



# **Streamlining the Sloan Digital Sky Survey Public Data Releases**

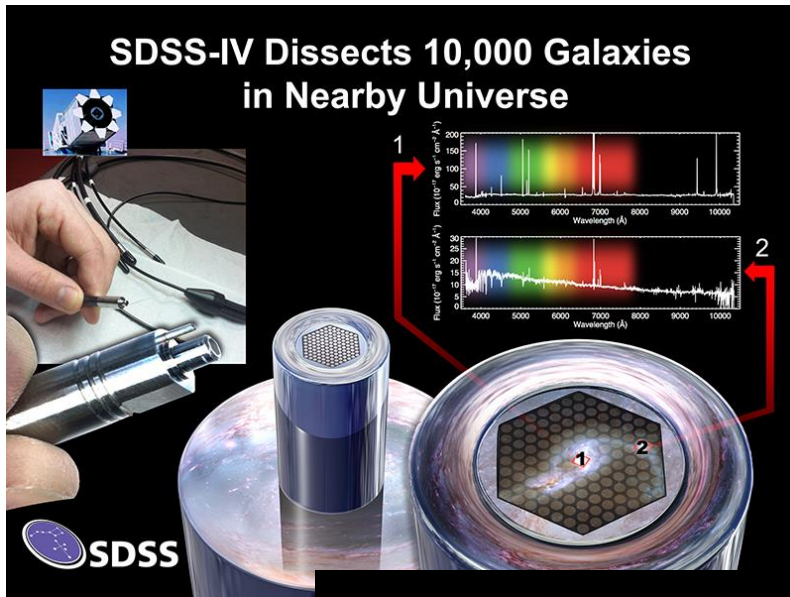
## **Changes Made, Lessons Learned**

Anne-Marie Weijmans  
University of St Andrews

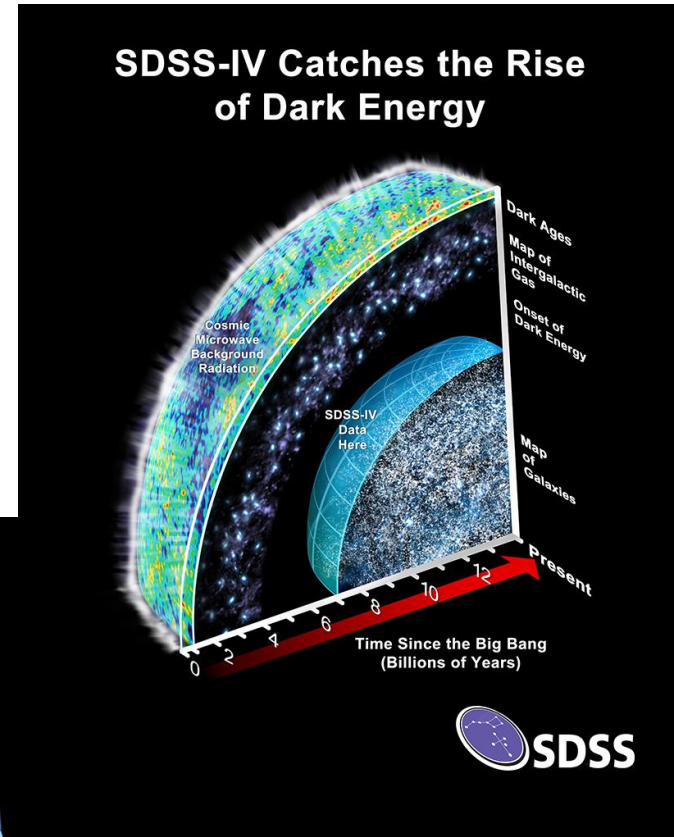
# The Sloan Digital Sky Survey

- Started in 1998 with imaging survey
  - dedicated 2.5m Sloan telescope, APO
- Continued with SDSS-II, III and now SDSS-IV
  - add optical, infrared and integral-field spectra
  - also observations from LCO
- More than 50 member institutes and groups
  - spanning 4 continents and almost all time zones
  - almost 1000 active scientists

