New approaches to Volume and Velocity challenges of Modern Astronomy

Dany Vohl

Big Data's "3 Vs"

Volume

information
required to be **stored**,
processed and
transfered

Velocity

at which
information needs
to be **stored**
or **accessed**

Variety

of
data formats
(FITS, HDF5, …)

Big Data's "3 Vs"

Volume

information
required to be **stored**,
processed and
transfered

Velocity

at which
information needs
to be **stored**
or **accessed**

Variety

of
data formats
(FITS, HDF5, …)

Fundamental problems of Petascale Astronomy Era

Storage, Transfer, Visualisation, & Analysis

Spectral Cube Surveys

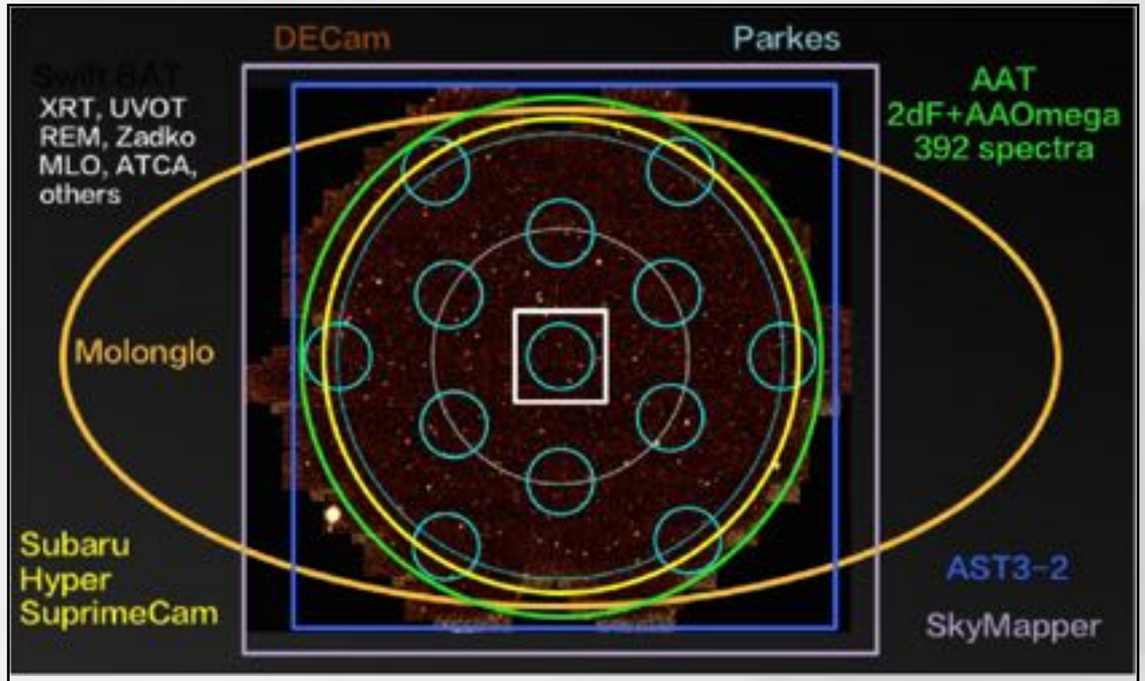
Apertif Cube Dimensions

0.25 TB / cube (Verheijen et al. 2009)

20,000 spectral-cubes

each containing ~100 sources

(Punzo et al. 2015)



Andreoni & Cooke (2018)

Time-critical remote surveys

(Cooke et al. In prep; Andreoni et al. 2017; Meade et al. 2017; Vohl et al. 2017a)

Fundamental problems of Petascale Astronomy Era

Storage, Transfer, Visualisation, & Analysis

Spectral Cube Surveys

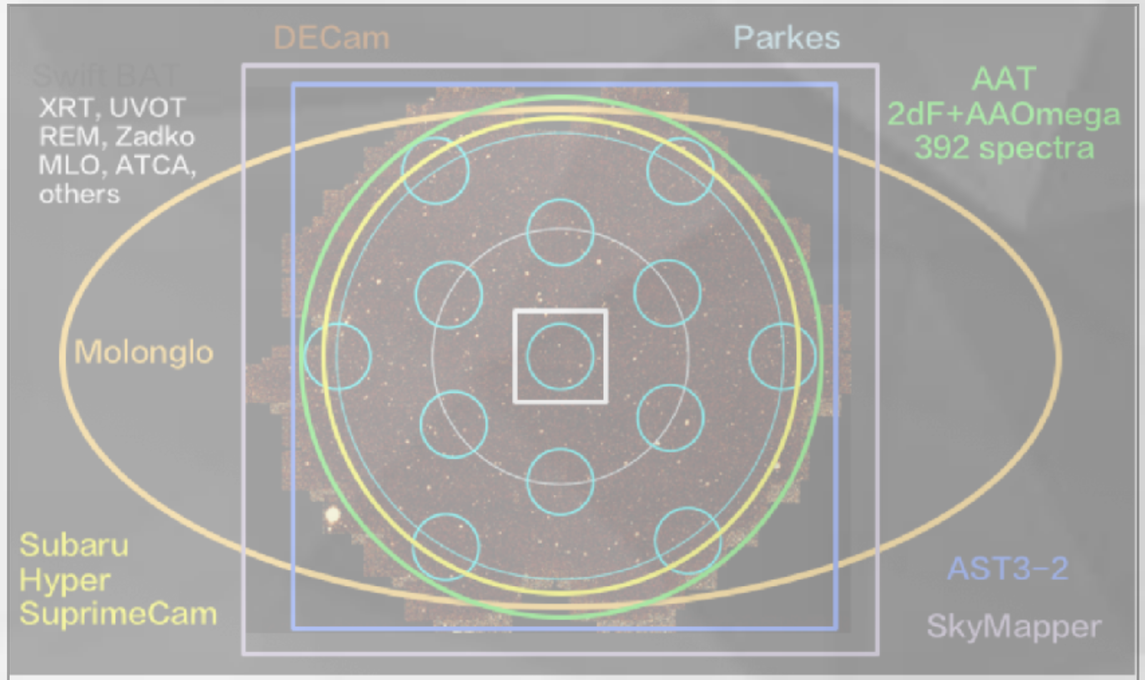
Apertif Cube Dimensions

0.25 TB / cube (Verheijen et al. 2009)

20,000 spectral-cubes

each containing ~100 sources

(Punzo et al. 2015)

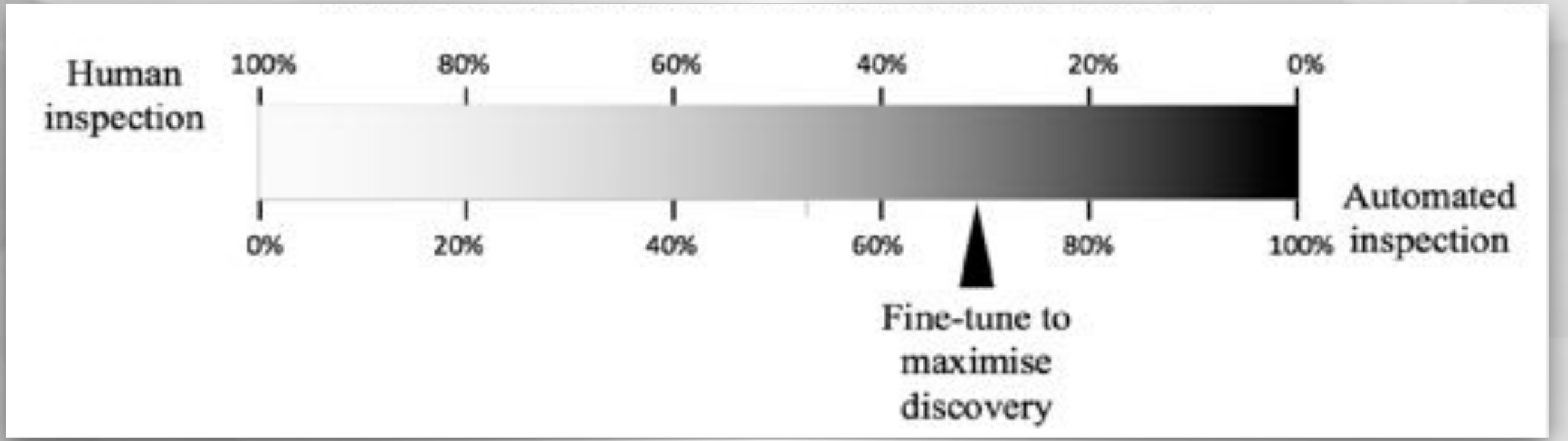


Andreoni & Cooke (2018)

