

# Streamlining MaNGA data with Marvin



Brian Cherinka (JHU)  
Brett Andrews (Pitt)  
José Sánchez-Gallego (UW)

# Outline

MaNGA overview

Astronomy Bottlenecks

Introducing Marvin

Marvin Live Demo

# MaNGA overview

Overview on MaNGA

# Science motivation

Metallicity

Stellar / gas kinematics

AGN

Star  
Formation  
History

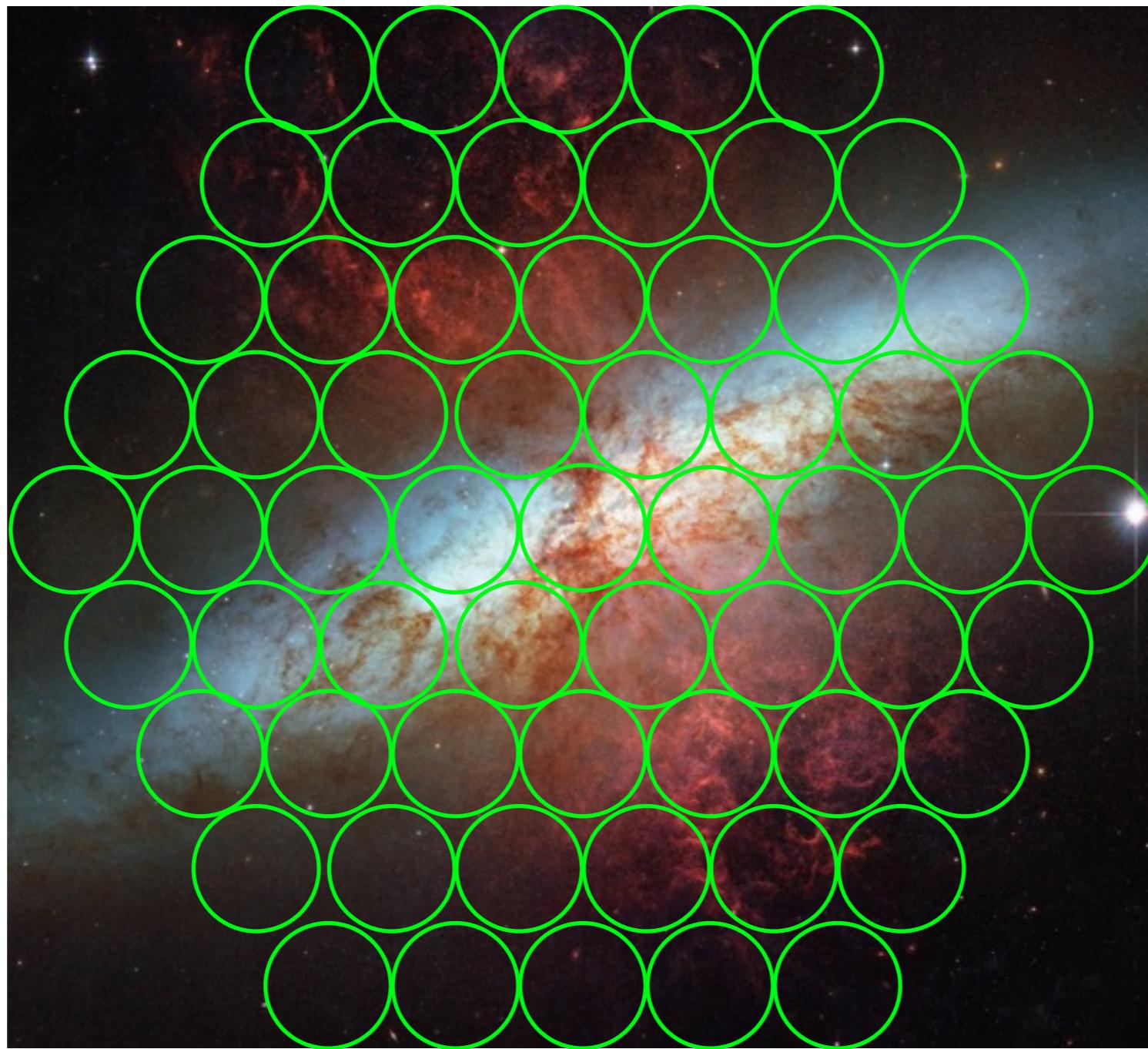


SFR

Chemical  
abundances

# Science motivation

Disk growth



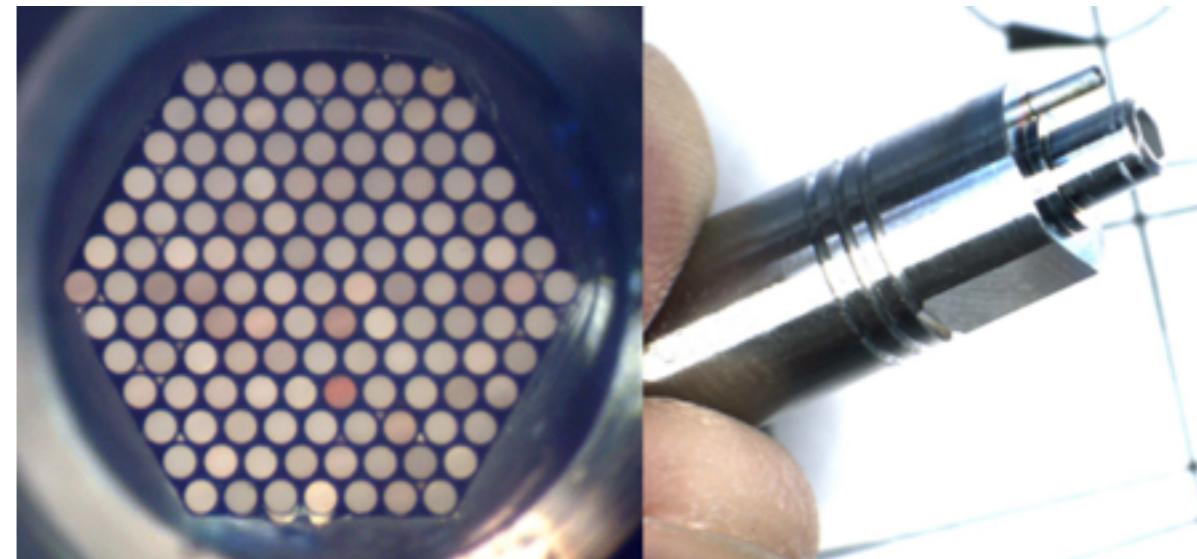
SF regulation

Spheroid growth

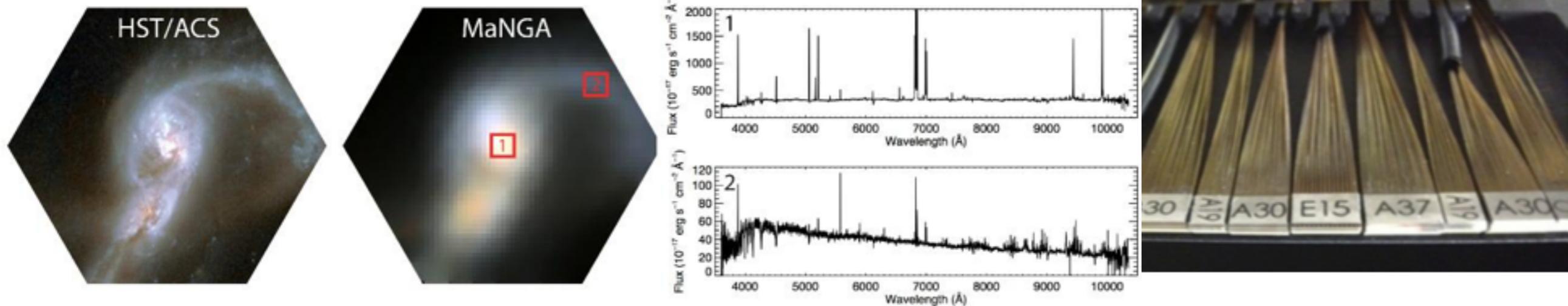
Mass and  
ang. mom.  
distribution

# MaNGA overview

- MaNGA: Mapping Nearby Galaxies at APO
- PI: Kevin Bundy. Over 160 members in 50+ institutions.
- Part of SDSS-IV (2014-2020)
- IFU observations of 10,000 galaxies (4000 already observed!)
- Stellar library during bright time
- $0.01 < z < 0.15$
- $\lambda \sim 3600\text{--}10300 \text{\AA}$
- $R \sim 1400\text{--}2600$  (115-215 km/s)
- Spatial resolution 1.3-5.1 kpc