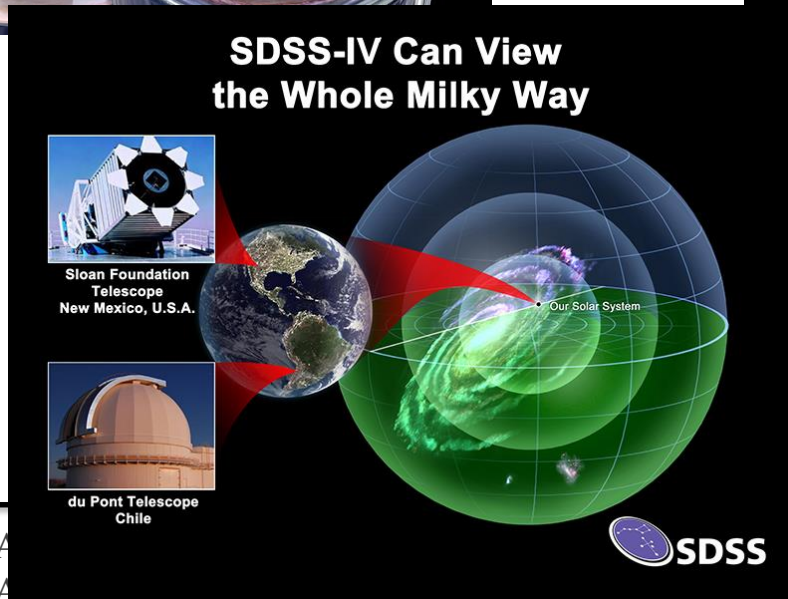
# Current surveys in SDSS-IV



MaNGA



eBOSS



APOGEE-2

# Data in SDSS-IV

Joel Brownstein  
Univ. of Utah



Index of /sas/dr15/

File Name ↓	File Size ↓	Date
Parent directory/	-	-
apo/	-	2018-
apogee/	-	2017-
casload/	-	2018-Jun-28 19:22
eboss/	-	2017-Apr-02 19:29
env/	-	2019-Feb-13 21:40
manga/	-	2018-Jul-17 18:40
marvels/	-	2017-Jun-14 00:39
sdss/	-	2017-Jun-06 15:01

SDSS-IV Science Archive Server (SAS)