Time-critical remote surveys

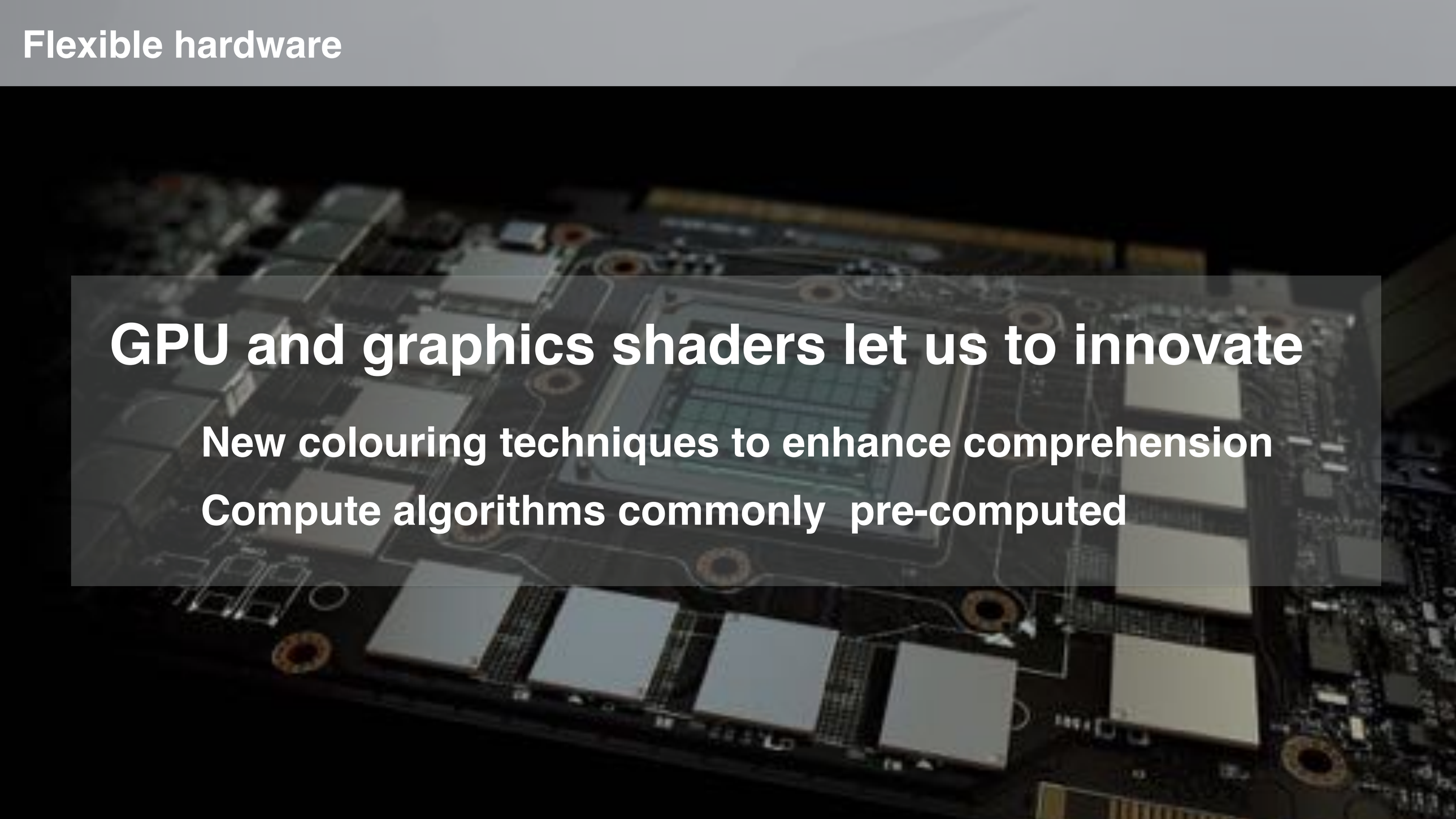
(Cooke et al. In prep; Andreoni et al. 2017; Meade et al. 2017; Vohl et al. 2017a)

Data-driven discovery



Fluke, Parrington, Hegarty et al. (2017)