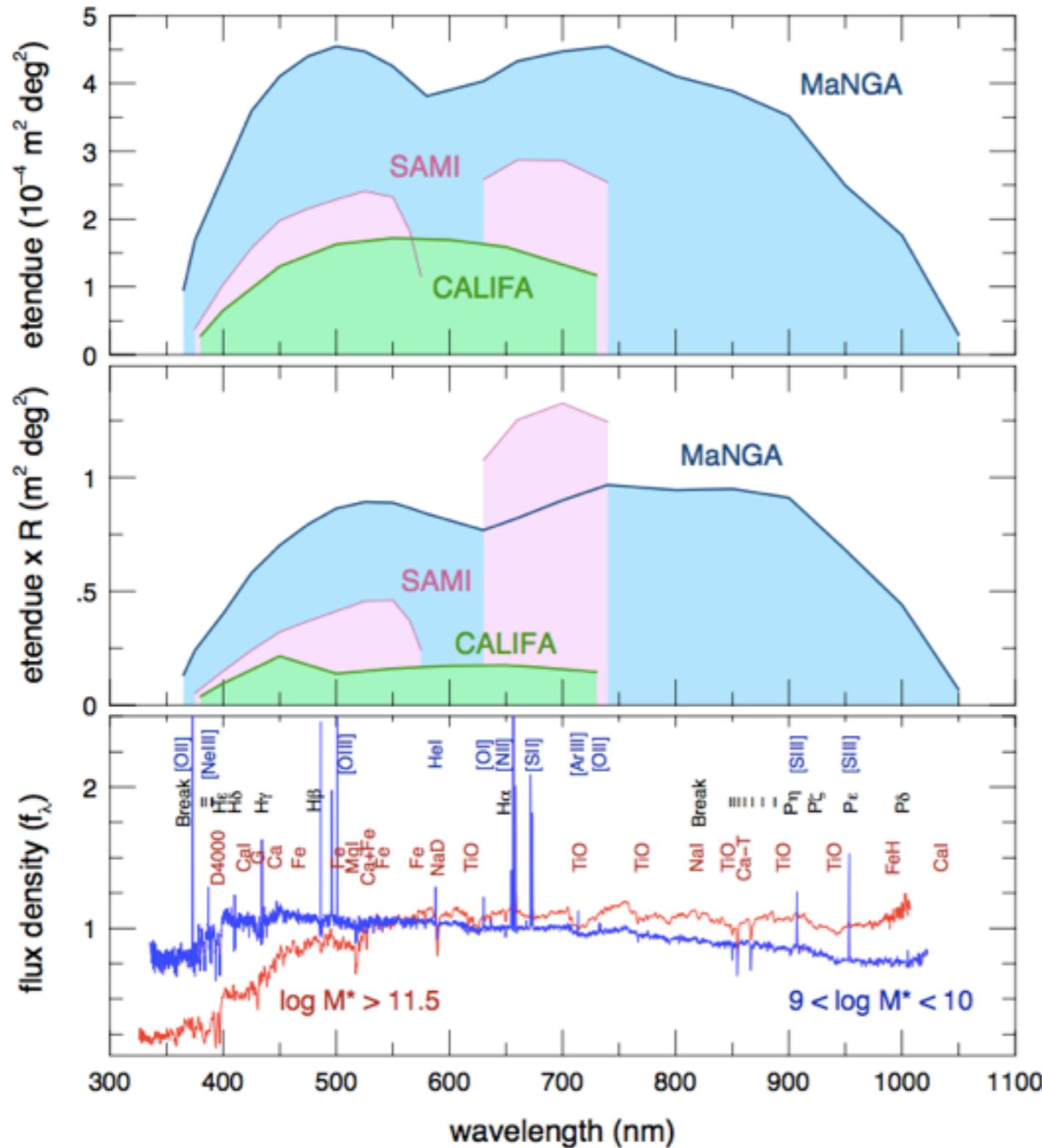


Mrk 848: SDSS-IV/MaNGA First-Article Data Cube



# MaNGA overview

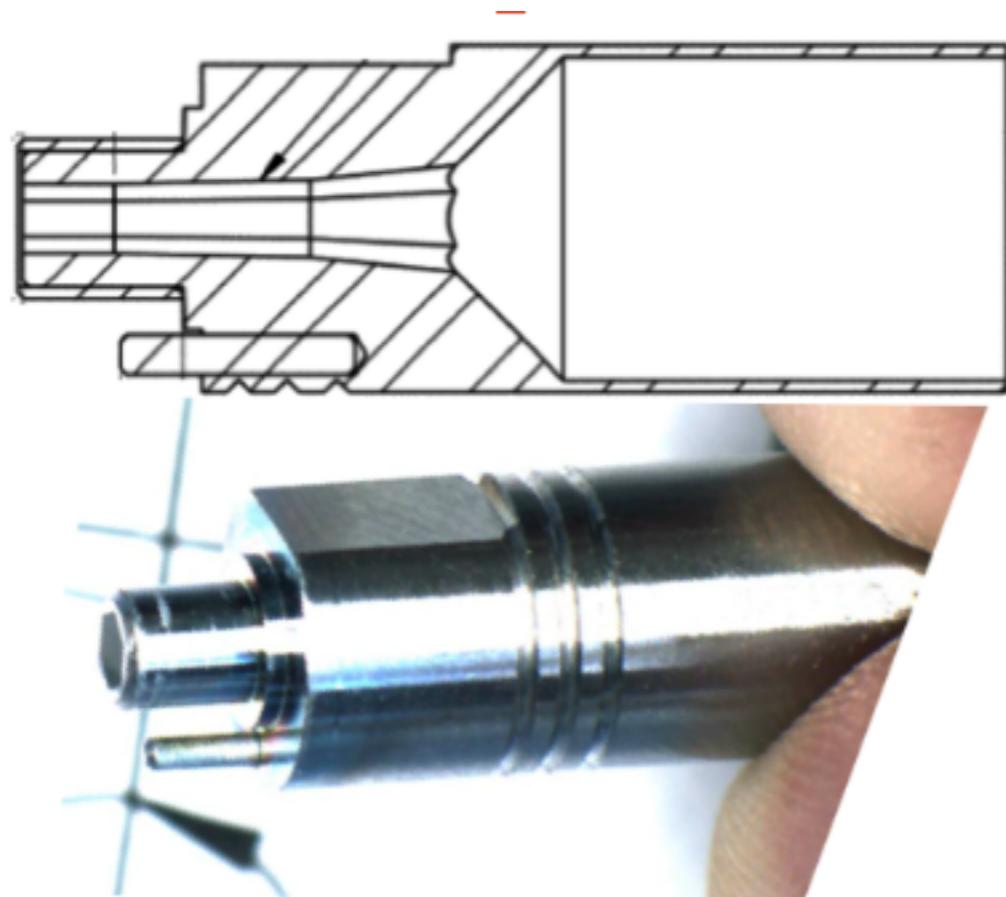
## MaNGA vs the World



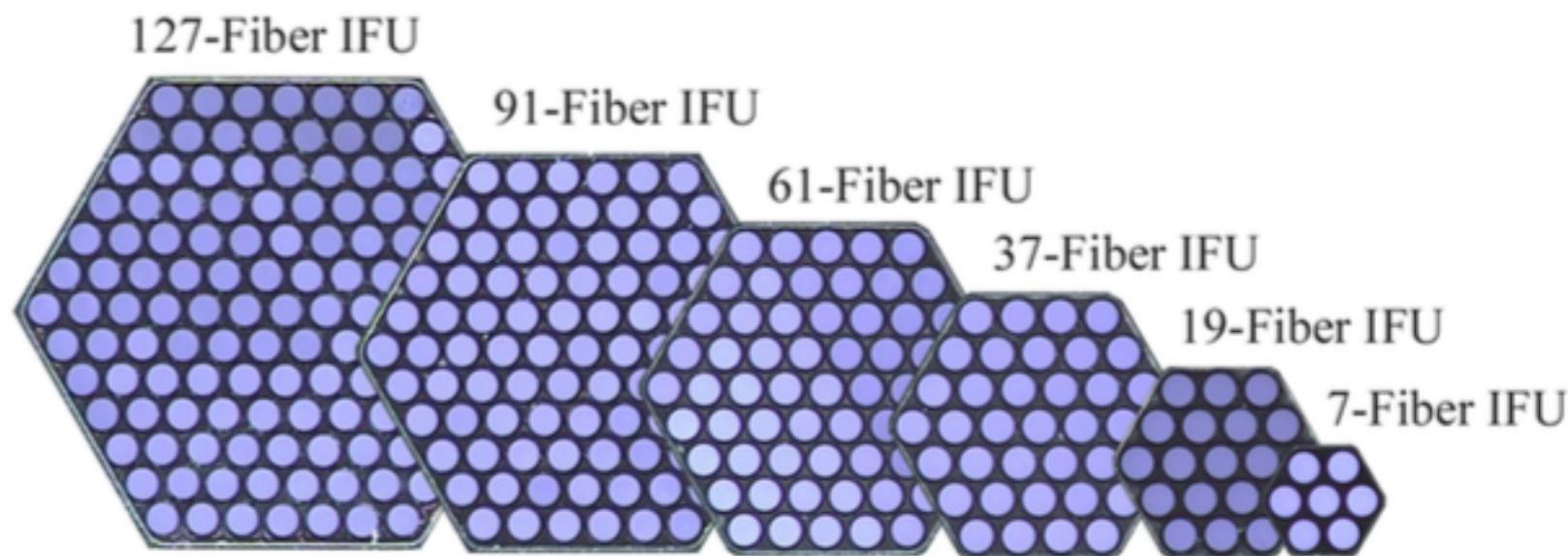
Bundy et al. (2015)