<https://dr15.sdss.org/sas/dr15/>



Ani Thakar  
JHU

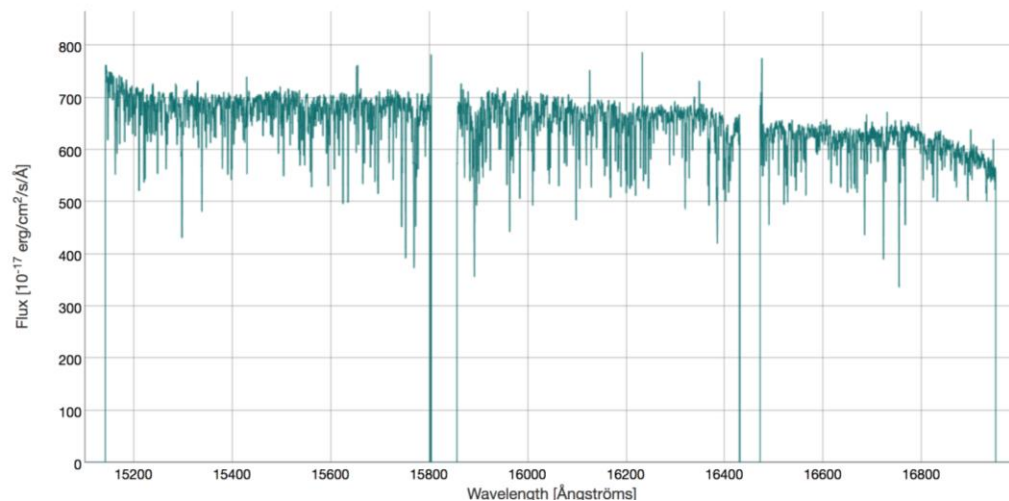
<http://skyserver.sdss.org>



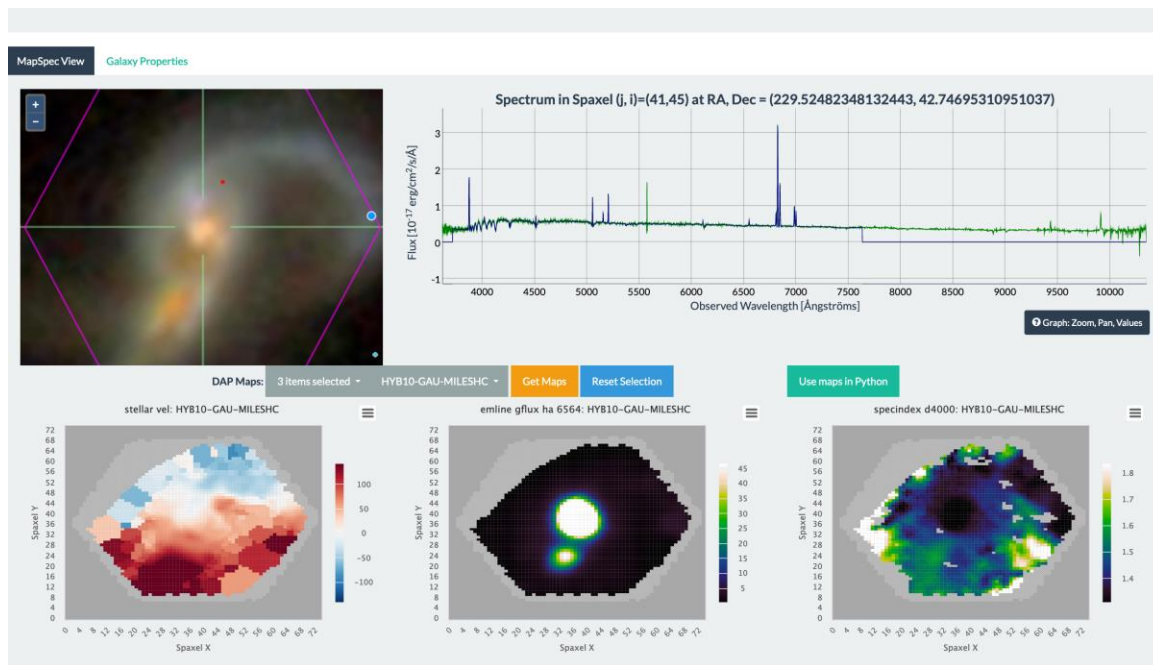


# Data Interaction

<https://dr15.sdss.org/>



Marvin



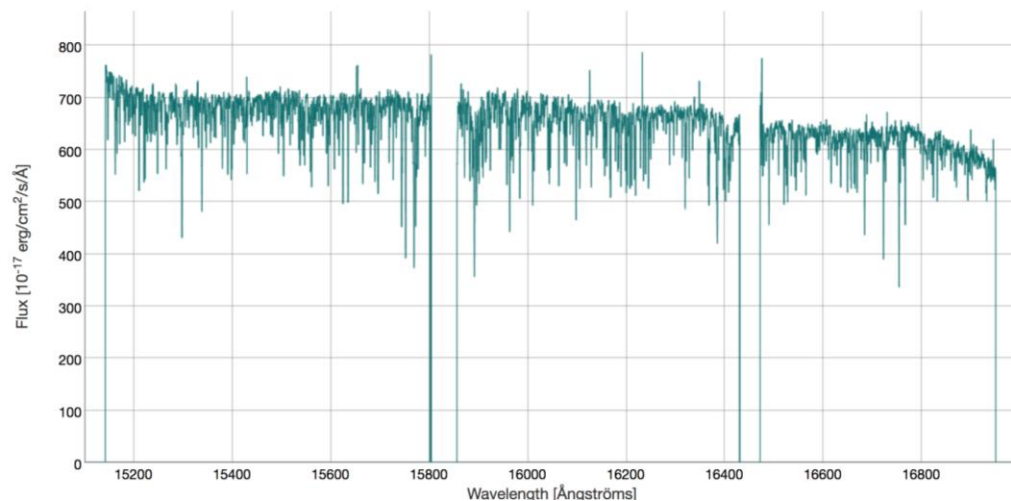
Science Archive Webapp



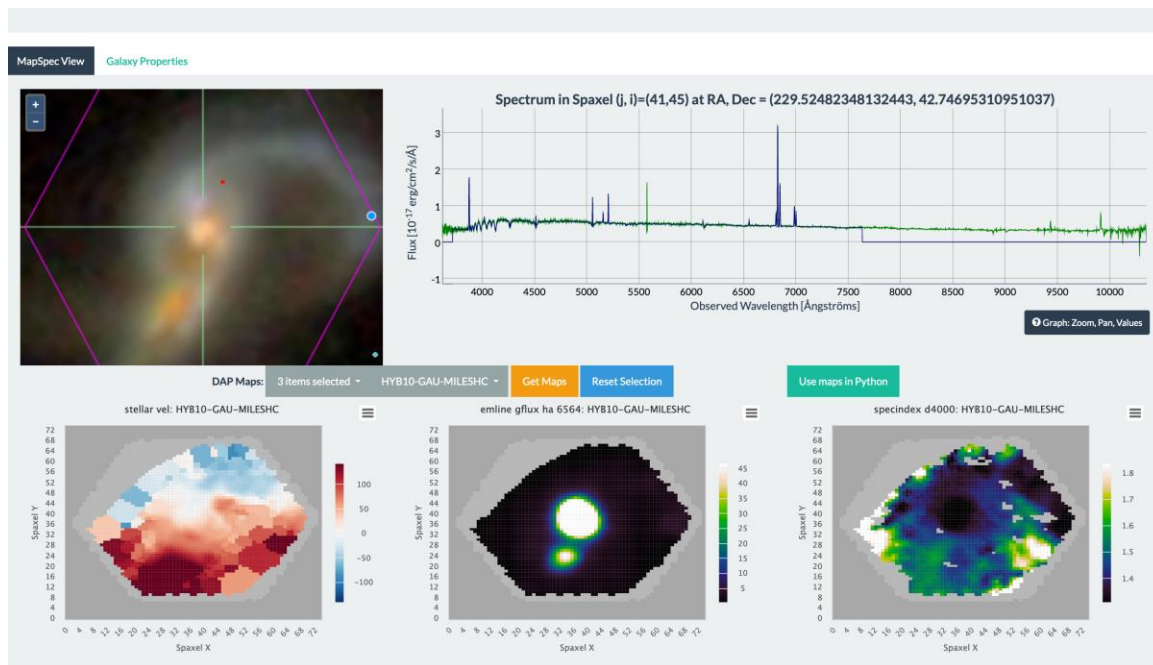
@SDSS\_Marvin

# Data Interaction

<https://dr15.sdss.org/>