Flexible hardware

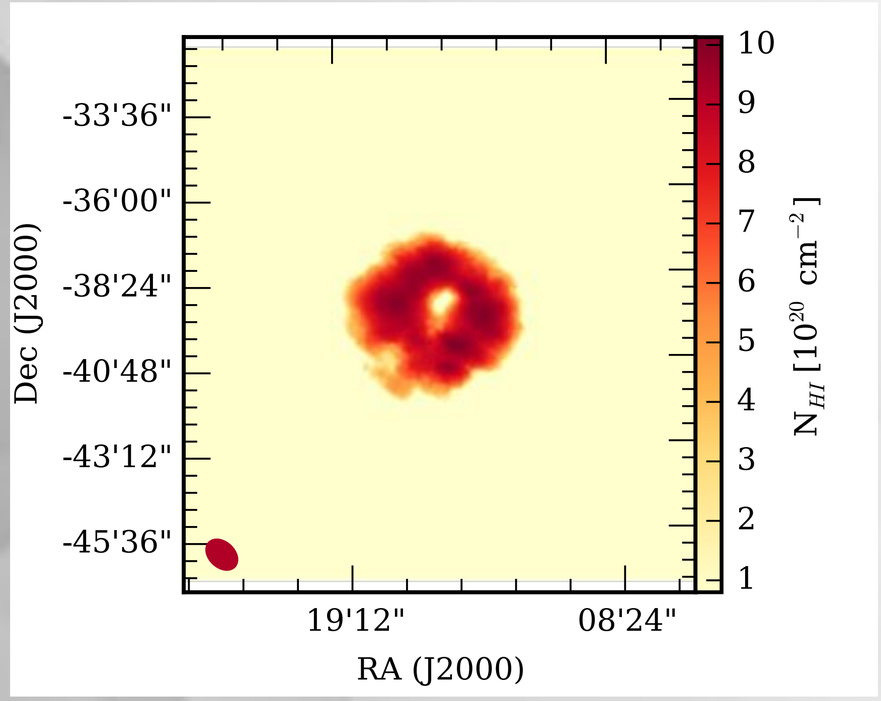


GPU and graphics shaders let us to innovate

New colouring techniques to enhance comprehension

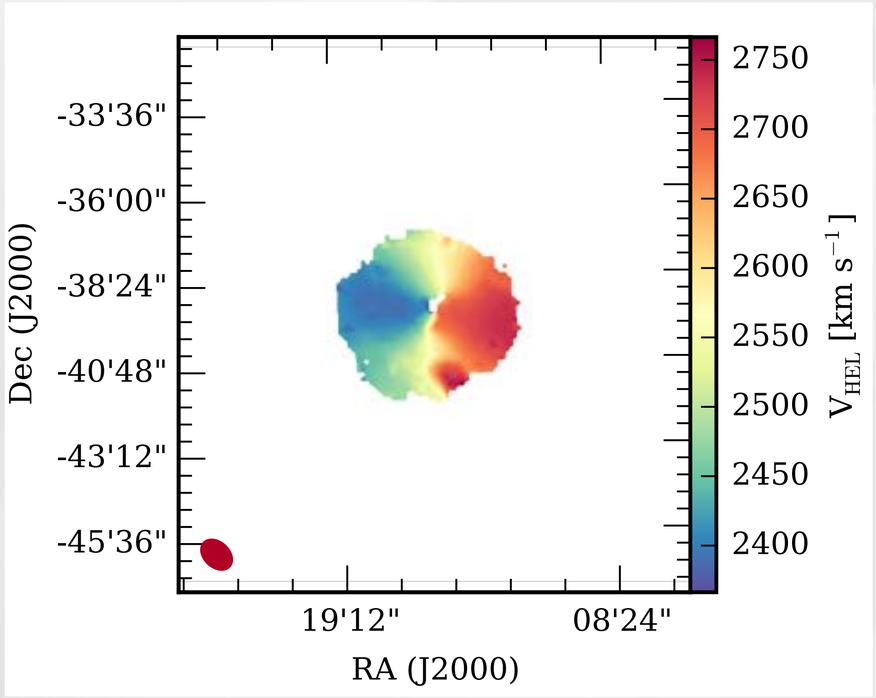
Compute algorithms commonly pre-computed

Moment maps



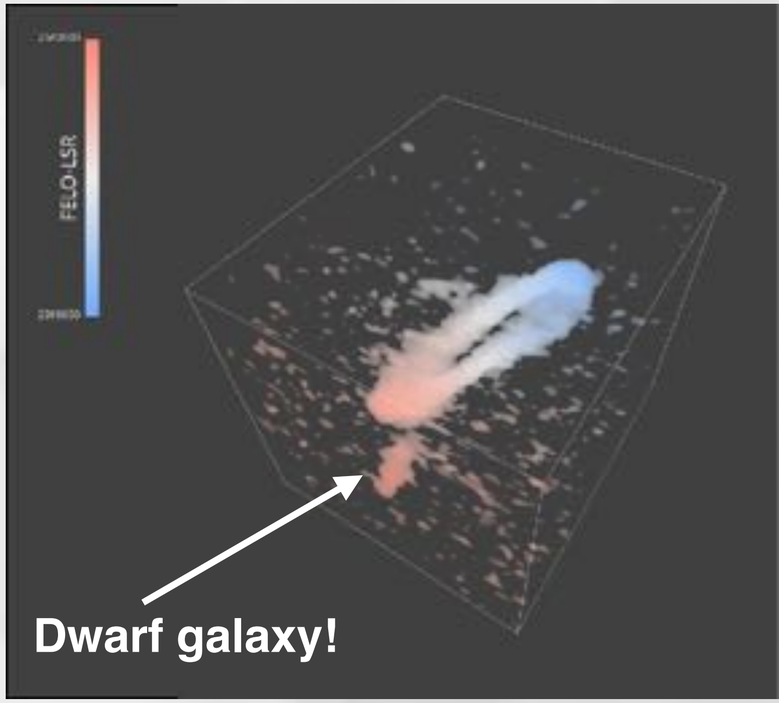
0th moment
overall gas distribution

$$M_0 = \Delta v \sum A(v)$$



1st moment
gas velocity field

$$M_1 = \frac{\sum vA(v)}{\sum A(v)}$$



Vohl et al. (2017a)