# MaNGA overview

## Hardware



- 17 IFU bundles
- 5 bundle sizes ranging from 19 to 127 fibres in hexagonal pattern
- 12 x 7-fibre mini-bundles for spectrophotometric calibration
- 92 single fibres for sky subtraction

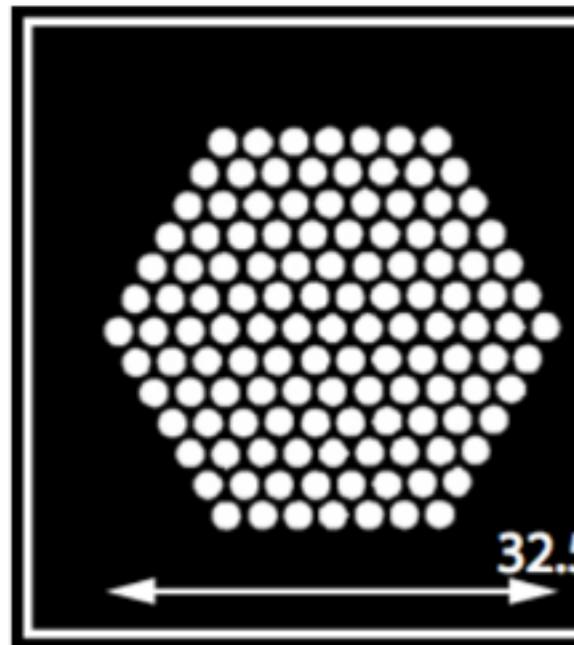


Drory et al. (2015)