Marvin



Science Archive Webapp



Brian Cherinka, STScI

# SDSS Public Data Releases: Why?

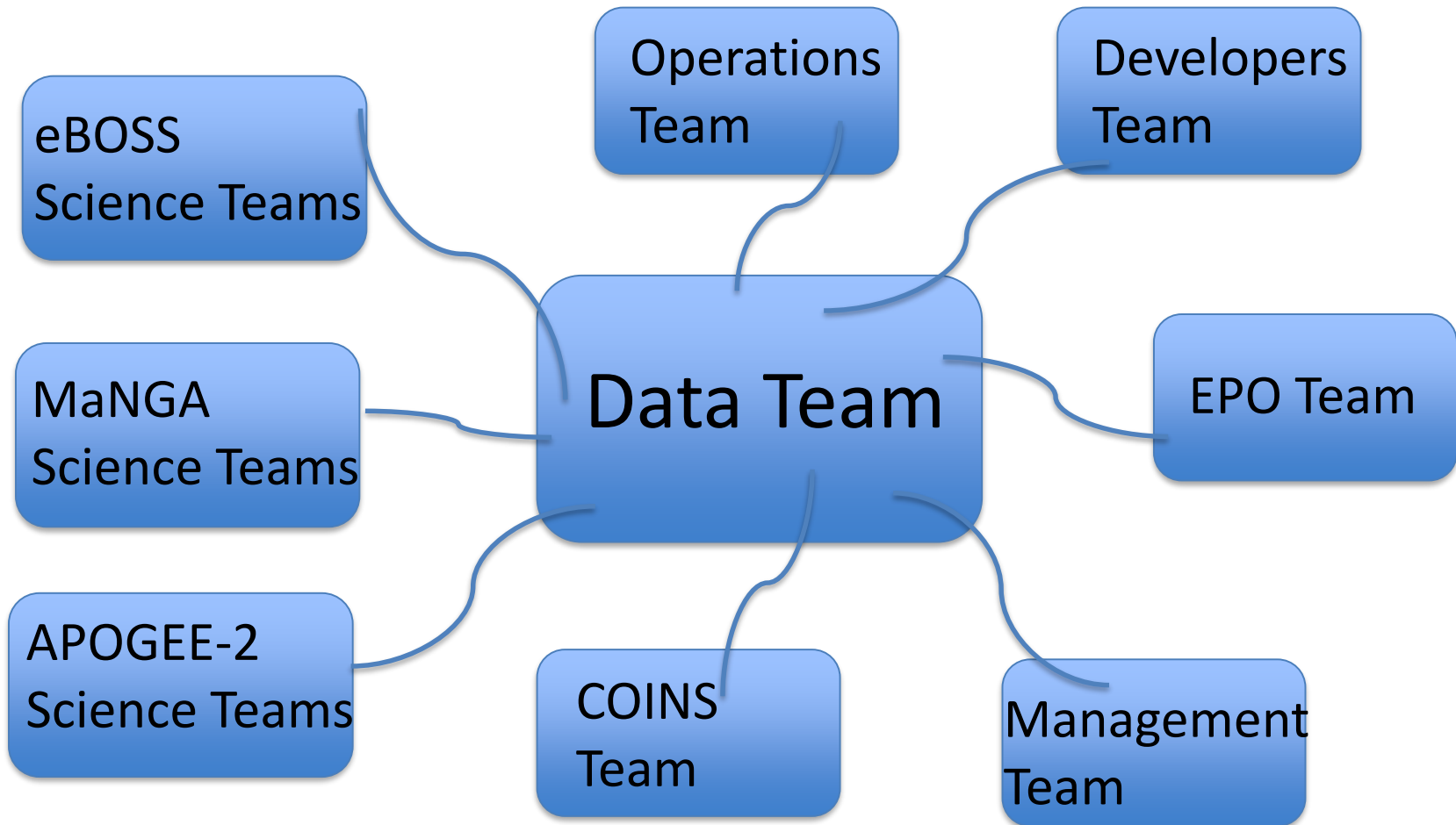
- To increase impact of our science
- To enable reproducibility of our results
- To increase the longevity of the data
- To enable new projects by non-SDSS scientists
  - and provide input for EPO projects
- To keep ourselves honest
- Because that's how we roll

