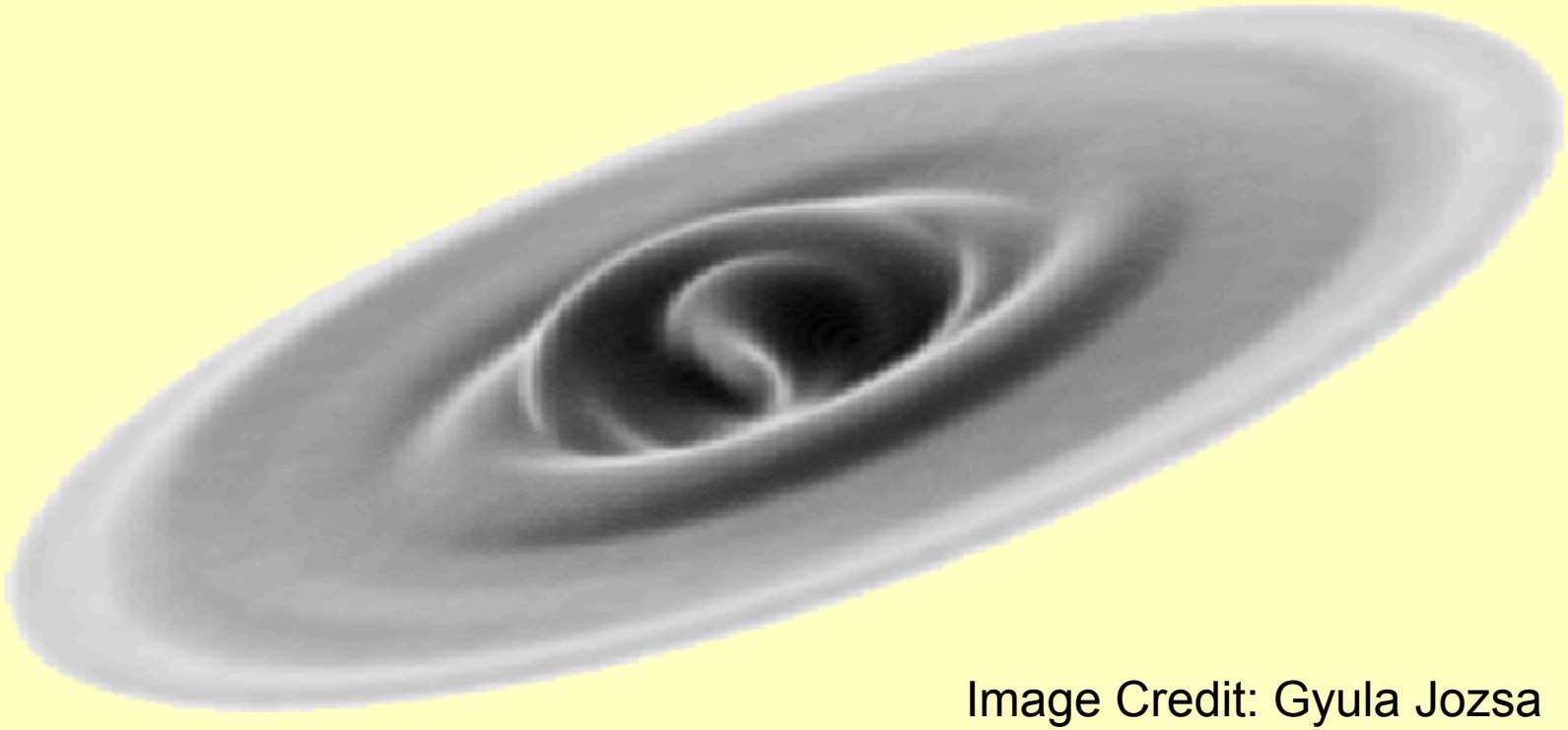


Image Credit: Gyula Jozsa

www.csiro.au

Leveraging visualisation for WALLABY

Russell Jurek

WALLABY Simulations Fest 2011,
November 23rd 2011

CSIRO

WALLABY/DINGO WG5

- Baerbel Koribalski (PI)
- Lister Staveley-Smith (PI)
- Martin Meyer (PI)
- Russell Jurek (co-chair)
- Chris Fluke (co-chair)
- Amr Hassan
- Andreas Wicenec
- Gerhardt Meurer