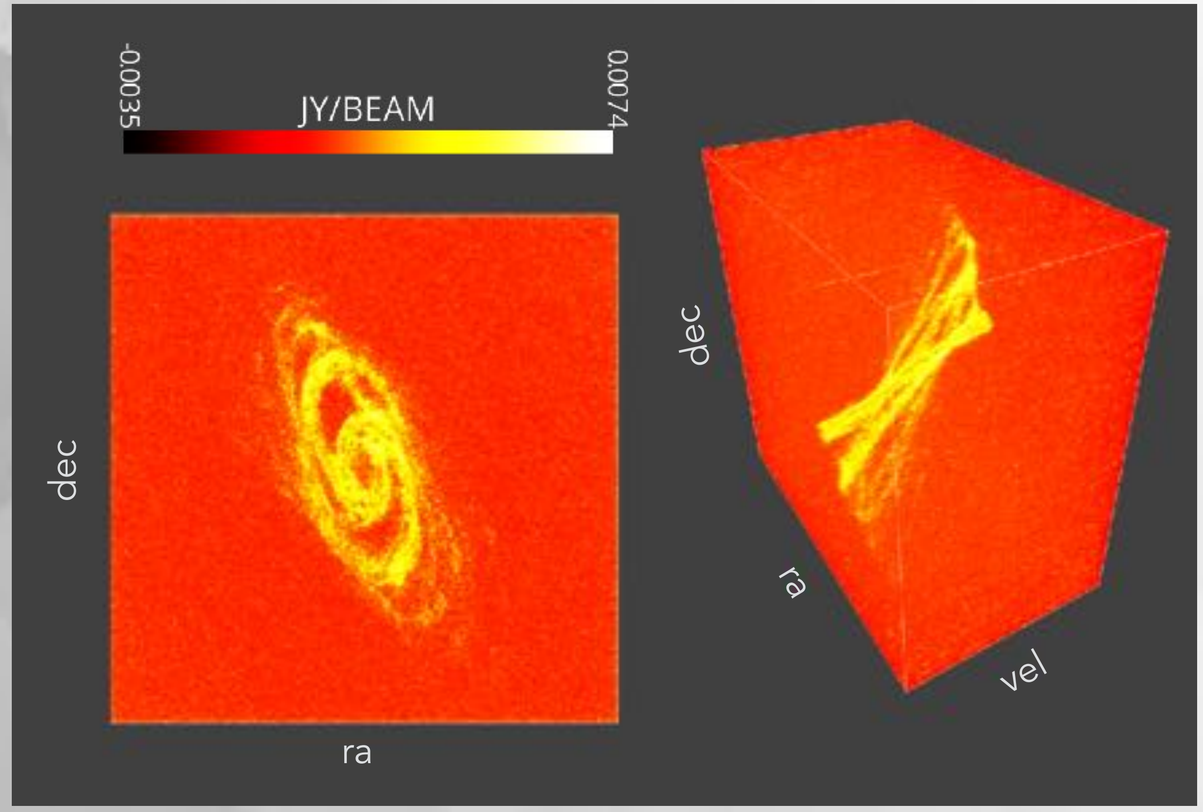
NGC 3261

Data : Katharina Lutz

Interactive & meaningful visual exploration

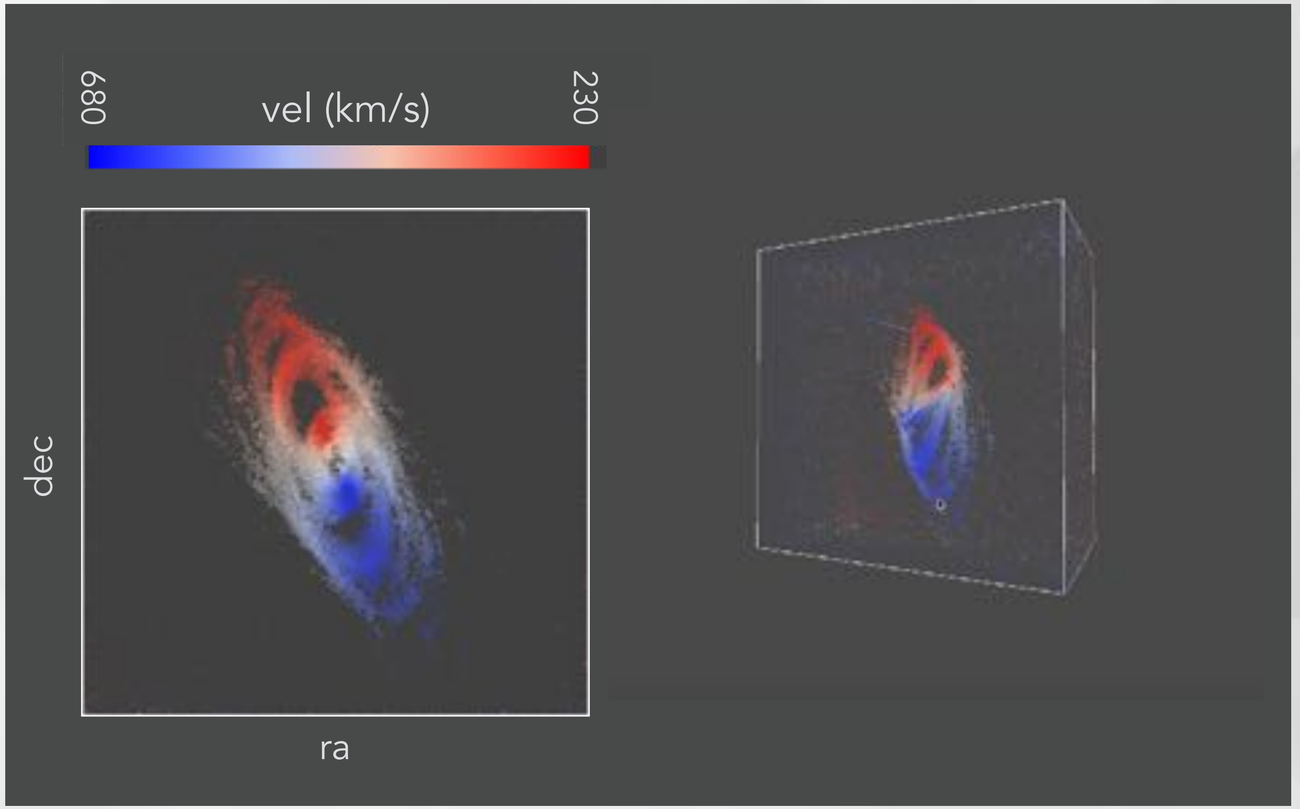
Domain knowledge to colour

0th moment-inspired



overall gas distribution

1st moment-inspired



gas velocity field

Vohl et al. (2017b)

Display ecology and **visual analytics**

Chung (2015)

Chung et al. (2014)

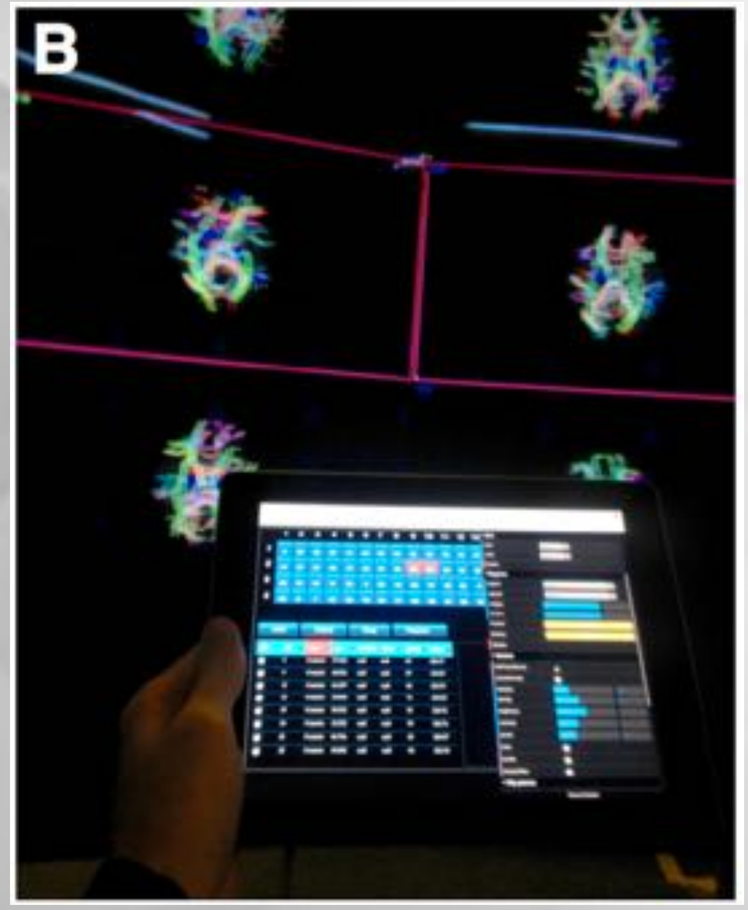
Different tasks require different tools or environment

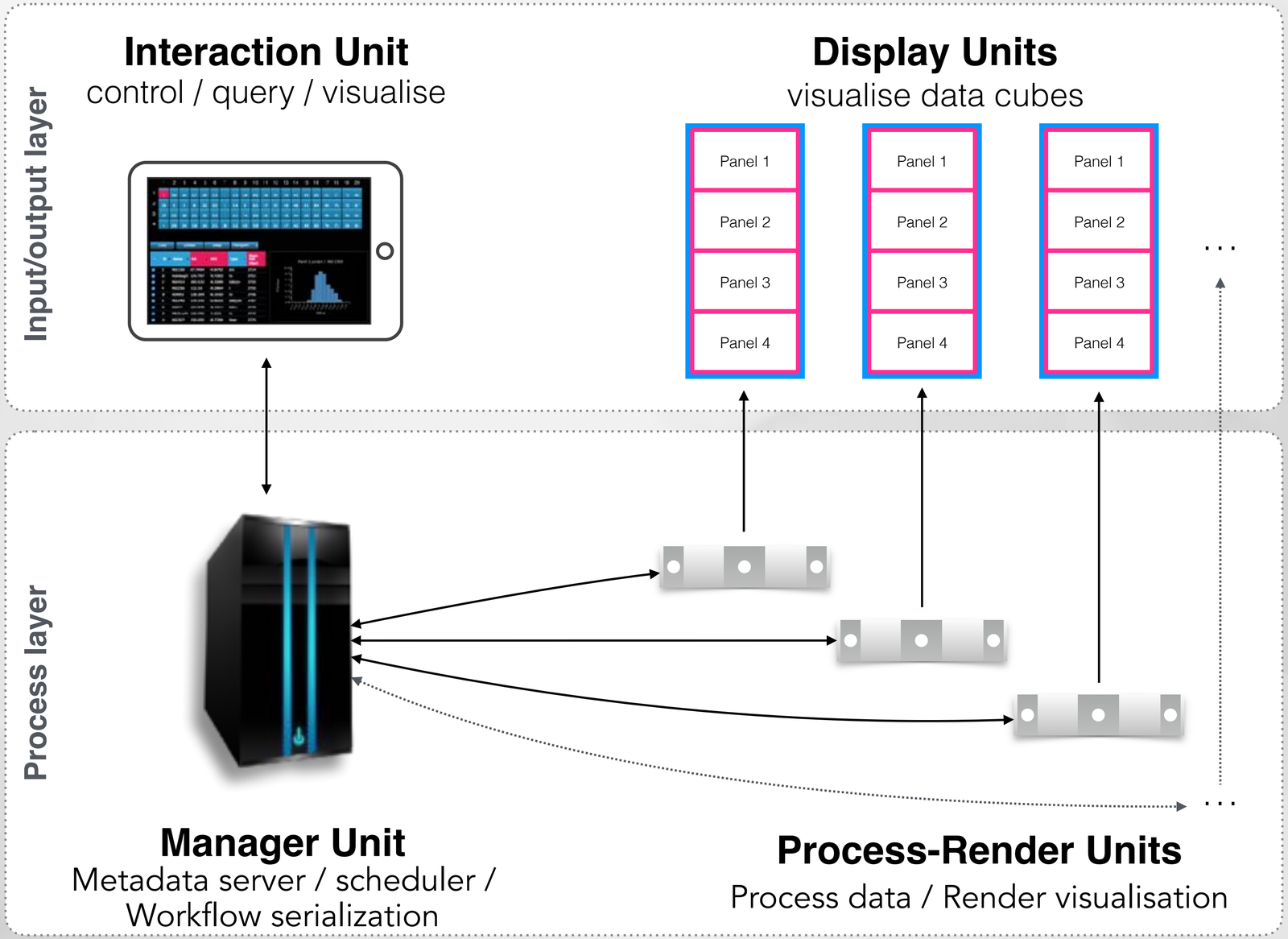
Immersive tiled-wall displays meets supercomputer

