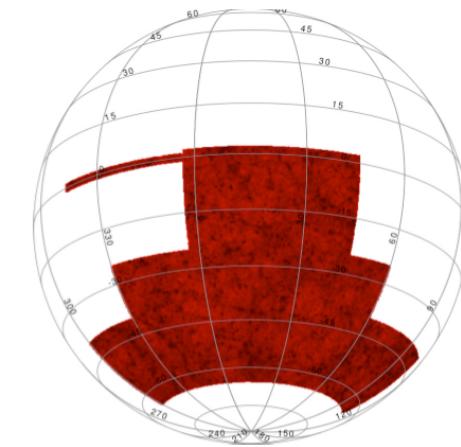
A key new ingredient of DES is its heavy use of simulations

- N-body + galaxy sims: Testing analysis codes → building pipelines, quantifying statistical/systematic errors.
- Factors of \sim few for the catalogs and CPU hours.
- Resource availability impediments.



<http://cosmohub.pic.es>

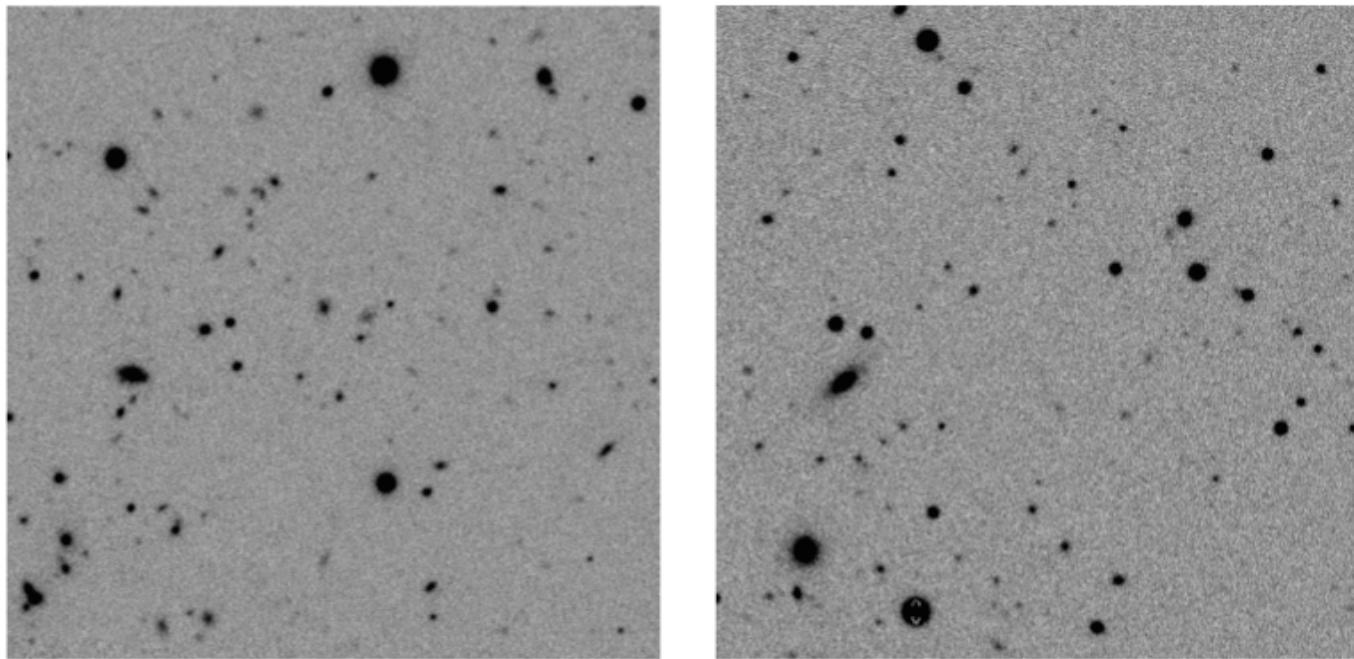
SLAC; MICE Collaboration; F.Sobreira et al. (Observatório Nacional)