- Fortnightly meetings starting soon

CSIRO

Talk Outline

- The trick to exploiting visualisation
- Visualisation uses
  - Data mining
  - Datacube quality control
  - Source finding
  - Testing parameterisation
  - Source classification
- Citizen science
- Examining weird objects
- Summary

CSIRO

The trick

The trick

- Visualisation is good for rapid, qualitative analysis
  - Visualisations are contextual and `information dense'
  - Visualisations provide perspective
- Images are engaging
- Leverages human pattern recognition capabilities
- Visualisation is the key to citizen science

CSIRO

Visualisation uses

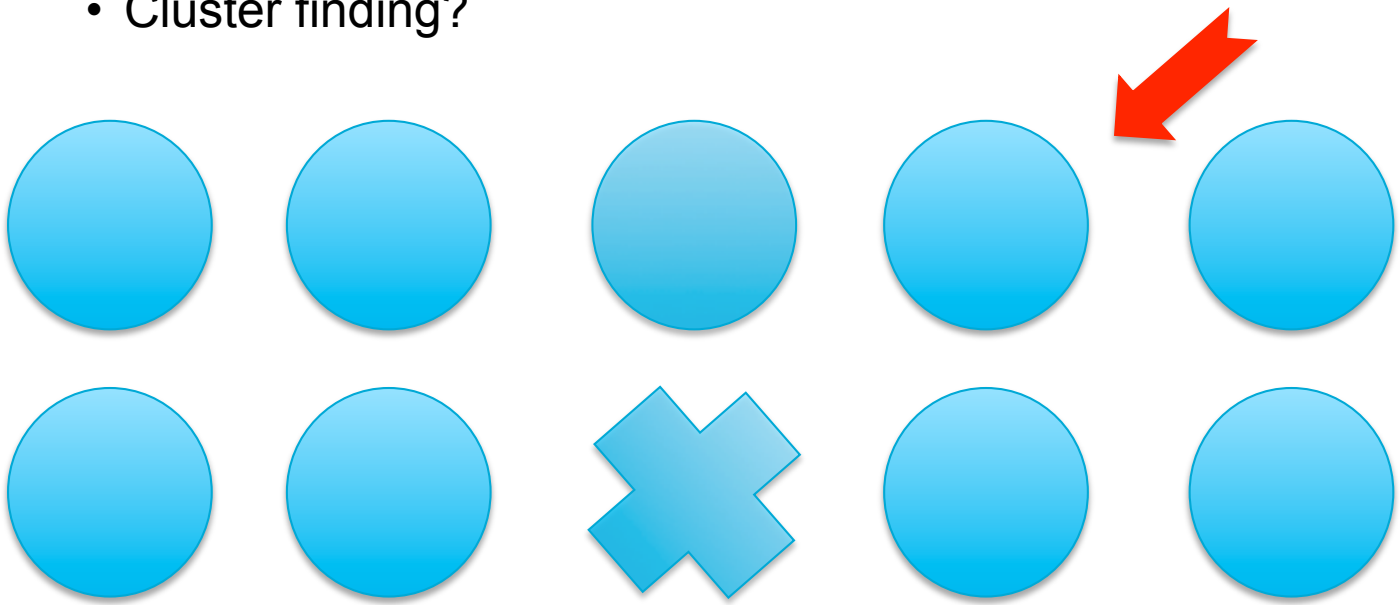
Visualisation uses

1. Data mining
2. Datacube quality control
3. Source finding
4. Testing parameterisation
5. Source classification

Data mining

- Exploit context
  - Easiest way to spot the weird and wonderful
- Exploit perspective
  - Identify large scale structure
  - Identify selection effects
- We need to be creative
  - Cluster finding?

Which of these is not like the others?