Image: D. Barnes (Monash)

Vohl et al. (2016)

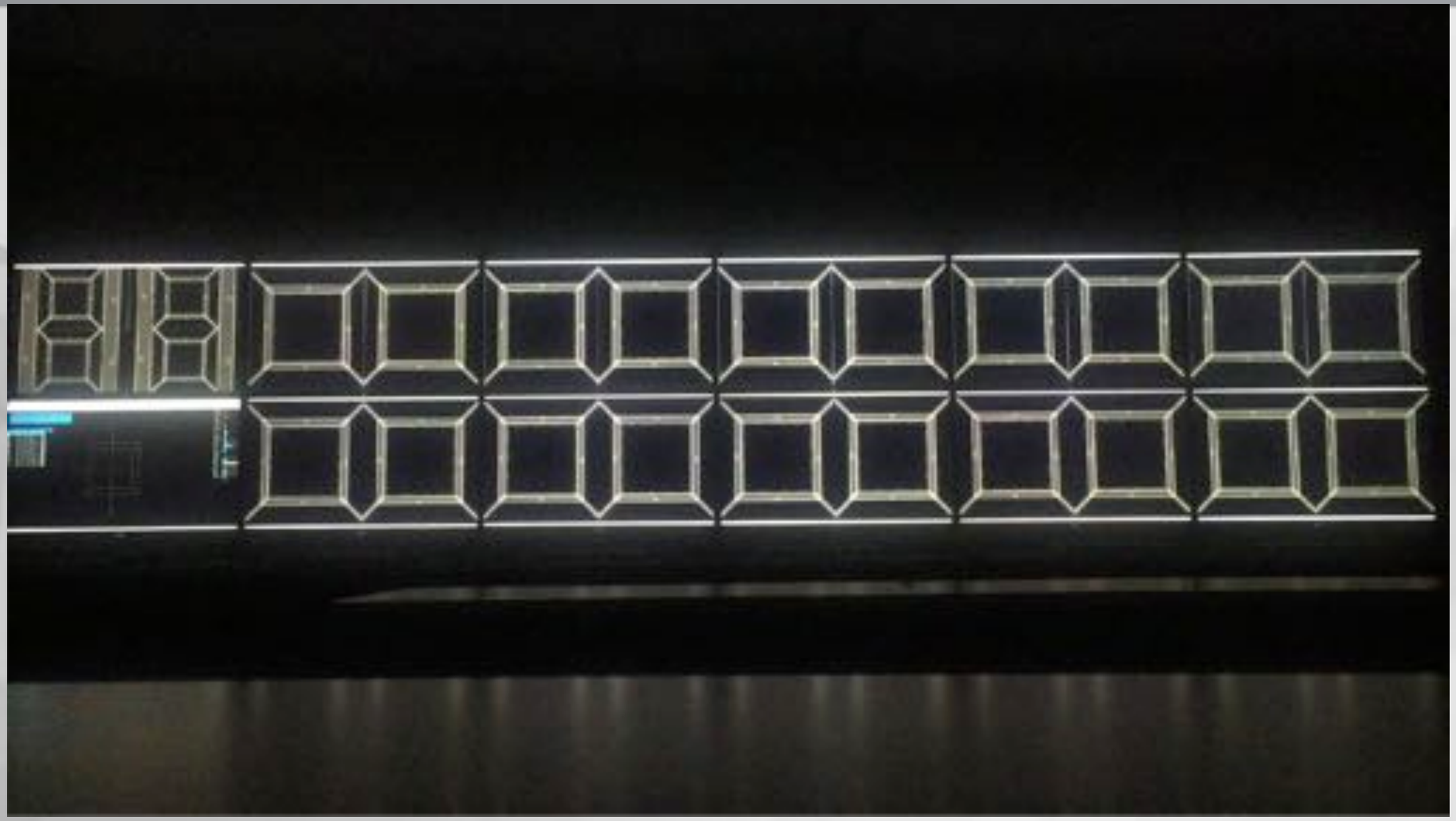
Load, Select, Sort, Query





Vohl et al. (2016)

encube: **Single instruction** Multiple visualisation / Multiple queries



Dany Vohl | AI 2018

https://youtu.be/OC2mW1oS94M

Vohl, Fluke, et al. In prep.

Fundamental problems of Petascale Astronomy Era

Storage, Transfer, Visualisation, & Analysis

Spectral Cube Surveys

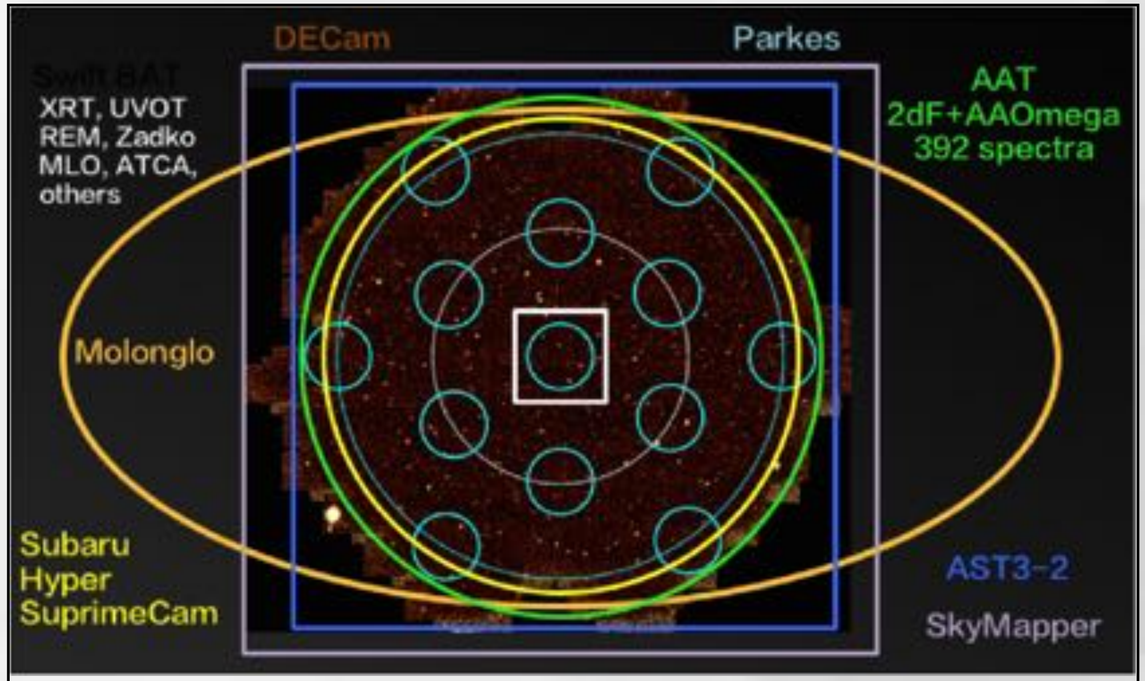
Apertif Cube Dimensions

0.25 TB / cube (Verheijen et al. 2009)

20,000 spectral-cubes

each containing ~100 sources

(Punzo et al. 2015)