ANALYZING

A key new ingredient of DES is its heavy use of simulations

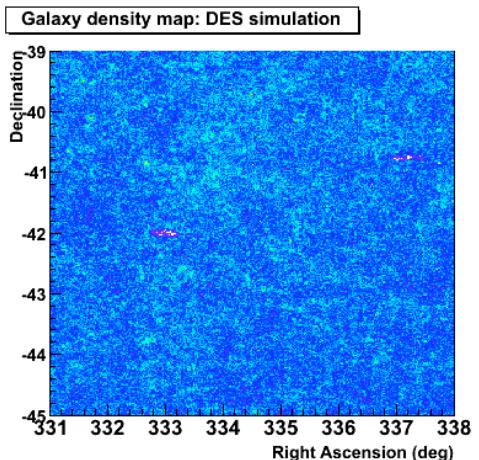
- Full image simulations: Testing DESDM pipeline, lower-level (pixel) systematics.



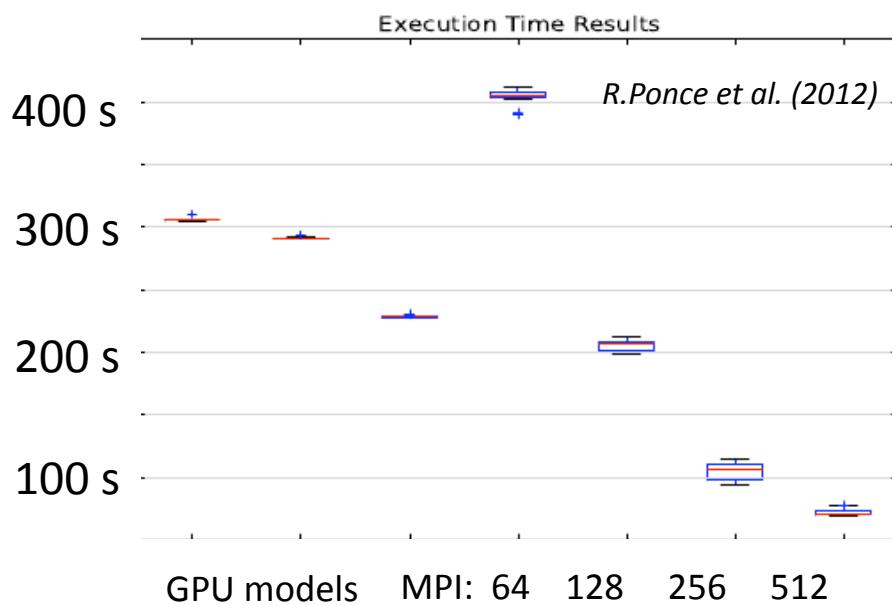
C.Chang et al. (2014) ; UFIG Bergé et al. (2013);
previous SW by DESDM+Fermilab, S.Desai et al. (2012)

ANALYZING

Estimating cosmological parameters from galaxy positions/shapes require costly N-point correlations



$$\omega(\theta) = 1 + \left(\frac{N_{random}}{N_{real}} \right)^2 \cdot \frac{DD(\theta)}{RR(\theta)} - 2 \cdot \left(\frac{N_{random}}{N_{real}} \right) \cdot \frac{DR(\theta)}{RR(\theta)}$$



Full two-point computation can be achieved in reasonable times using GPUs.

R.Ponce, I.S. et al. (2012) arXiv:1204.6630
M.Cárdenas-Montes, I.S. et al. (2014)
(CIEMAT)

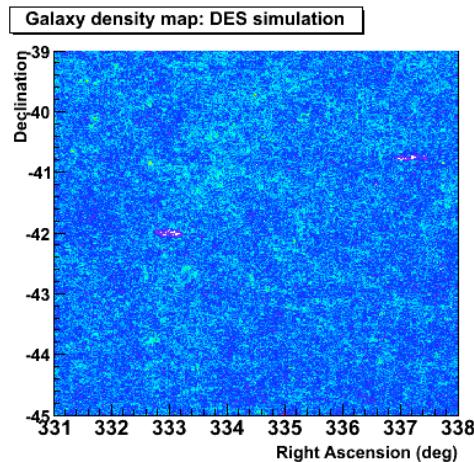
nsevilla@gmail.com

github.com/nsevilla/pointpoint

github.com/nsevilla/shearshear

ANALYZING

Estimating cosmological parameters from galaxy positions/shapes require costly N-point correlations