Visualisation uses

1. Data mining
2. Datacube quality control
3. Source finding
4. Testing parameterisation
5. Source classification

Datacube quality control

- 2 levels of datacube quality control
- Internal QC
  - Part of ASKAP pipeline
  - Automated tests with quantitative results
- External QC
  - Semi-automated and manual tests
  - Incorporates visualisation
  - Addresses weaknesses of automated tests

CSIRO

Internal QC

- Measure global noise level of datacube
- Test sensitivity/noise variation of datacube
  - Measure RMS in running box
  - Compare to sensitivity model and survey requirements
- Test continuum subtraction of datacube
  - Compare continuum subtracted and blank line-of-sights
  - Are the voxel flux distributions the same?
    - Apply Kolmogorov-Smirnov test
    - Compare medians and interquartile ranges
- Test for presence of known sources
- Test if source count and distribution is sensible
  - Use predicted counts and predicted variance
- Test if source parameters are sensible

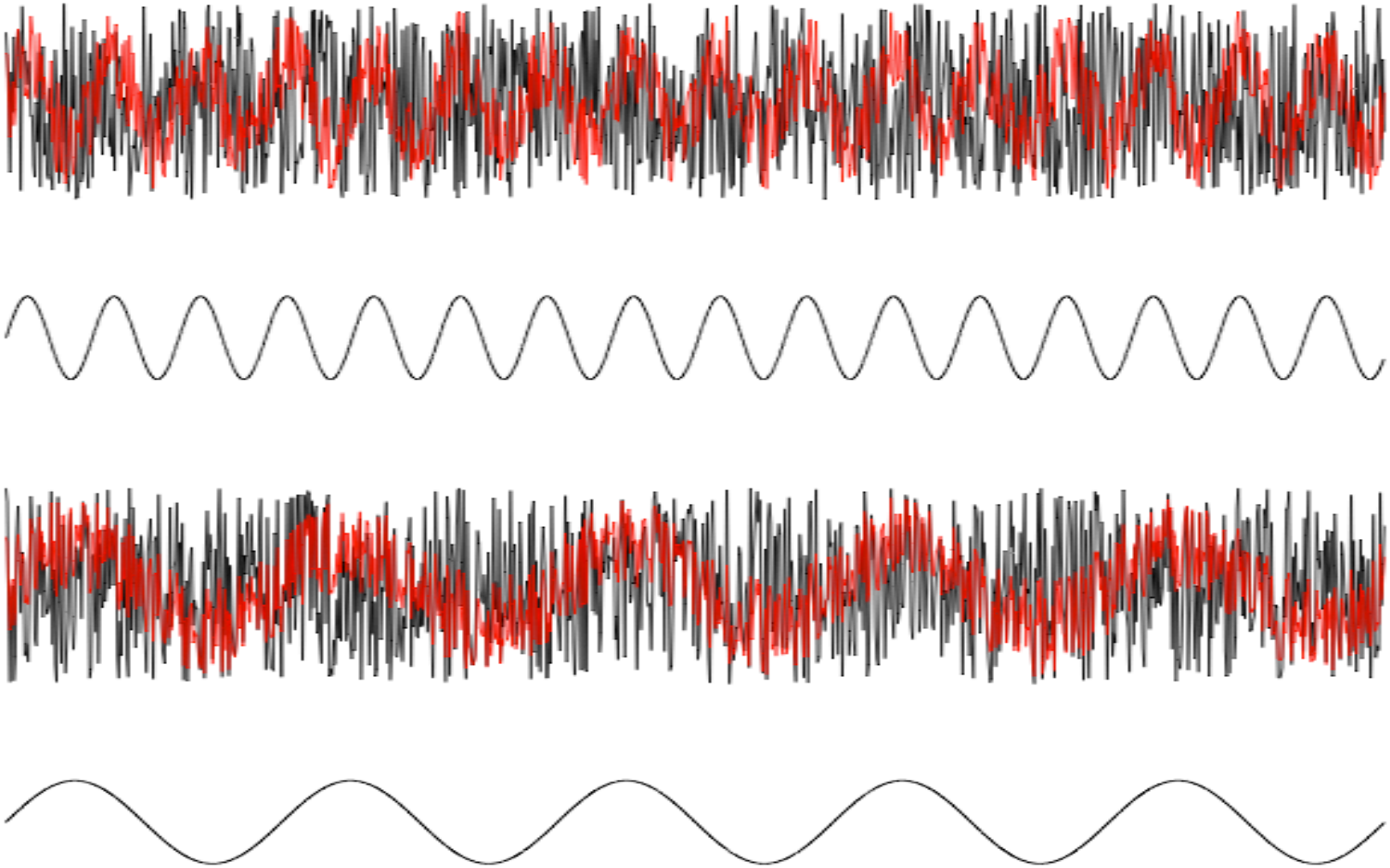
CSIRO

The problem with automated tests

- Measured quantities describe a macroscopic state
  - Potentially large microcanonical ensemble
  - Scale dependent
- Relies on picking the right test
- Effects might be intermittent
  - Another mechanism for scale dependence