# SDSS Public Data Releases: Why?

- To increase impact of our science
- To enable reproducibility of our results
- To increase the longevity of the data
- To enable new projects by non-SDSS scientists
  - and provide input for EPO projects
- To keep ourselves honest
- Because that's how we roll
- **Because it strengthens the team**



# Team Structure



# Value Added Catalogs (VACs)

- VACs are data products not provided by SDSS data reduction pipelines
  - e.g.: APOGEE distances, eBOSS large scale structure
- Included into public data releases
- Needs to conform to data standards
  - requires data model, documentation
  - accompanied by refereed paper

# VAC database

- New in SDSS-IV: VAC announcement system
  - team members fill in form to announce VAC
- Documentation available on how to prepare VAC for data release
  - file structure, data model, documentation
- VACs get published twice a year (on average)
  - with data release, or half year later in mini release
  - needs to be build on public data

# Effect of new VAC system

- Larger participation in VAC creation
  - SDSS-I, II, III → 11 VACs
  - SDSS-IV → additional 32 VACs (+16 more for DR16)
- More diverse VAC owners
  - team members with various levels of expertise are creating VACs
  - more connections science teams and data team
- Larger workload for data team
  - worth it, but need to think about scalability

# Documentation

- Data needs to be accessible, not just available
  - documentation is essential
  - each SDSS data product has a data model
- End-users need different types of documentation
  - astronomers, students, teachers, general public
  - provide tailored documentation and tutorials
- Writing documentation is a community effort
  - requires expertise of data, science and EPO teams



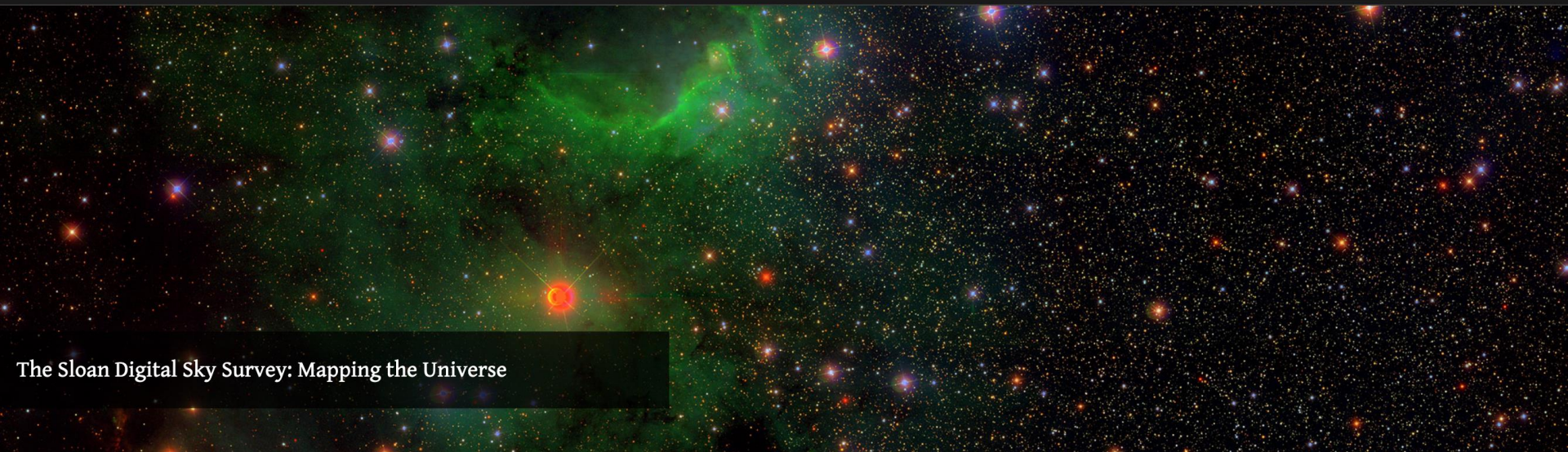


This is Data Release 15.