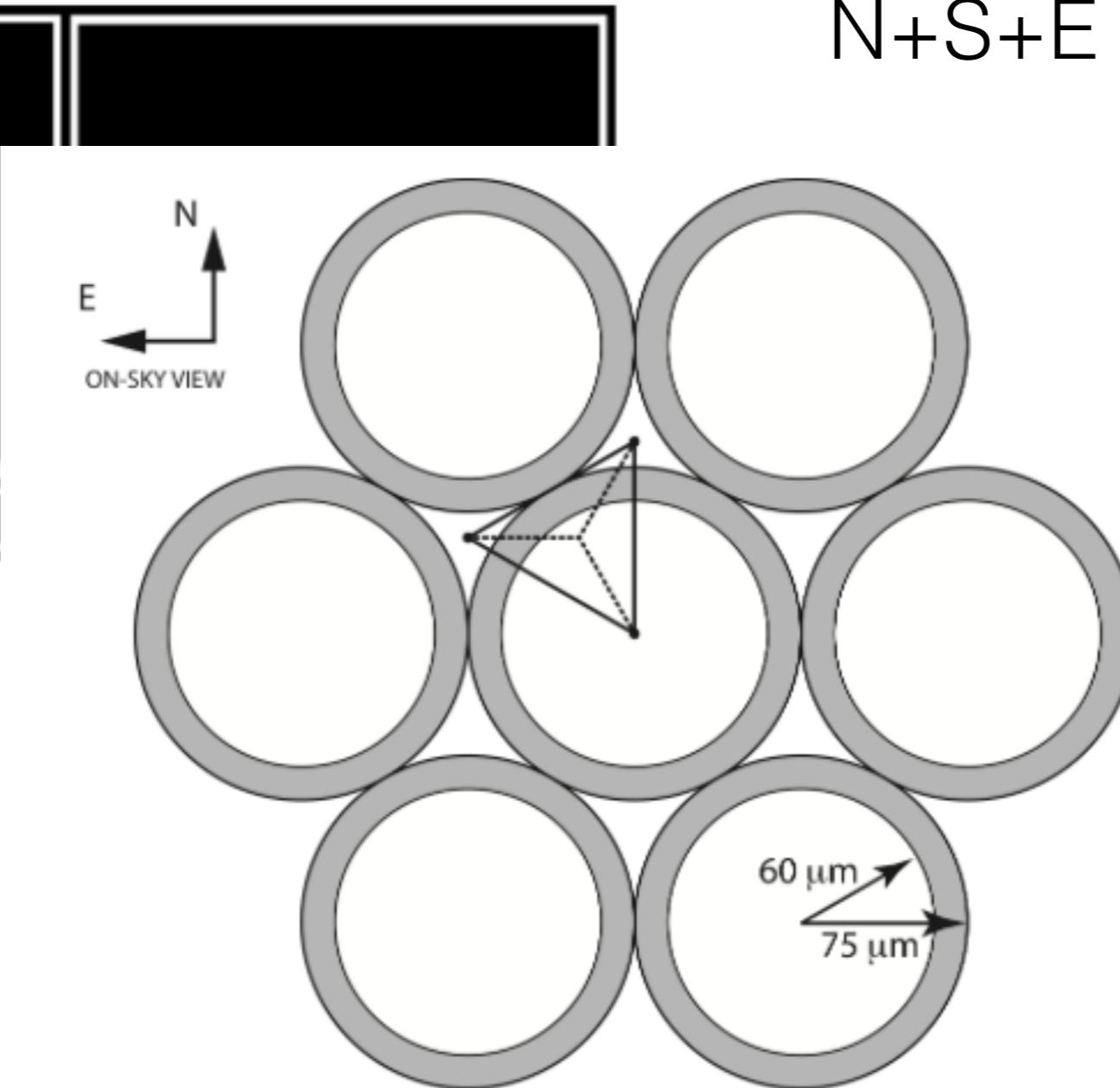
# MaNGA overview

## Dithering

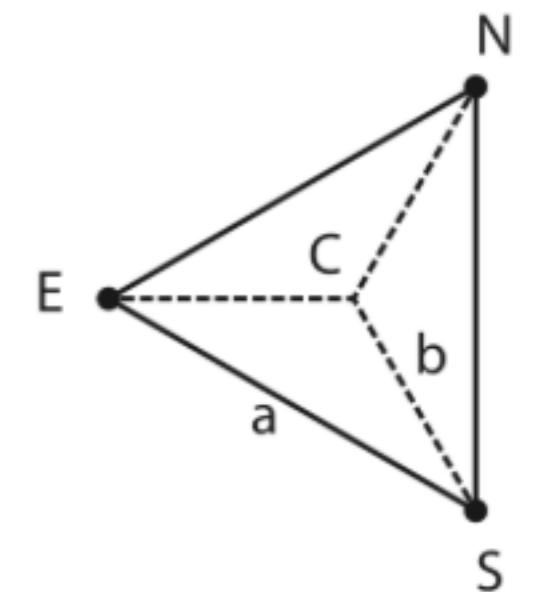
MaNGA



SAMI



**Set:** combination of  
N+S+E exposures

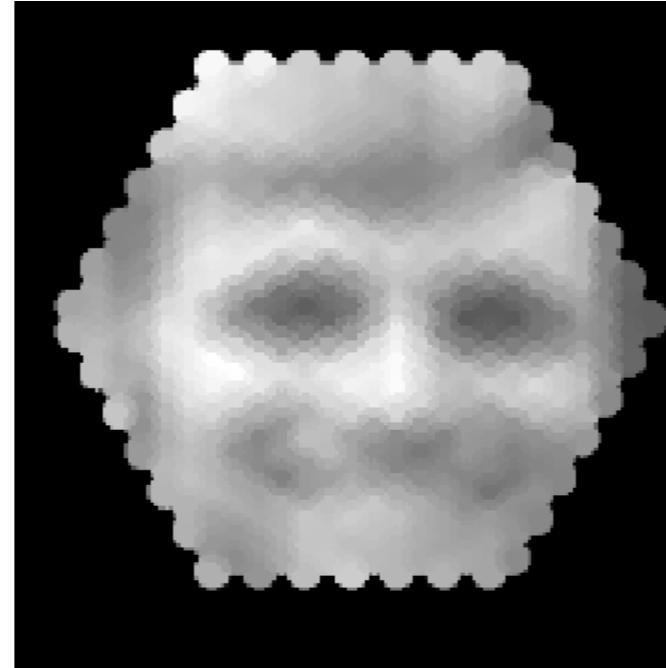
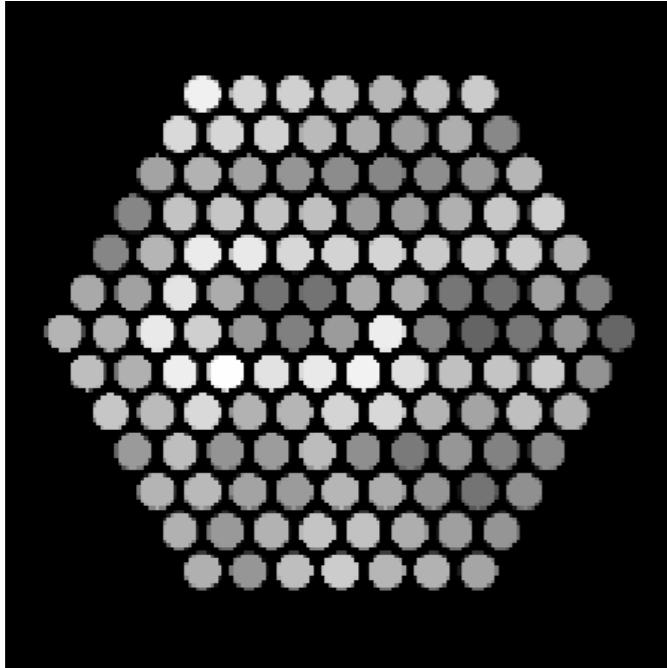


$$a = 86.6 \mu\text{m} = 1.44''$$
$$b = 50.0 \mu\text{m} = 0.83''$$

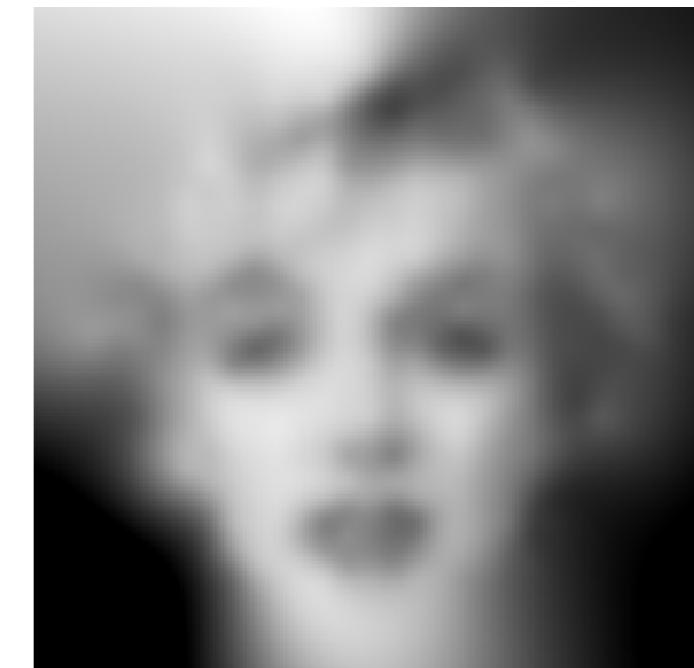
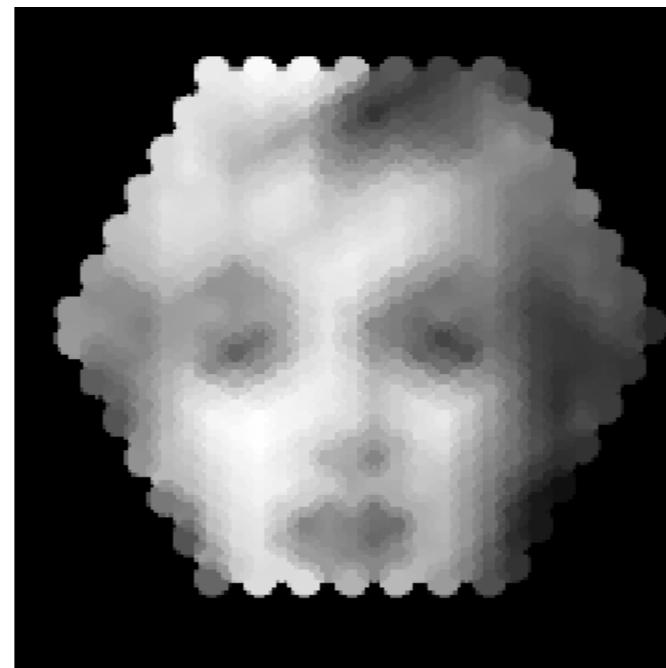
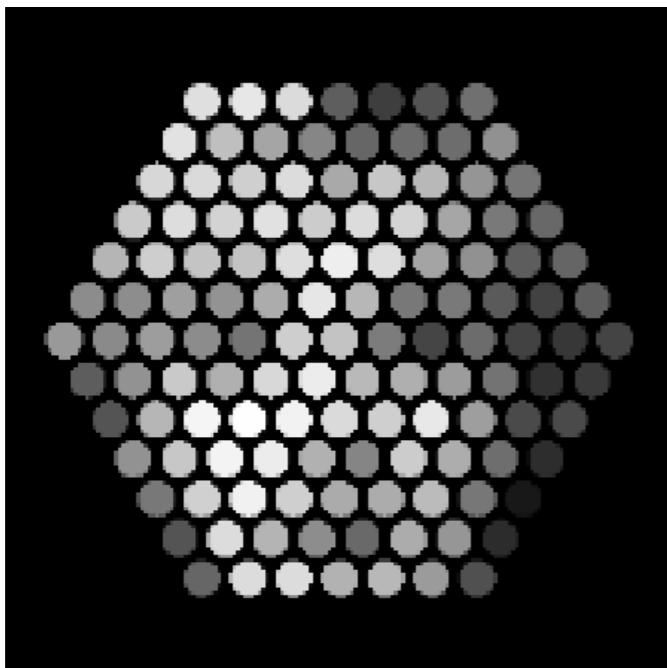
Law et al. (2015)