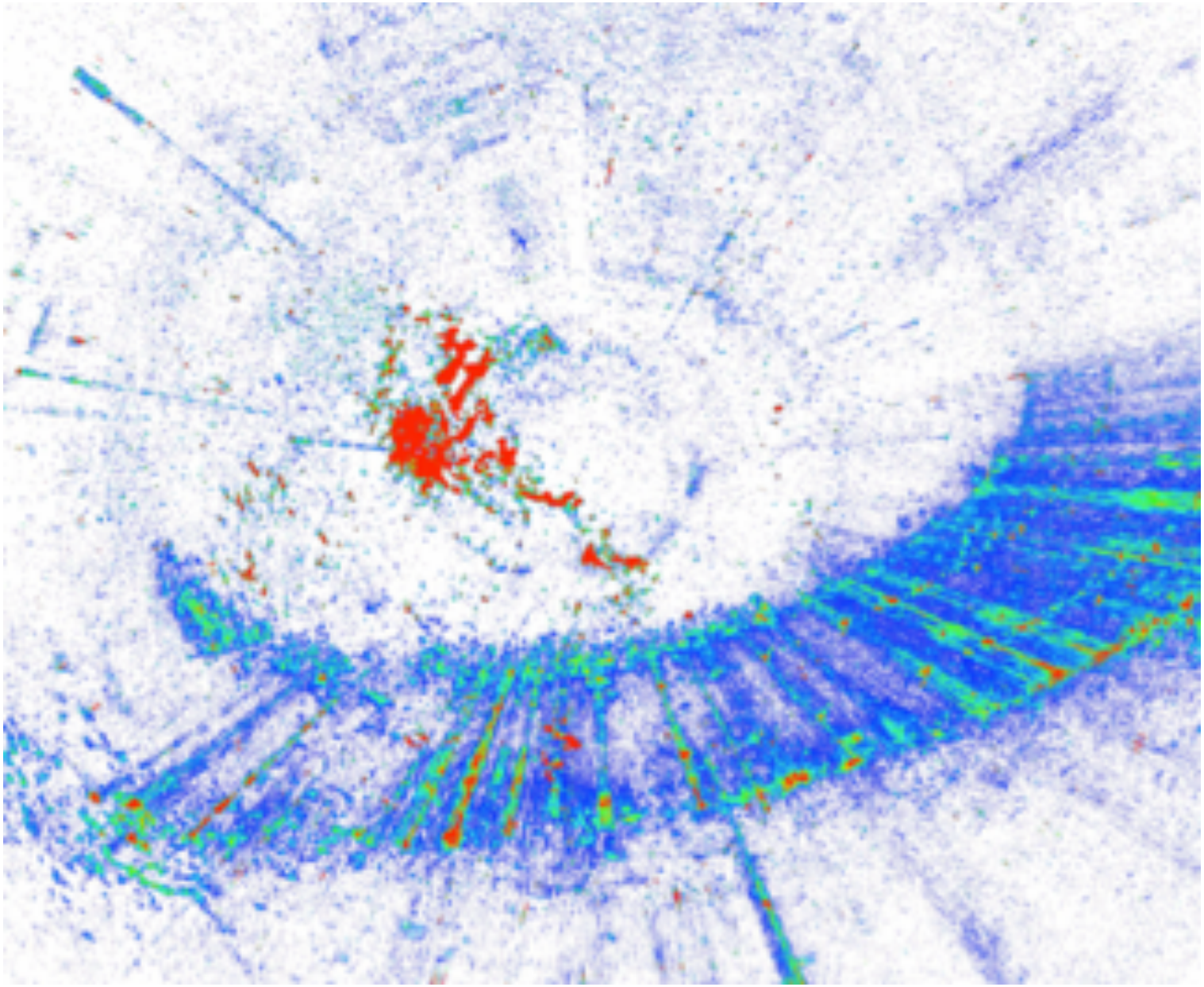
CSIRO

The problem with automated tests



CSIRO

The problem with automated tests



CSIRO

External QC

- Goal is to address issues with automated QC tests
- Make use of low resolution datacube
- Visualise datacubes or datacube sub-regions
- Visualise running RMS of datacubes
- Overlay catalogue, known sources and sky model on datacube visualisation
- Visualise calibration residuals
    1. Measure calibration residual for known sources
    2. Plot as a function of position in the datacube
    3. Look for trends/distinct regions

CSIRO

Visualisation uses

1. Data mining
2. Datacube quality control
3. Source finding
4. Testing parameterisation
5. Source classification

Source finding

- Citizen science?
- Best on low resolution datacubes?
- Procedure
  1. Remove sources from datacube
  2. Examine datacube for oddities/sources

CSIRO

Visualisation uses

1. Data mining
2. Datacube quality control
3. Source finding
4. Testing parameterisation
5. Source classification

Testing parameterisation

- Citizen science?
- Show overlap of sources and models/masks:
  - Compare TiRiFic model projection and cubelet
  - Compare velocity field and model velocity field
  - Compare rotation curve and model rotation curve
  - Compare source finding binary mask and cubelet
  - Compare integrated spectrum and cubelet integrated spectrum

CSIRO

Visualisation uses

1. Data mining
2. Datacube quality control
3. Source finding
4. Testing parameterisation
5. Source classification

Source classification