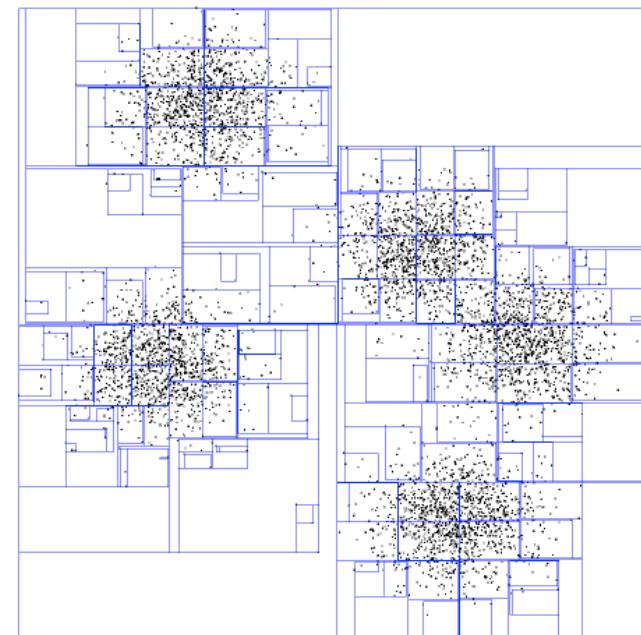


$$\omega(\theta) = 1 + \left(\frac{N_{random}}{N_{real}} \right)^2 \cdot \frac{DD(\theta)}{RR(\theta)} - 2 \cdot \left(\frac{N_{random}}{N_{real}} \right) \cdot \frac{DR(\theta)}{RR(\theta)}$$

Very commonly-used alternative is adaptive mesh refinement through kd-trees (Moore 2001).

Speeds can be enhanced by efficient management of cache memory and tree creation through arrays.

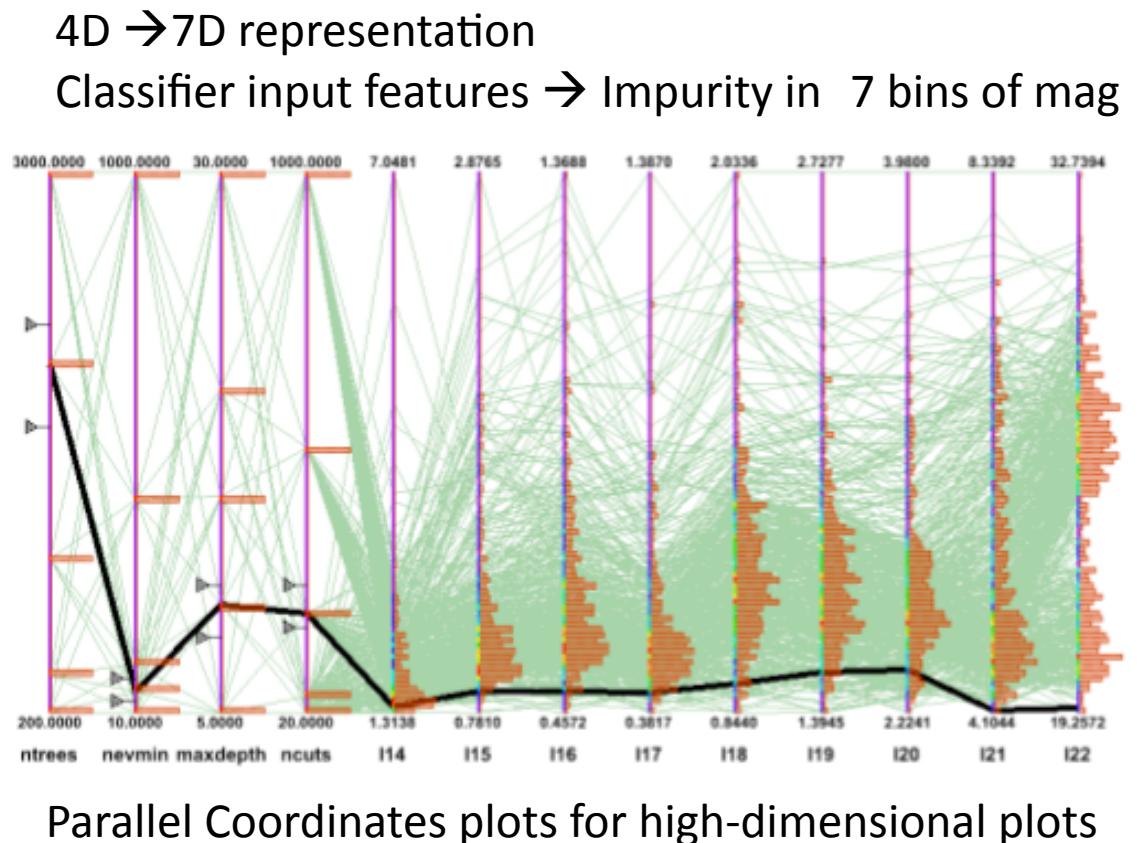
V. Reverdy, UIUC (vince.rev@gmail.com)