# MaNGA overview

## Dithering



Kevin ≠ Marilyn



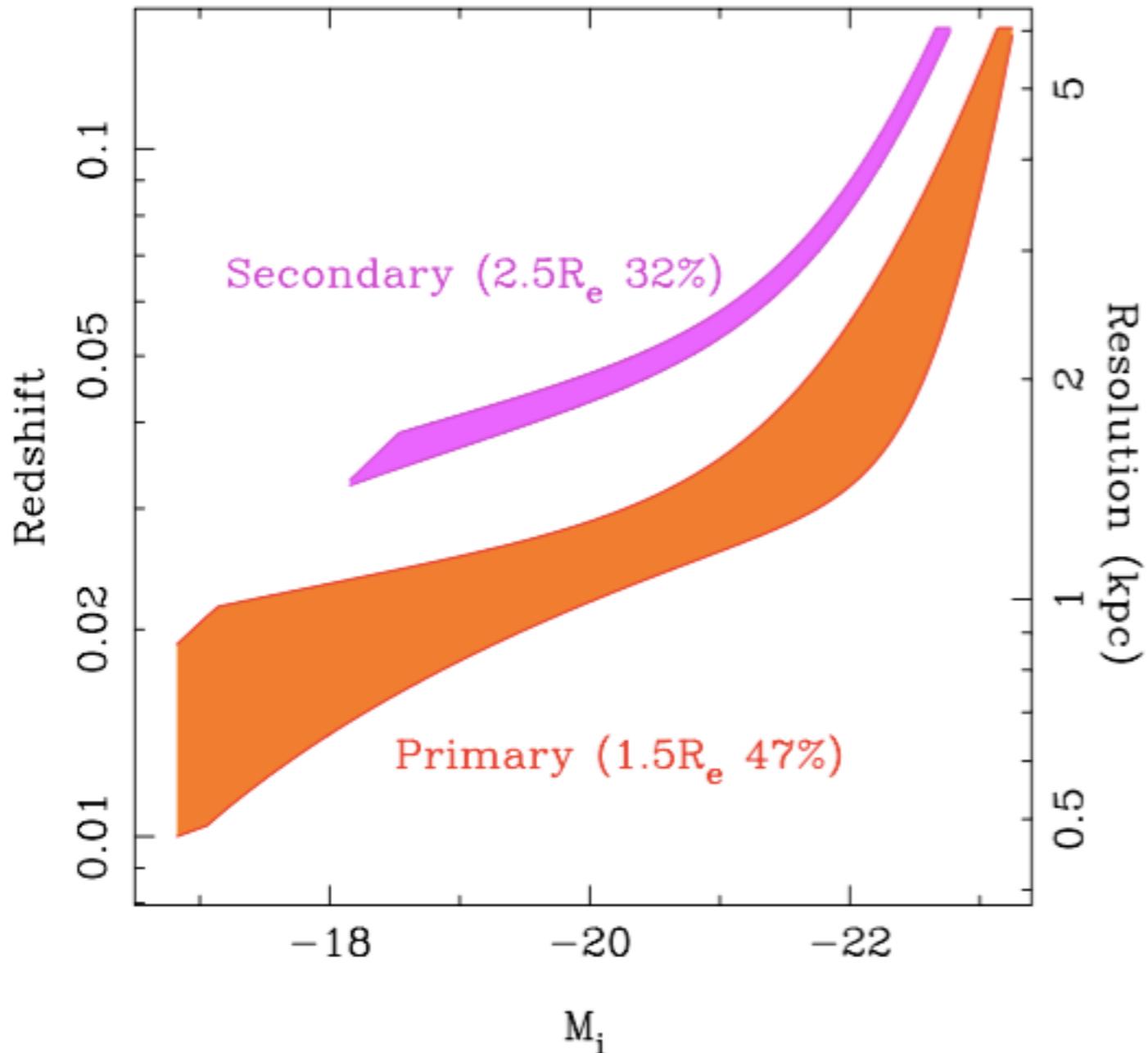
# MaNGA overview

## Hardware



# MaNGA overview

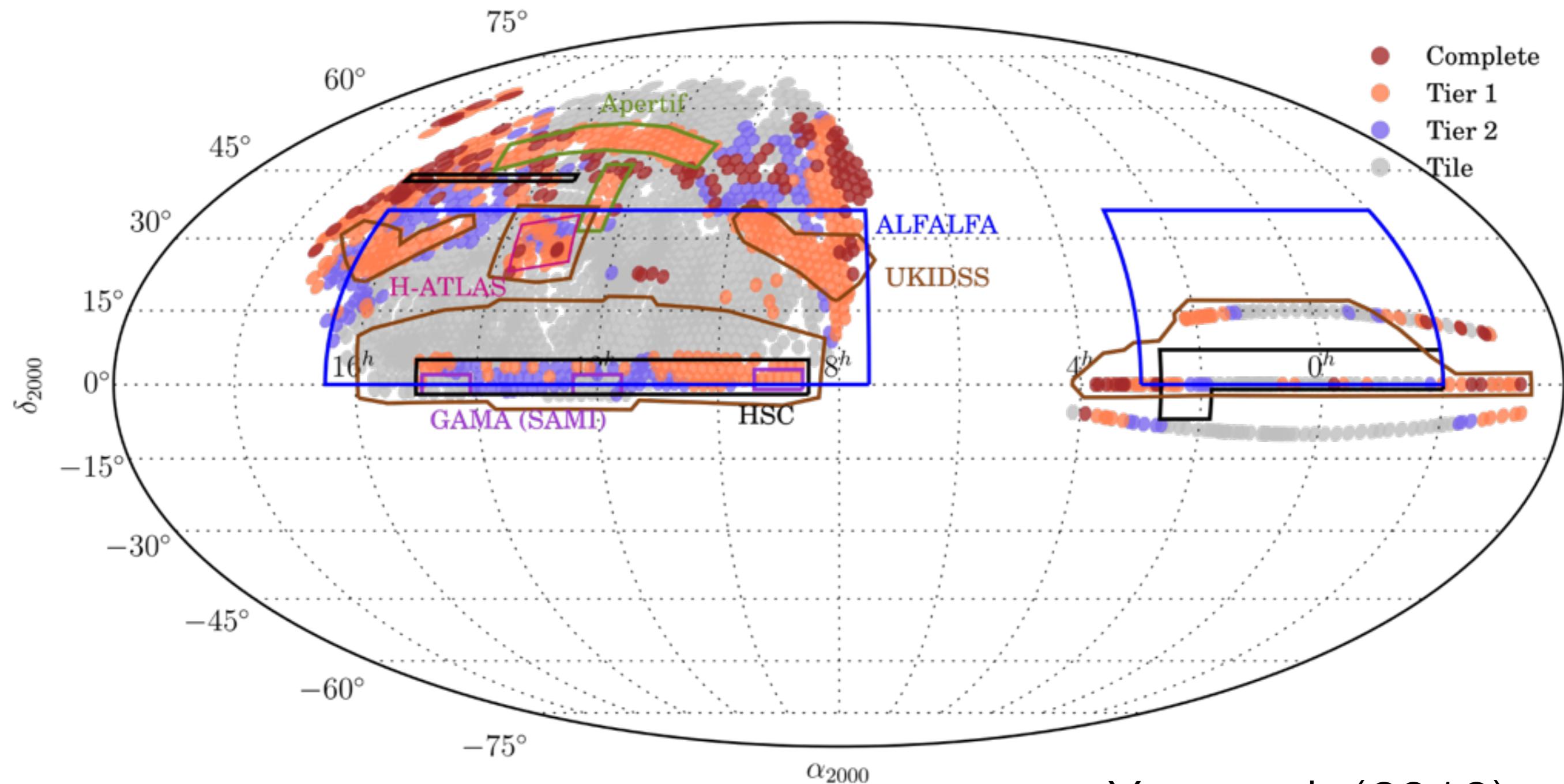
## Sample design



- $M_* > 10^9 M_\odot$
- Two main subsamples at 1.5 and 2.5  $R_{\text{eff}}$
- Flat distribution in  $M_*$
- Based on NASA Sloan Atlas v1
- 5-10% bundles allocated to ancillary programs

# MaNGA overview

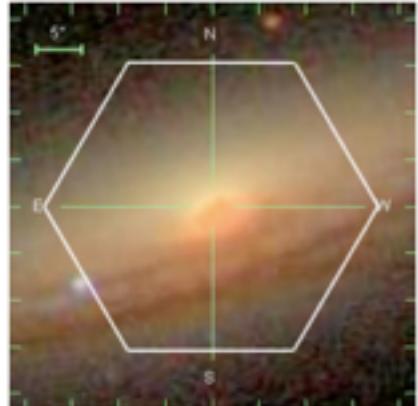
## Field selection



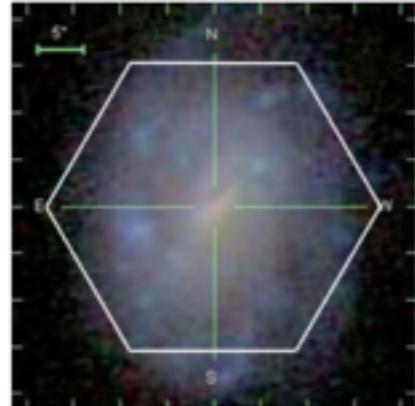
Yan et al. (2016)