[Data](#) [Surveys](#) [Instruments](#) [Collaboration](#) [Results](#) [Education](#) [The Future](#) [Contact](#)

Search [www.sdss.org](#)

Search



The Sloan Digital Sky Survey: Mapping the Universe



Jordan Raddick, JHU

The Sloan Digital Sky Survey has created the most detailed three-dimensional maps of the Universe ever made, with deep multi-color images of one third of the sky, and spectra for more than three million astronomical objects. Learn and explore all phases and surveys—past, present, and future—of the SDSS.



Bonnie Souter, JHU

# DocuFeest!



Image Credit: Jennifer Johnson



# DocuFeest!

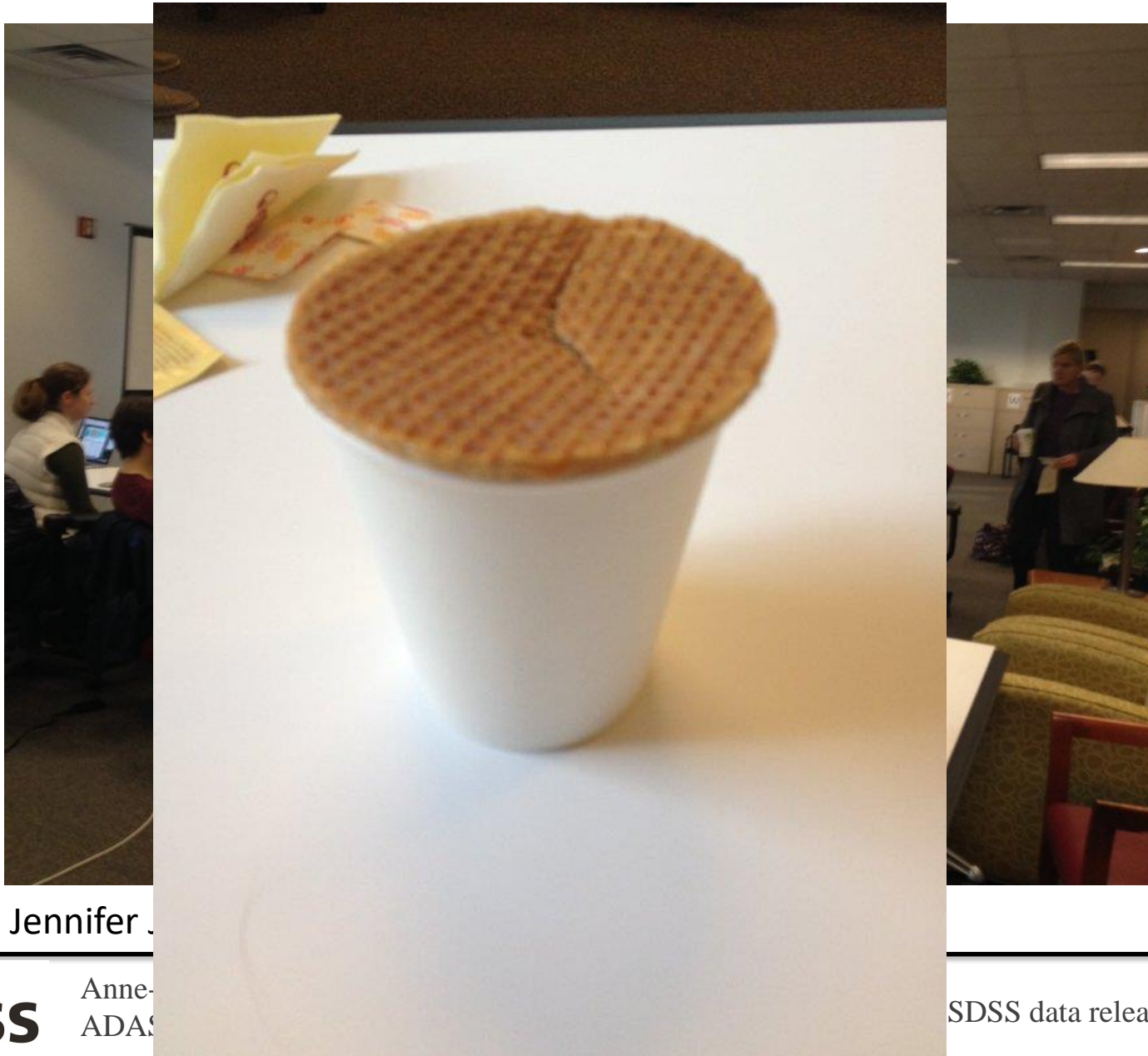


Image Credit: Jennifer J.