ANALYZING

Machine Learning for cosmology

- Photometric redshift
- Object classification
- Artifact identification
- Transient search

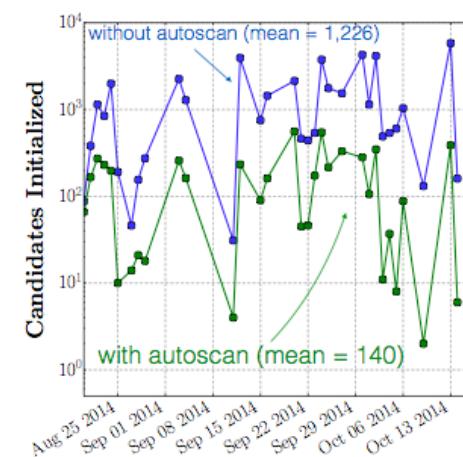
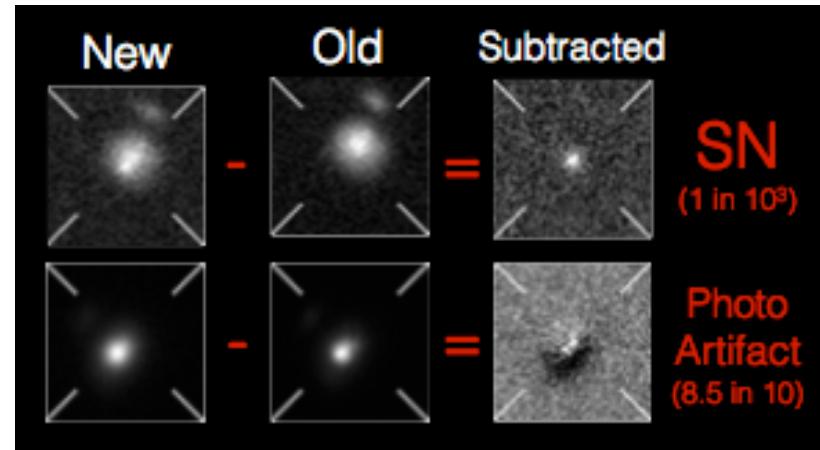


I.S., P.Etayo-Sotos (2015)

MACHINE LEARNING

Machine Learning for cosmology

- Photometric redshifts
- Object classification
- Artifact identification
- Transient search