# Science

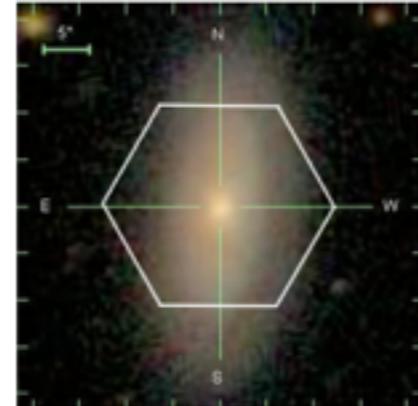
## Early science results



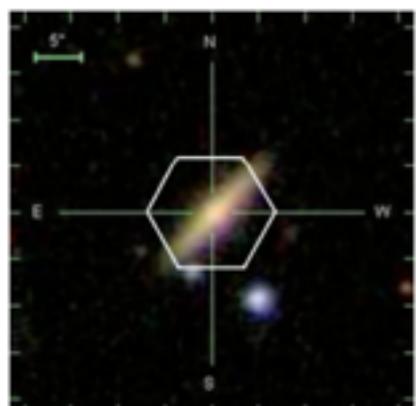
p11-127A



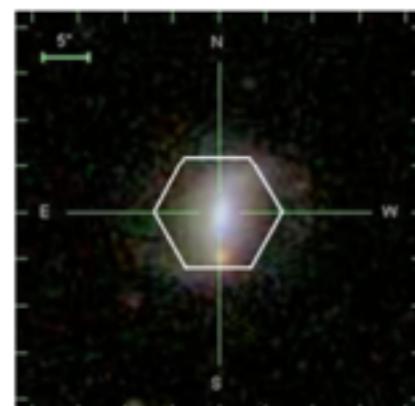
p11-127B



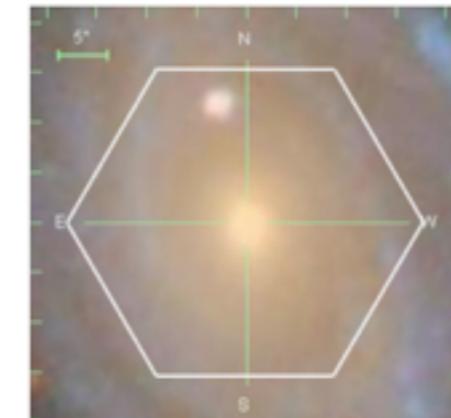
p11-61A



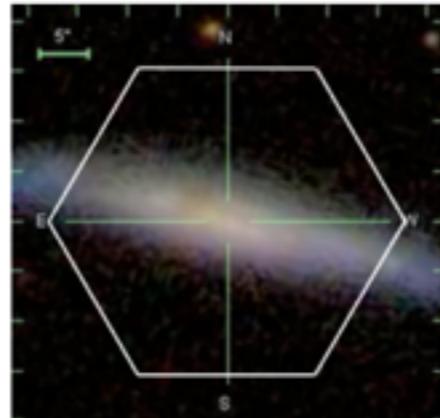
p11-19A



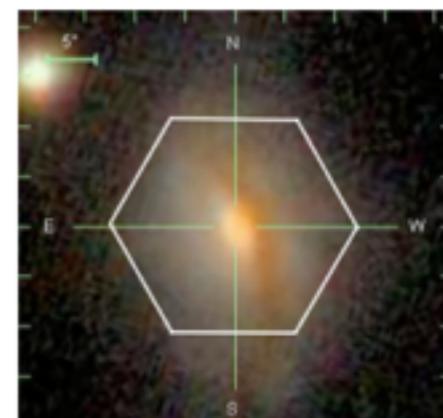
p11-19B



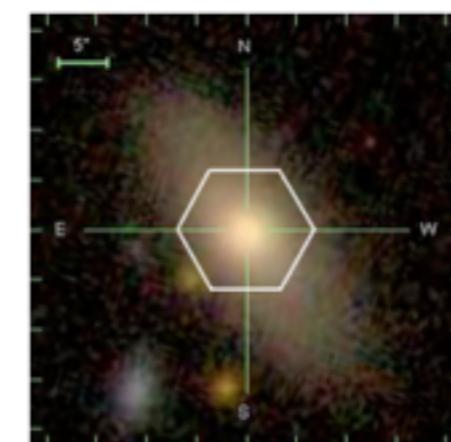
p9-127A



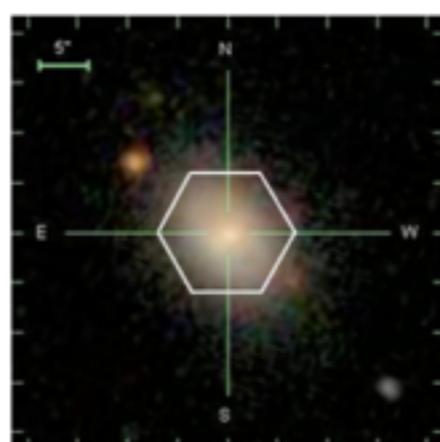
p9-127B



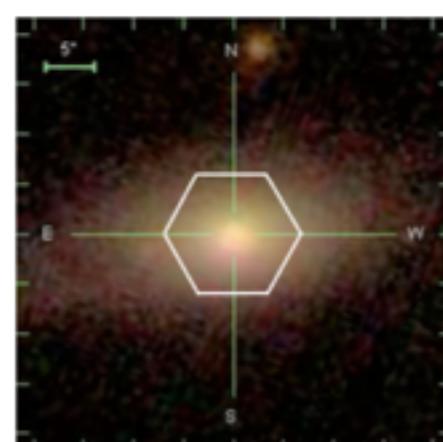
p9-61A



p9-19D



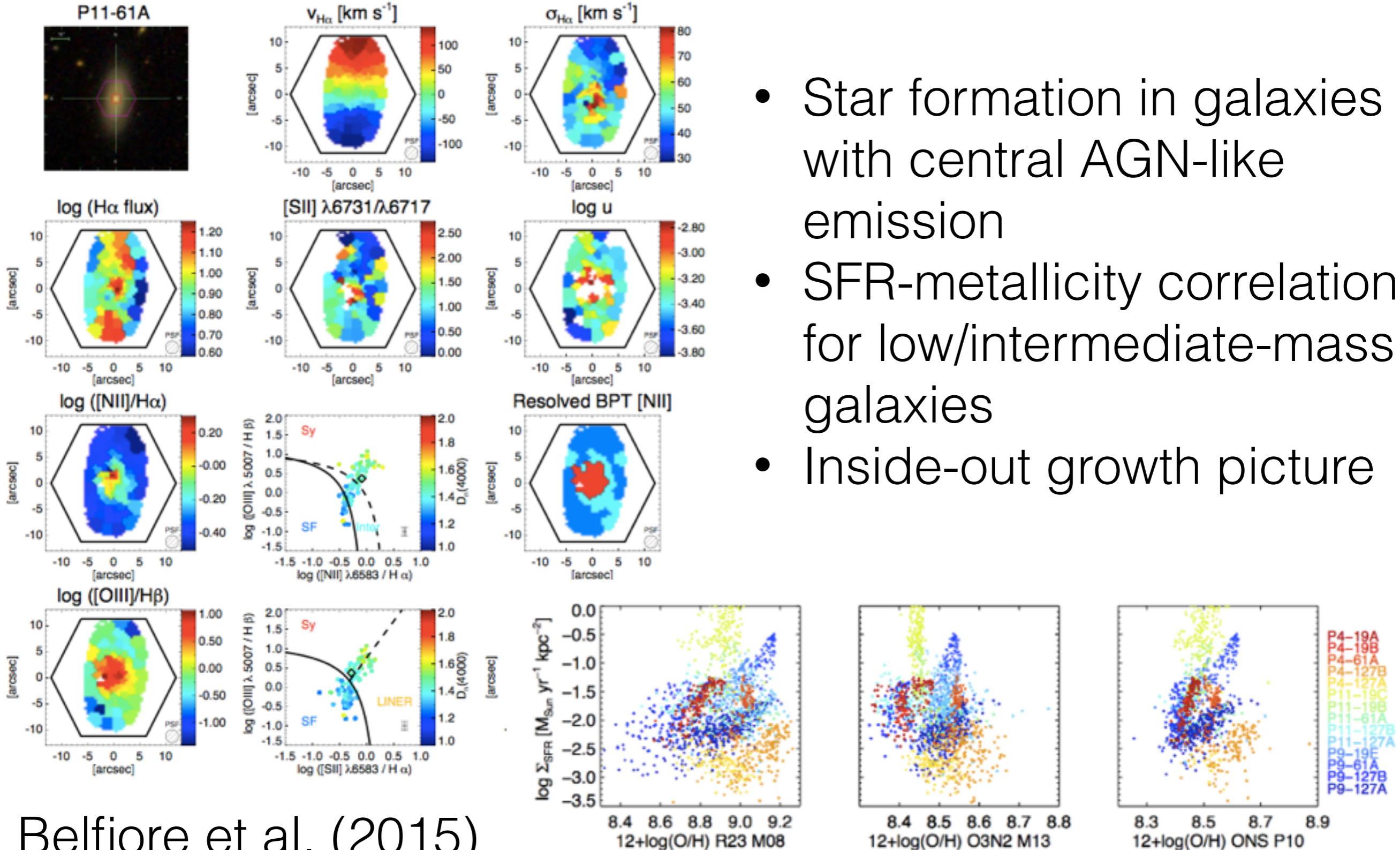
p9-19E



p9-19B

# Science

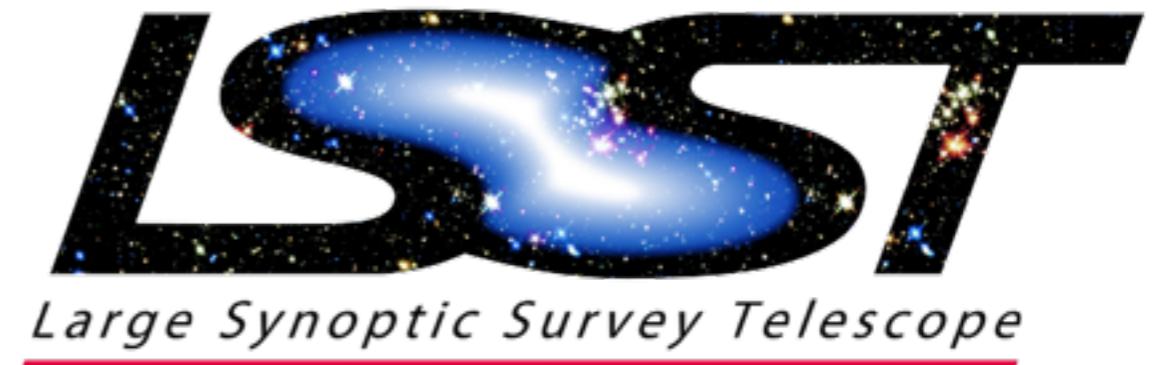
## Early science results: emission line diagnostics



Belfiore et al. (2015)

# Marvin Motivation

The Future: Survey Science Astronomy



- Homogeneous processing
- Statistically meaningful cross comparisons and population studies

# Marvin Motivation

## Limitations of Small Data Techniques

- Small data techniques: assume bandwidth and memory are not bottlenecks.
- At best inconvenient, at worst impossible.