# Documentation lessons

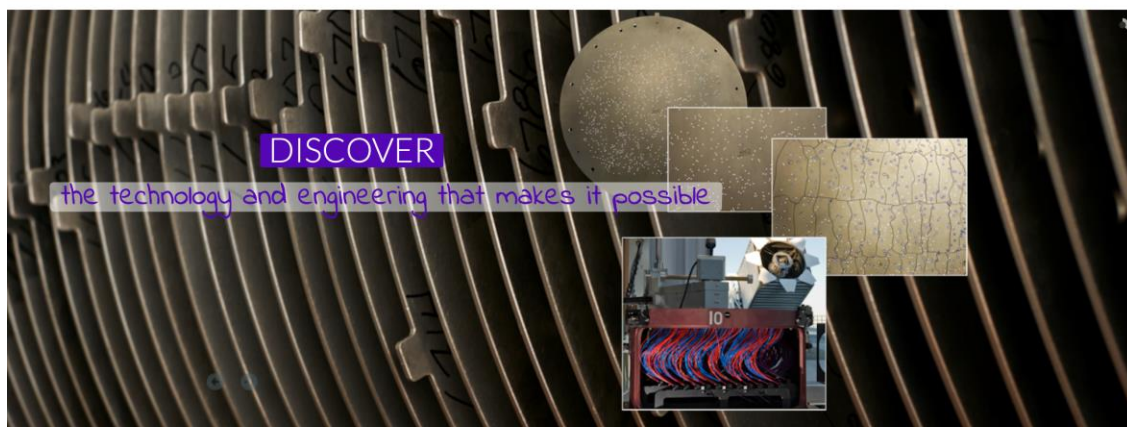
- Documentation is vital for a useful data release
  - but also: to help new team members!!!
- Documentation needs continuous effort
- Documentation is a team effort
  - not just the developer that writes documentation
  - ideally with dedicated tutorial writers
- Documentation is aimed at undergraduate level
  - what about the general public?

# Education and Public Outreach



<http://voyages.sdss.org/>

Jordan Raddick, JHU



<http://sciserver.org>, see booth!



Rita Tojeiro  
St Andrews



Britt Lundgren  
UNC Asheville

Anne-Marie Weijmans  
ADASS XXIX– Groningen – 9 Oct 2019



Sloan Digital Sky Surveys  
@sdssurveys

Following

```
import SciServer.CasJobs as CasJobs
import matplotlib.pyplot as plt
q = 'select modelmag_u, modelmag_r,
modelmag_u-modelmag_r as ur, z from
specphoto where z between 0.04 and
0.05 and modelmag_r > 0'
r= CasJobs.executeQuery(q,'dr14')
plt.scatter(r['modelmag_r'], r['ur'])
```

[Translate Tweet](#)