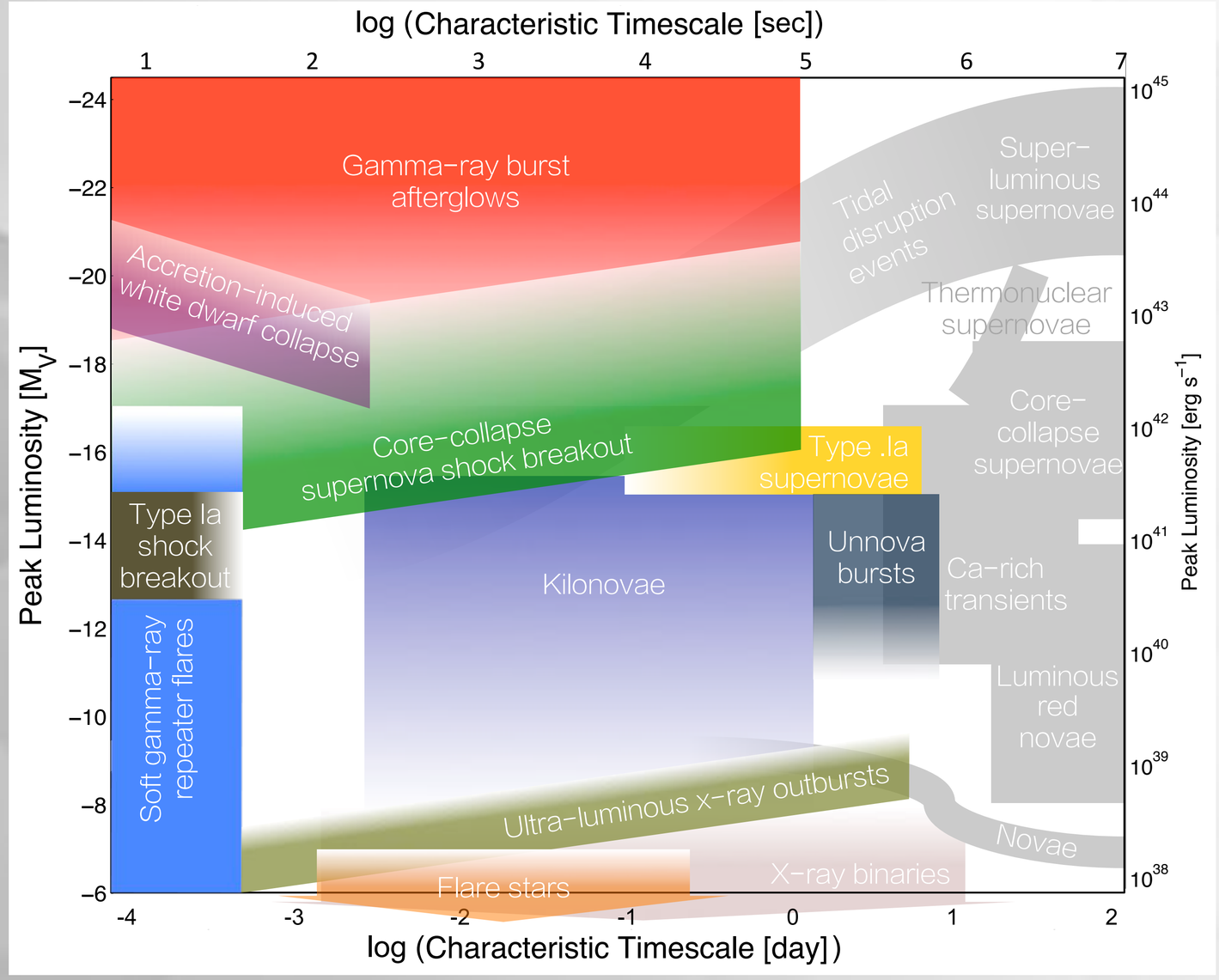


Andreoni & Cooke (2018)

Time-critical remote surveys

(Cooke et al. In prep; Andreoni et al. 2017; Meade et al. 2017; Vohl et al. 2017a)

Velocity Time-Critical Geographically-Dispersed Surveys



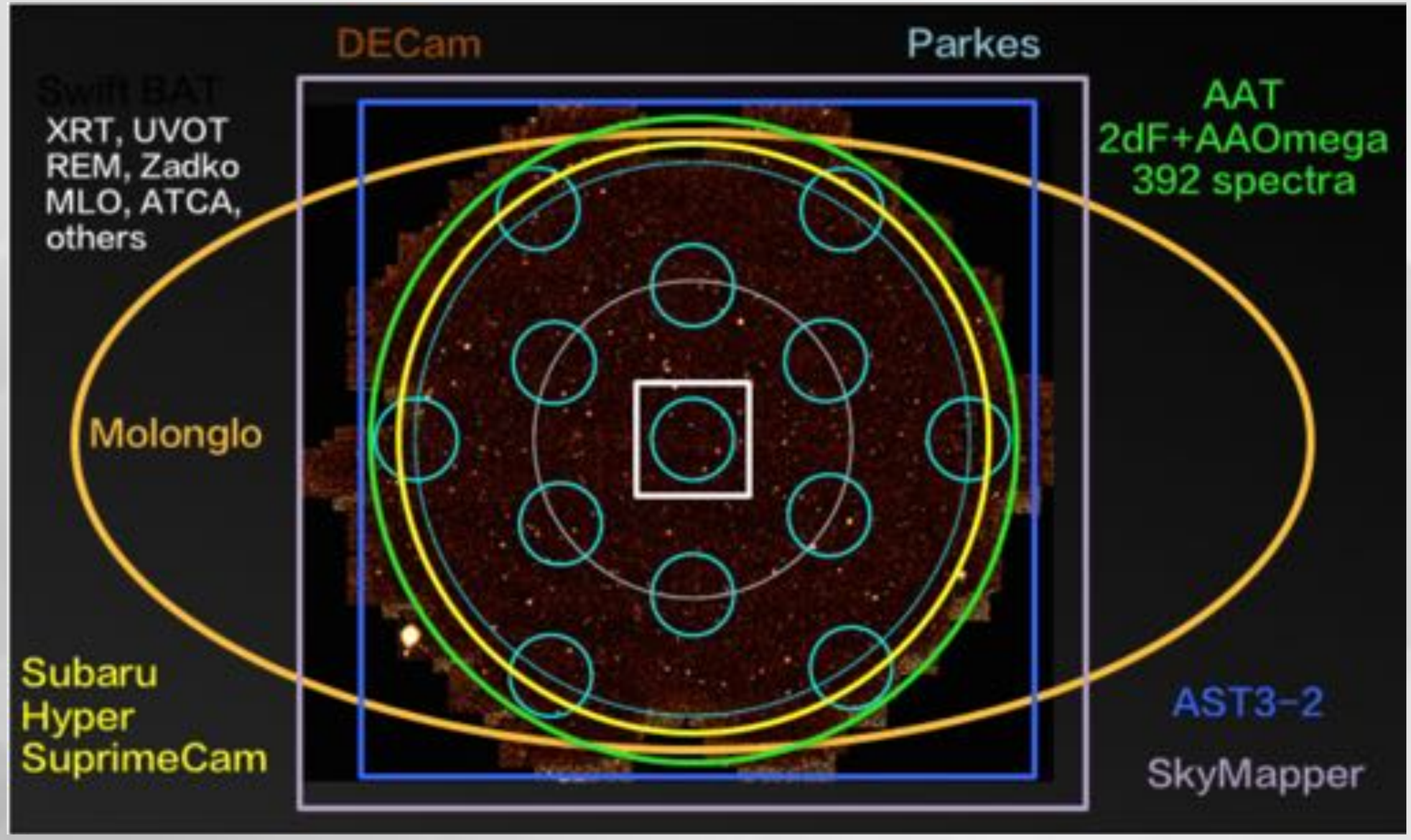
Cooke et al. In prep.