# Marvin Motivation

## Limitations of Small Data Techniques



- Data access is bandwidth limited.
- Hampers exploration and reduces number of iterations.
- Results in simpler analyses that marginalize over interesting axes for logistical purposes.
- May not fit into RAM.
- Updated analyses multiply volume several fold.

# Marvin Motivation

## Logistics

- Significantly more managing of datasets and formats.
- Large barrier to entry for new users
- Logistics not taught in grad school and technologies evolve rapidly.
- Not aware of or don't have time to learn big data techniques.

# Marvin Motivation

## Logistics

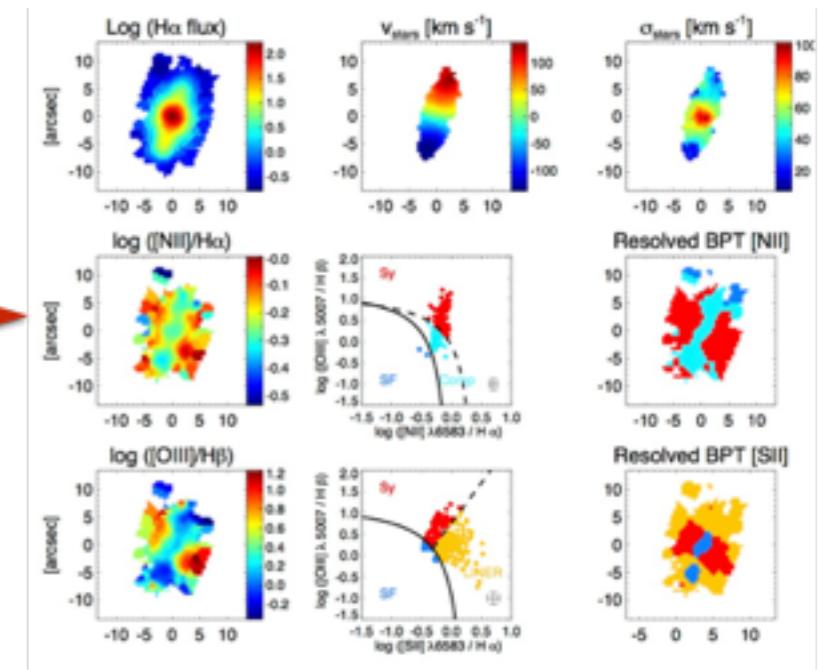
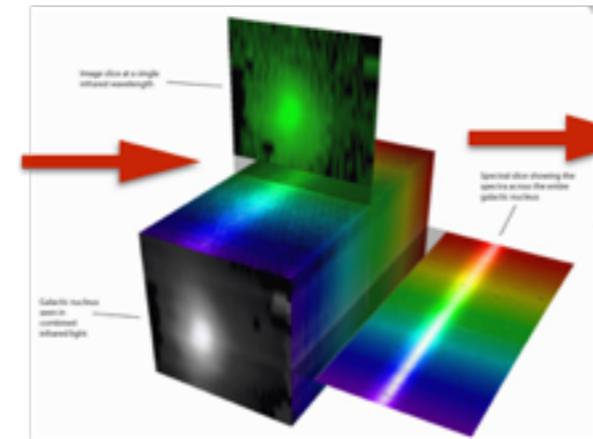
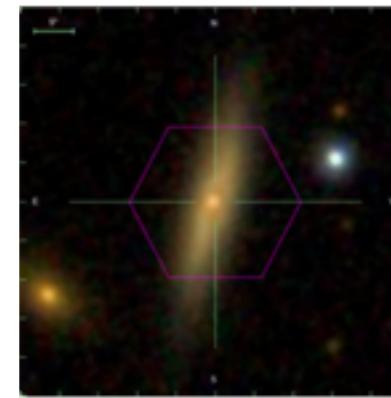
HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?  
(ACROSS FIVE YEARS)

		HOW OFTEN YOU DO THE TASK					
		50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
HOW MUCH TIME YOU SHAVE OFF	1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
	5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
	30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
	1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
	5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
	30 MINUTES	6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS	
	1 HOUR	10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS	
	6 HOURS			2 MONTHS	2 WEEKS	1 DAY	
	1 DAY				8 WEEKS	5 DAYS	

- Everyone has the same problems.
- Solve it once. Solve it right.

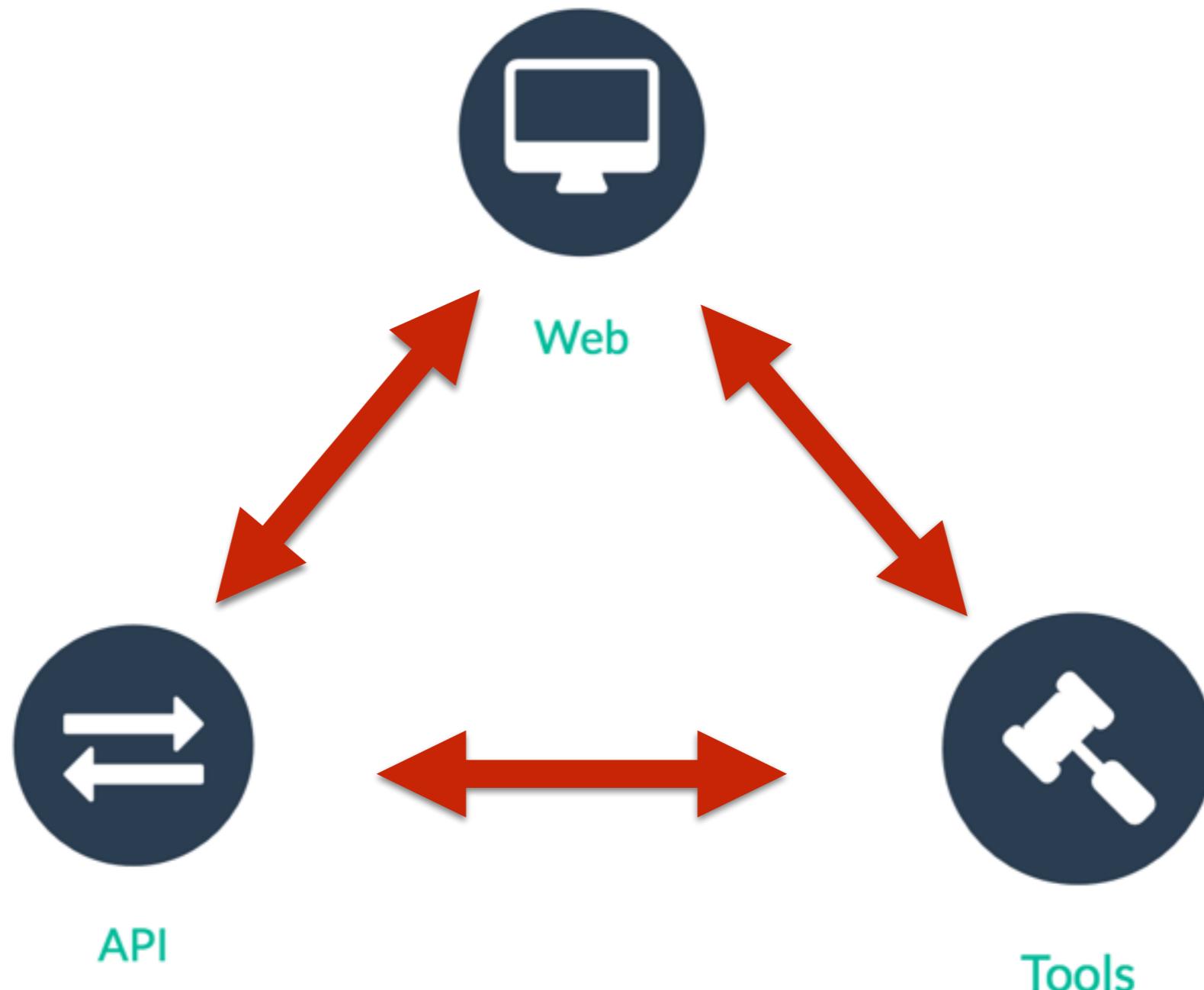
# Manga Data

- Assuming 6 years of survey time



- ~10,000 galaxy cubes
- ~3,000 spaxels per cube
- ~100 measured properties + 10,000 spectral elements per spaxel
- ~300 billion total data elements