1:38 AM - 16 May 2018

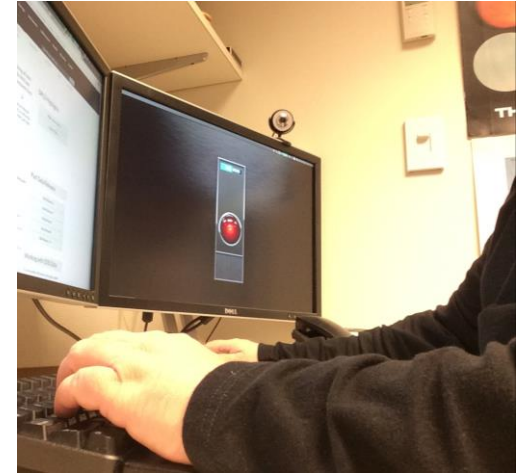




# Help? Helpdesk!

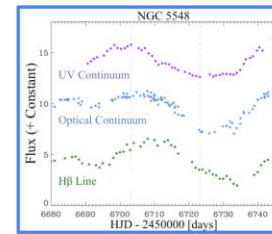
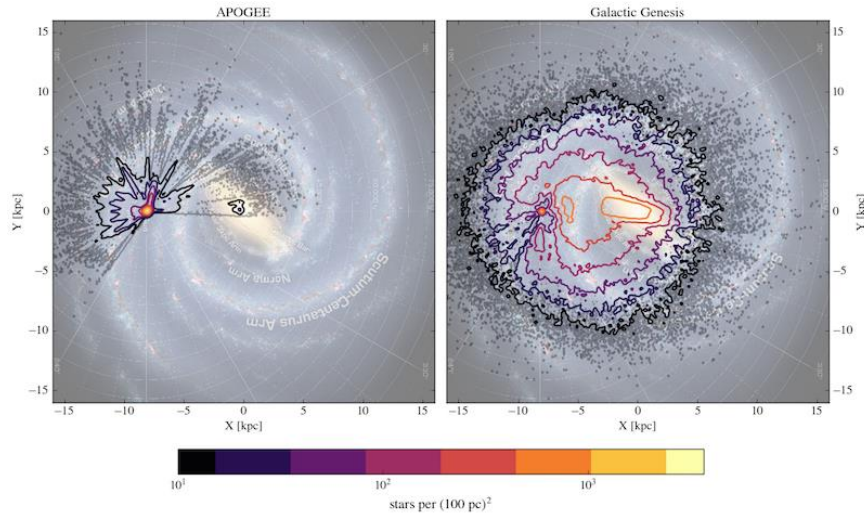
Ben Weaver, NOAO

- Need help? Send e-mail to helpdesk, advertised on webpage
- About 2-3 e-mails per week
  - more traffic just after data release
- E-mails get triaged and sent to experts if needed
- Concerns:
  - scalability? loss of expertise?
- Thinking about platform based helpdesk
  - input from science teams and community

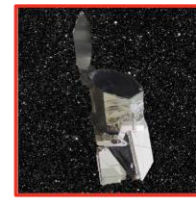


# The Future: SDSS-V

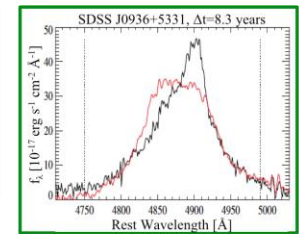
## Milky Way Mapper



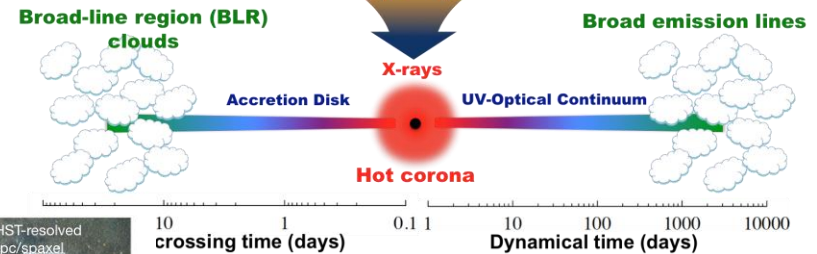
**Reverberation Mapping**  
Measuring BLR sizes and BH masses



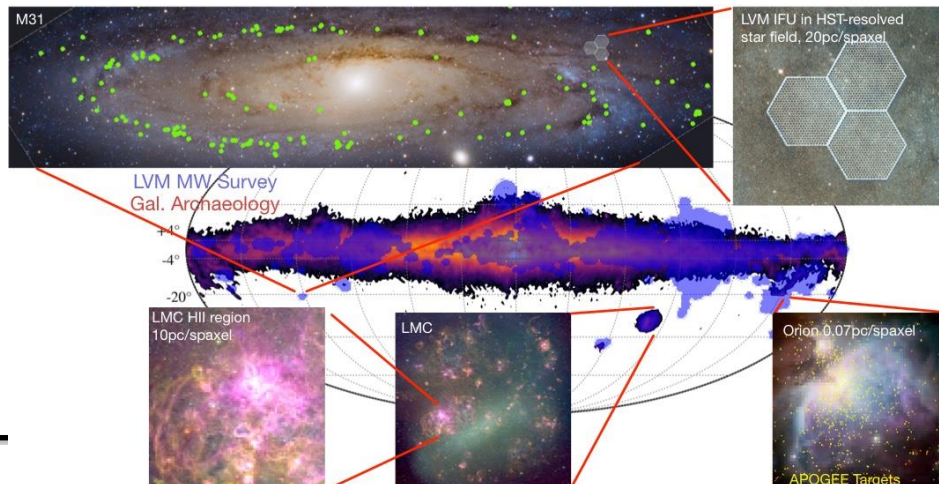
**eROSITA Follow-up**  
Probing hot X-ray coronae



**Multi-epoch Spectroscopy**  
Probing dynamical changes in the BLR



## Black Hole Mapper



## Local Volume Mapper

# Conclusions

- Public data releases are team work
  - and can be used to make your team stronger and more connected
- Public data releases serve both your team and the community
  - documentation needed also for new team members and students!
- Public data releases are hard work, but worth it!
  - direct impact + long-time archiving