- **Deeper Wider Faster**
- PI: Jeff Cooke (Swinburne)
- Looking for the fastest transients
- Now 17 different telescopes synchronized
- And 25 telescopes ready for follow-up observations
- Heaps of people involved!

Velocity Time-Critical Geographically-Dispersed Surveys

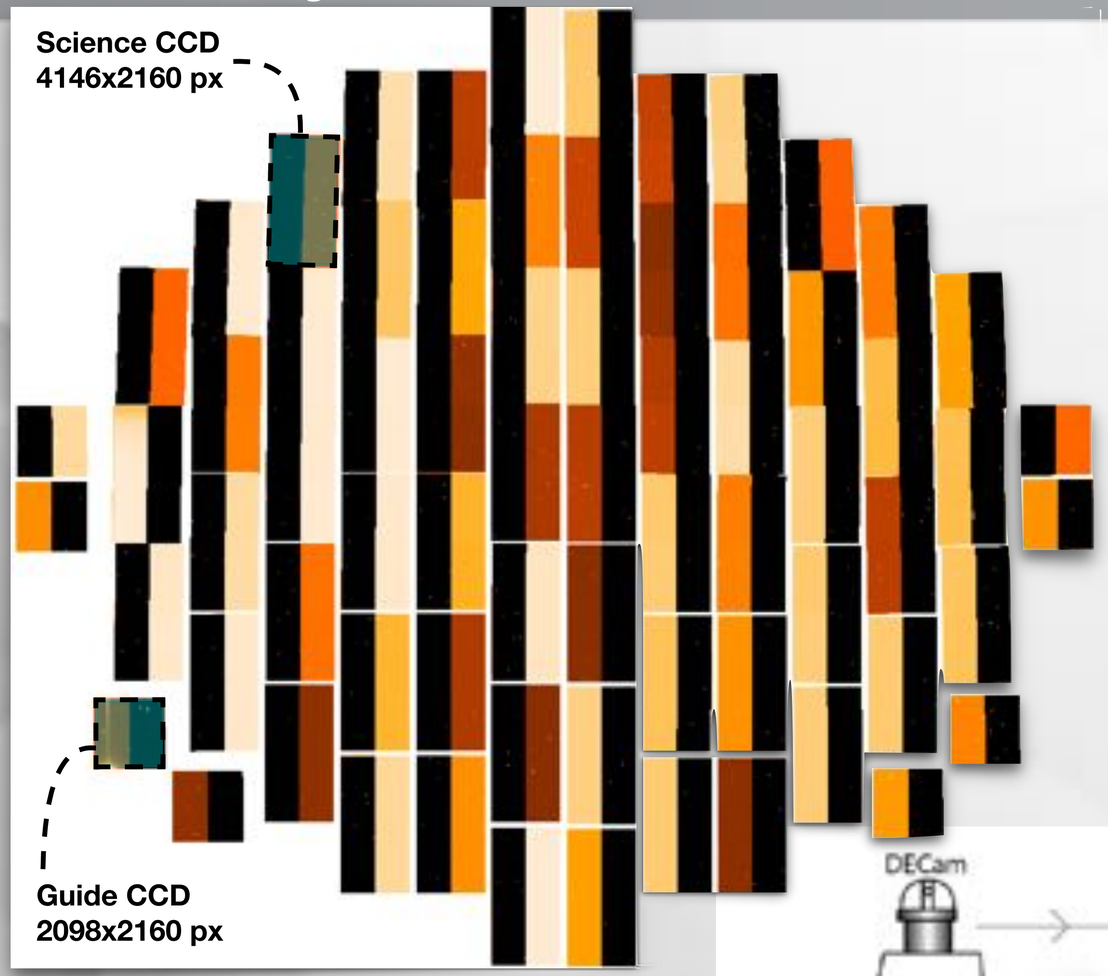
- **Deeper Wider Faster**
- PI: Jeff Cooke (Swinburne)
- Looking for the fastest transients
- Now 17 different telescopes synchronized
- And 25 telescopes ready for follow-up observations
- Heaps of people involved!

Velocity Time-Critical Geographically-Dispersed Surveys



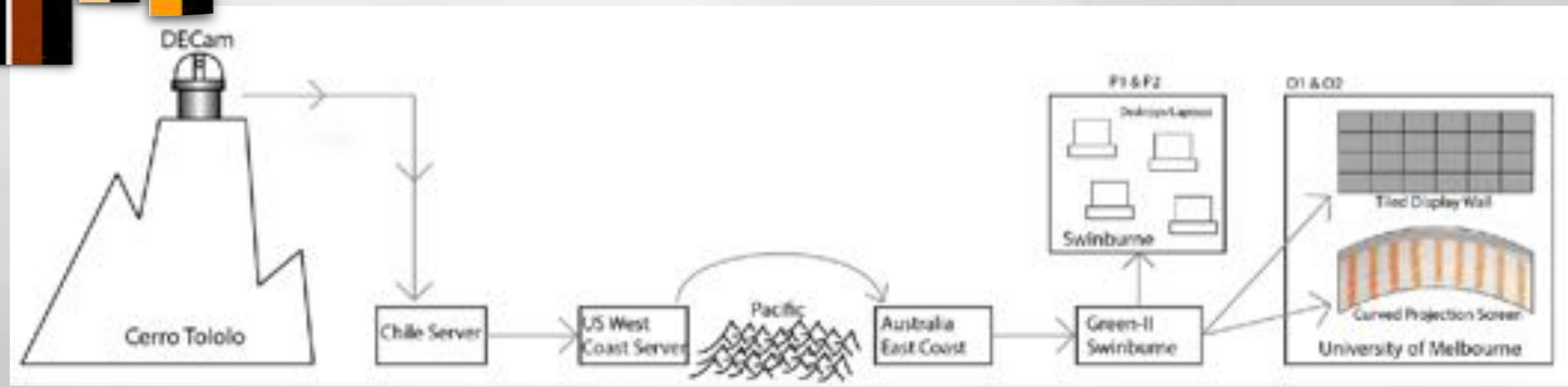
Andreoni & Cooke (2018)

Velocity Time-Critical Geographically-Dispersed Surveys



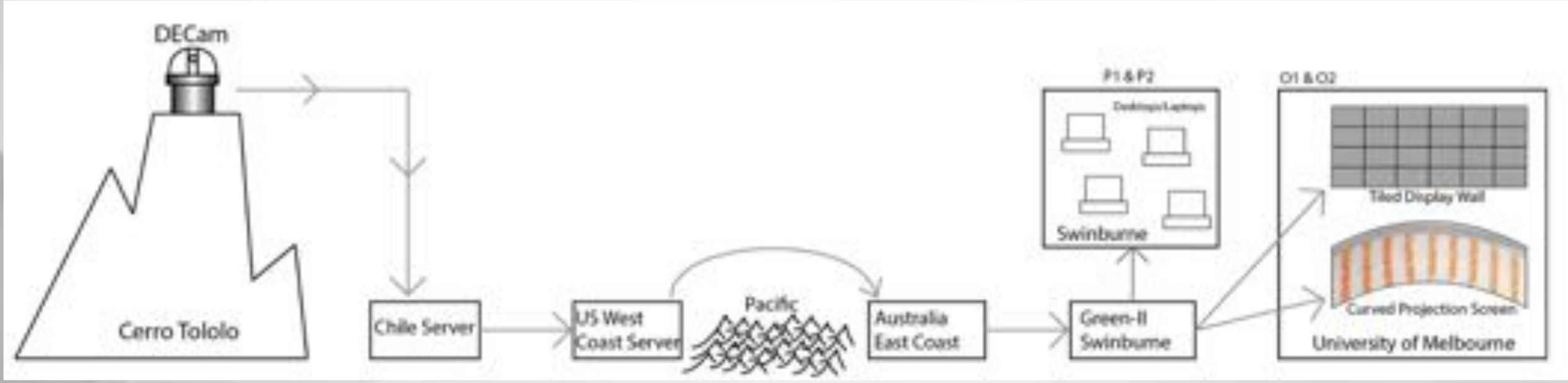
Vohl et al. (2017a)