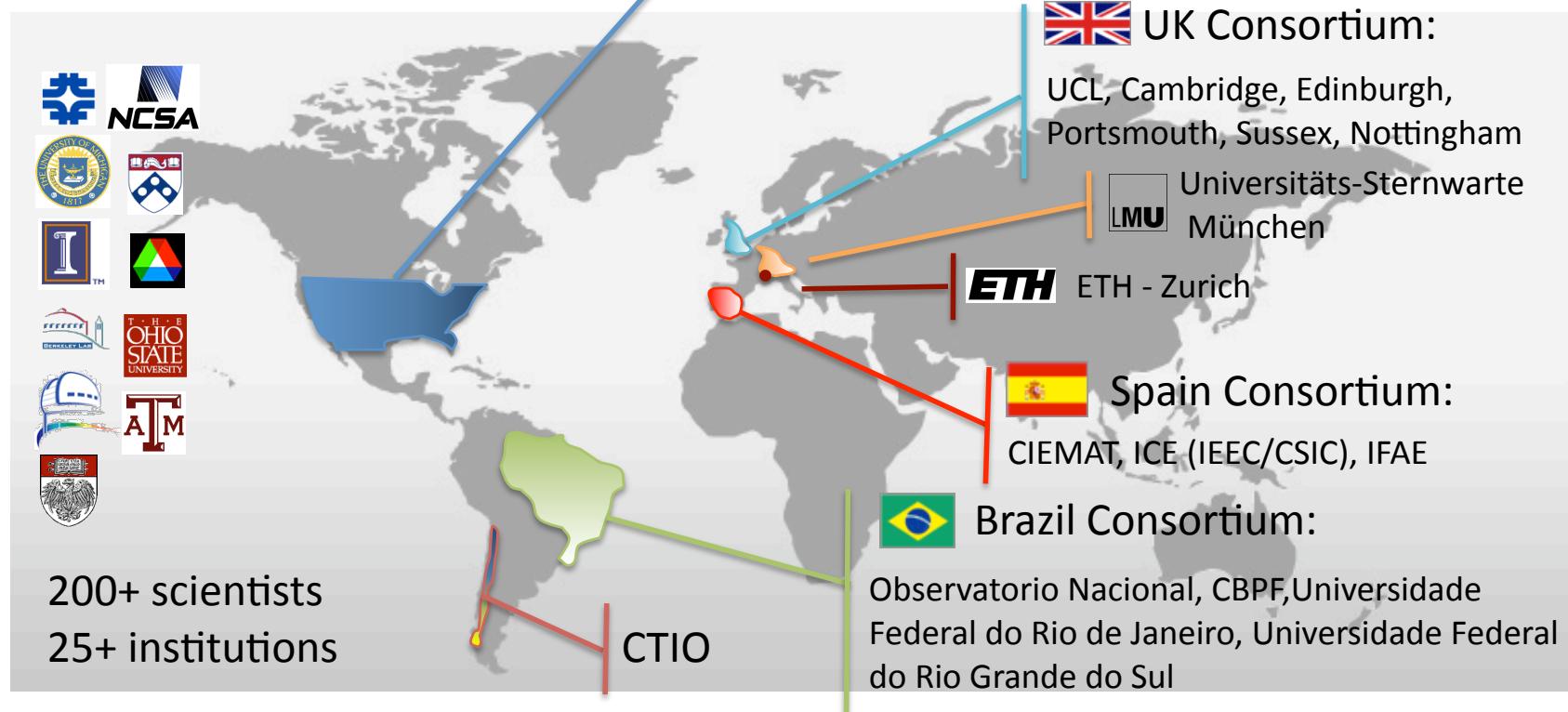
D. Goldstein et al. (UCBerkeley) in prep. (dgold@berkeley.edu)

MACHINE LEARNING

The inevitable collaboration slide

- Big Collaborations → Big Data problem too
- Coping with multitude of use cases, data access modes (occasional vs power users – ‘proxies’)
- Eliminate overheads for scientists to exploit the data fully

Fermilab, UIUC/NCSA, University of Chicago, LBNL, NOAO, University of Michigan, University of Pennsylvania, Argonne National Laboratory, Ohio State University, Santa-Cruz/SLAC/Stanford Consortium, Texas A&M University



DES is about the most robust measurement of dark energy and an extraordinary legacy

But also about preparing for next generation surveys.

- DESDM@NCSA performing exceptionally with heterogeneous resources.
- Lots of smart people doing interesting developments.
- Simulations are a big player.
- Data storage, resource availability (queue throughput) important technical issues.
- People/site management are part of Big Data.