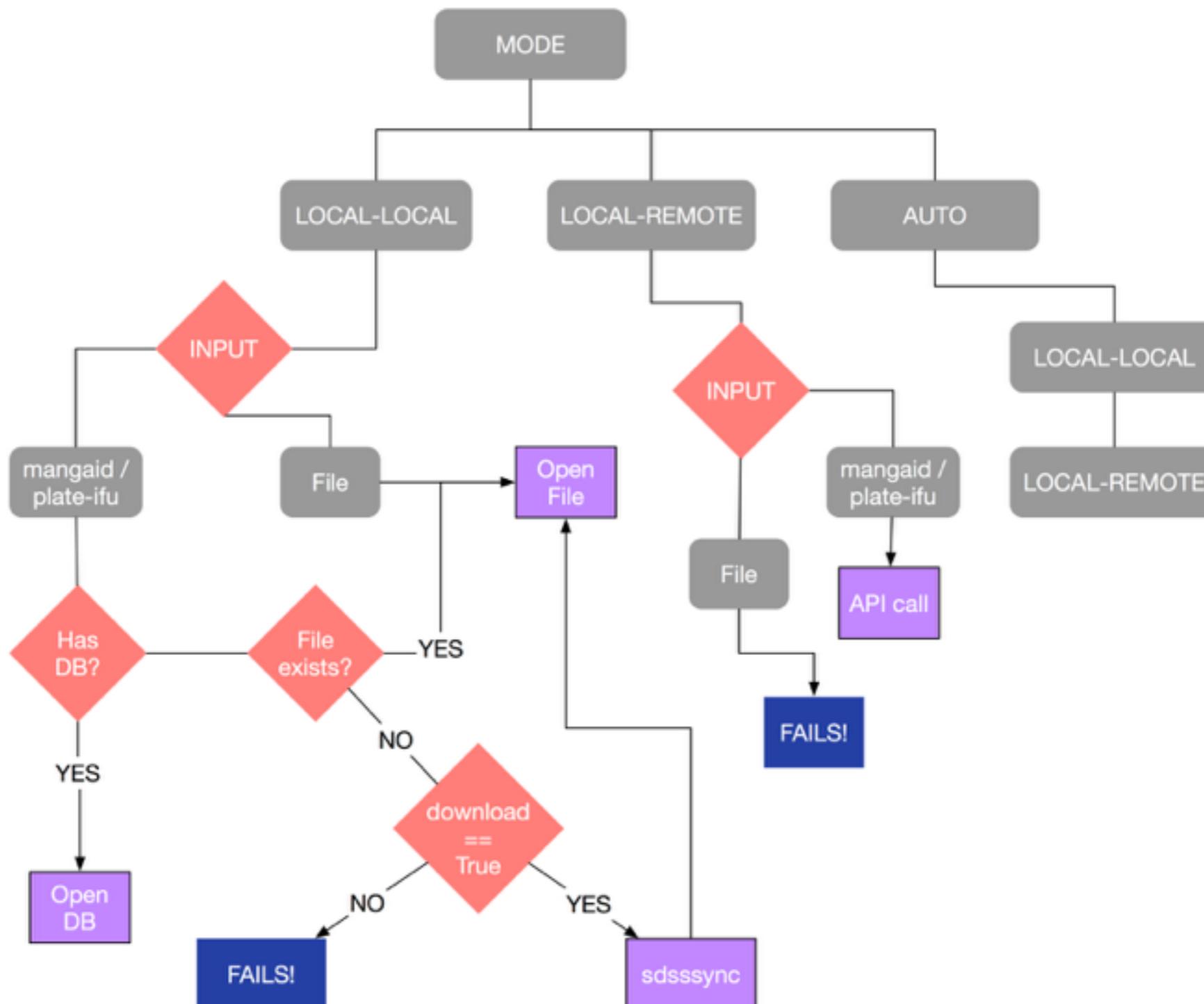
# Marvin Ecosystem



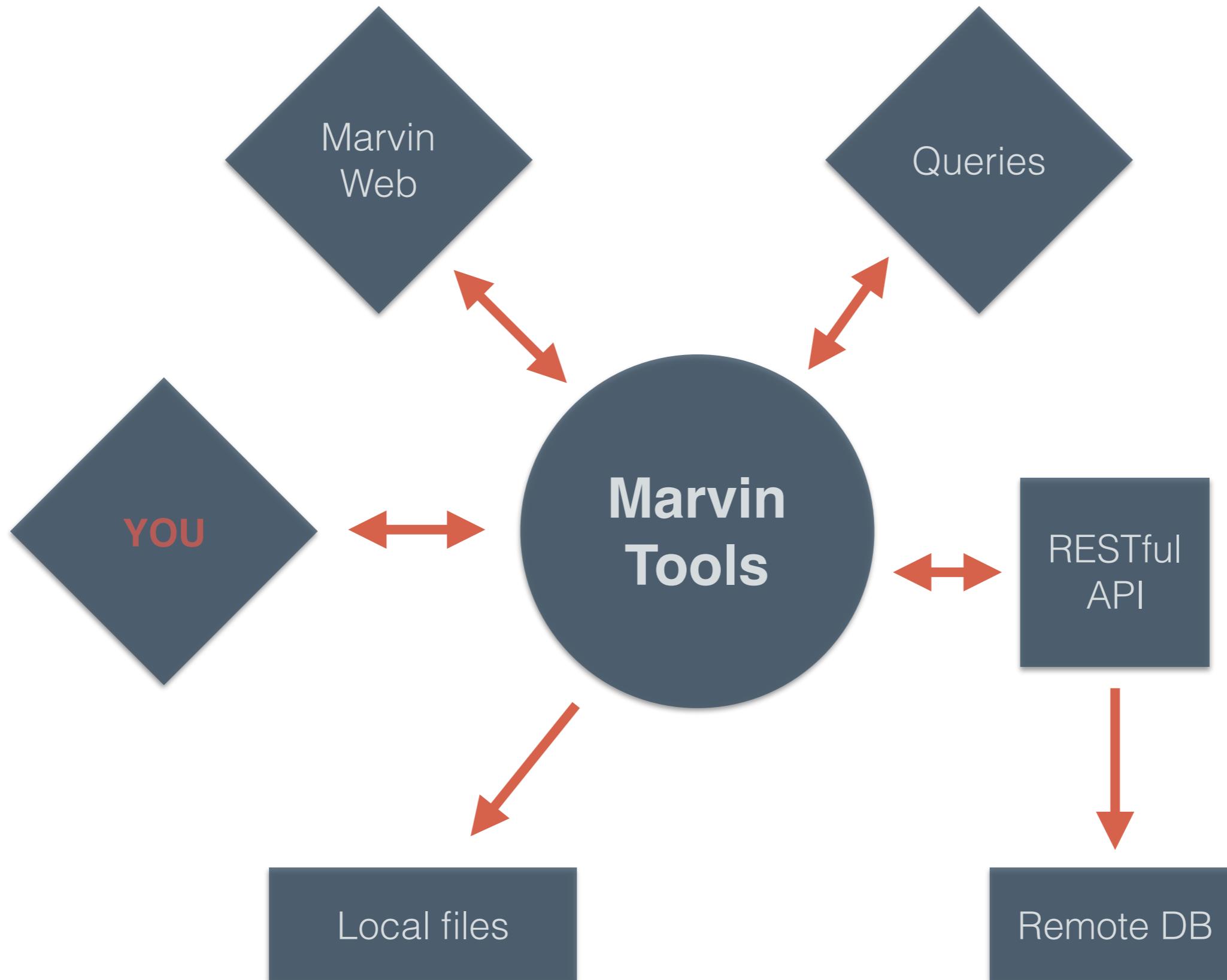
**Core Idea :** Smart Multi-Modal Data Access System

# Decision Tree

Local-Local vs Local-Remote decision tree



# Tools



# Queries

PostgreSQL



- Backend database
- Allows inter- / intra- galaxy searching
- Intuitive yet powerful query syntax
- Abstracts away the need to know SQL and database design

# Web



- The Front-End for new users into MaNGA
- Warm, Streamlined, Intuitive, Accessible (like these bullets)
- Interactive Visual Exploration
- Uses the same Marvin Tools for robustness (or consistent failures)

# Demo

- “Let’s do it live...”