- Citizen science
- Repeat galaxy zoo with WALLABY
- Additional classifications:
  - Multi-wavelength overlaps
    - Degree of overlap
    - Degree of overlap correlation
  - Check/Find object merging

Citizen Science

The benefit of citizen science

- 1,500,000 sources = 13 FTEs (using CSIRO definition)
- 1 FTE = 8 FHRs (5 hour regulars, 5hr/week for 48 weeks)
- 1,500,000 sources = 104 FHRs

Eyeballs	FHRs
10^6	70
10^7	695
14.41 Million	1000
10^8	6945

MANY HANDS MAKE LIGHT WORK

CSIRO

Implementation

- Most of our operations are comparisons
  - We can easily set up fuzzy answers
  - A single tool is possible
- Server that provides HI images and a random overlay
- Overlays correspond to different comparisons
- Generate catalogue of classifications
  - Identifies the weird and wonderful

IT'S FEASIBLE ➔ EXCITING

CSIRO

Examining weird objects

WALLABY HI Analysis Tool

- What do we do with the weird stuff?
  - HI poor and HI rich galaxies
  - No obvious multi-wavelength counterpart
- Let's do the simple thing
  - Leverage citizen science to refine the list of weird things
  - Look at every object in this (hopefully) small list
- WALLABY HI Analysis Tool (WHAT)
  - Single source analysis tool
  - Single source visualisation
  - Multi-wavelength overlays
  - S'Finder mask overlay?
  - Relative position in parameter space

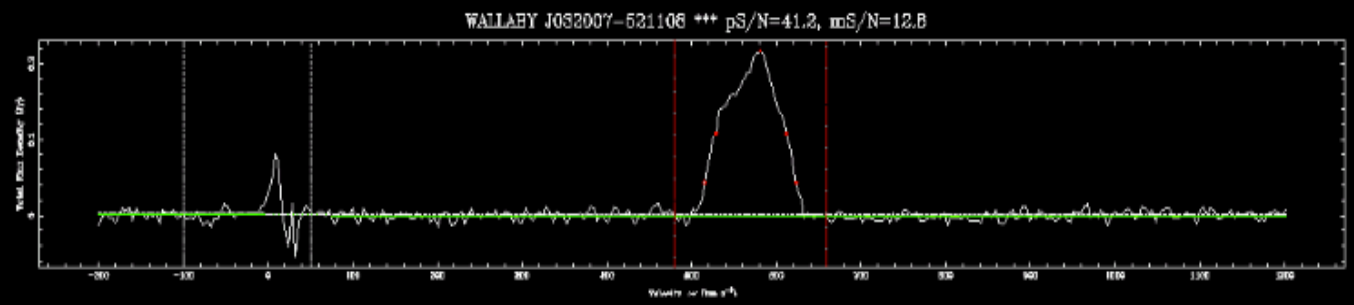
CSIRO

Image credit:
Baerbel Koribalski

WALLABY HI Analysis Tool

Wallaby HI Sources | hello :) | candidate list | Upload *.txt file | Start Row # | 321 | exit | HELP

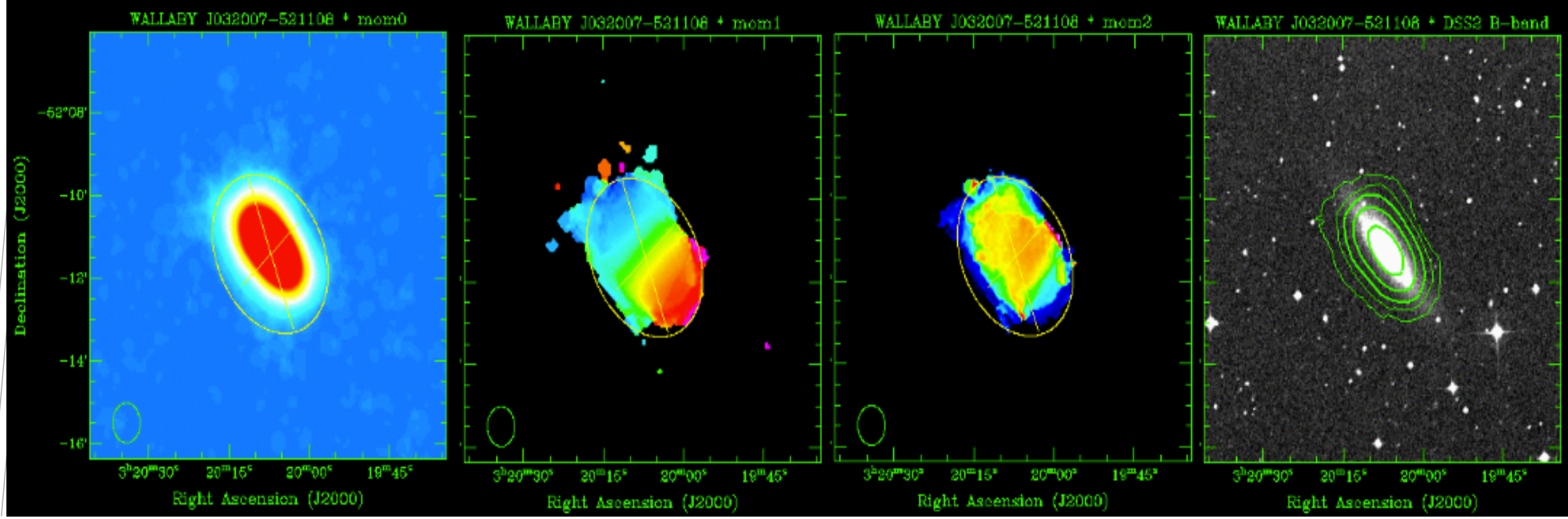
Cube	#	Ra, Dec (J2000)	velocity range (km/s)	dimensions (arcsec^2)	M_{HI} (Msun)	Quality Flag	NED-ID (best guess)	Comments
1	321	03:20:07 -52:11:08	440 - 660	250 x 200 pa=60	5.2 x 10^8	Excellent	NGC 1311	



Measured HI properties:
vsys = 570.3 +- 0.3 km/s
mom1 = 570.2 +- 0.3 km/s
w50 = 83.1 +- 0.7 km/s
w20 = 108.5 +- 1.5 km/s
FHI = 16.13 +- 0.42 Jy km/s

ACCEPT | CHECK | REJECT

Smooth HI spectrum: Hanning-1 * Adjust spatial extent: Point Source Choose overlay: DSS2 Blue



Summary

Summary

- Visualisation complements automated QC tests
- Visualisation is the key to unlocking citizen science
- We can leverage citizen science to find the weird (interesting) things, which we can analyse in detail
  - WALLABY HI Analysis Tool (WHAT)

CSIRO