- **DECam**
  - ~60 Science CCDs
- Image every 40s
  - Weeks at a time
- **~1.2GB / image**
- Raw image: **~17 minutes / file**



Meade et al. (2017), Andreoni et al. (2017)

Velocity Time-Critical Geographically-Dispersed Surveys



Lossy data compression

Convert FITS to JPEG2000

Reduced transfer time to seconds (enabling realtime work)

Vohl et al. (2017b)

Candidate identification

Mary pipeline, Machine learning

Andreoni et al. (2017)

Collaborative candidate confirmation

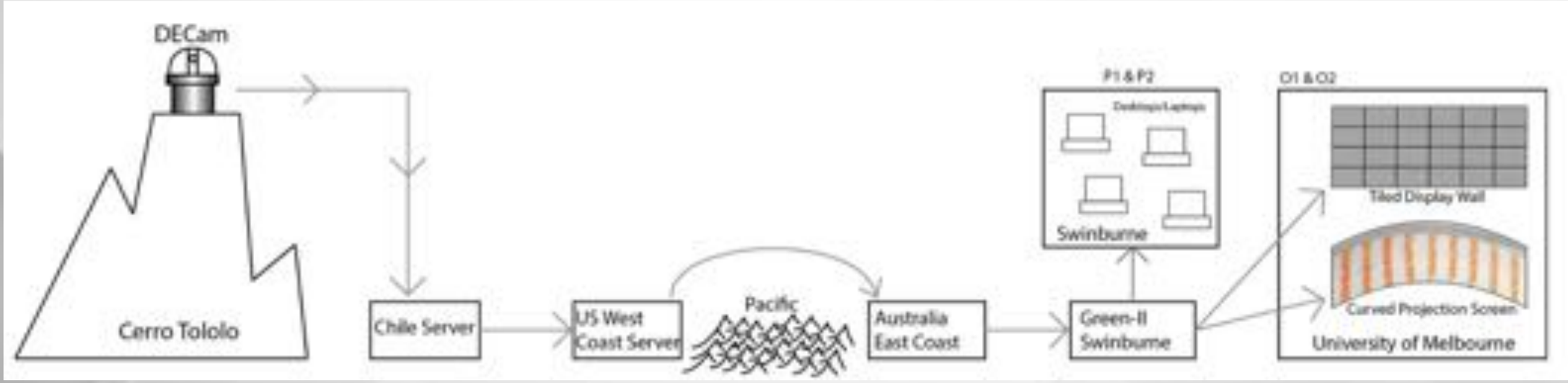
On site: Heterogeneous display ecology

Meade et al. (2017)

On site or remotely: PerSieve (web portal)

Hegarty et al. (In prep.)

Velocity Time-Critical Geographically-Dispersed Surveys



Lossy data compression

Convert FITS to JPEG2000

Reduced transfer time to seconds (enabling realtime work)

Vohl et al. (2017b)

Candidate identification

Mary pipeline + Machine learning (to reject artefacts)

Andreoni et al. (2017)

Collaborative candidate confirmation

On site: Heterogeneous display ecology

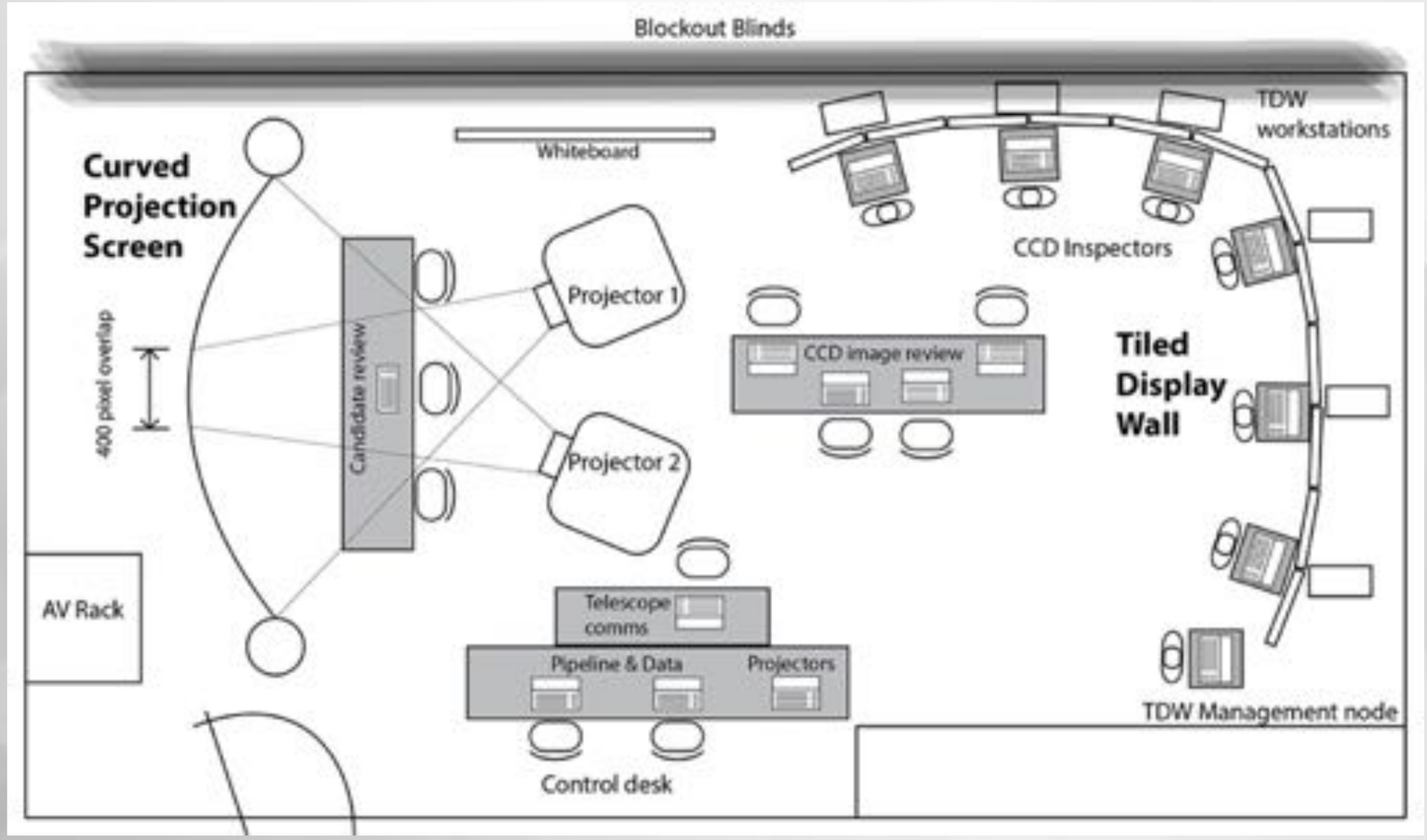
Meade et al. (2017)

On site or remotely: PerSieve (web portal)

Hegarty et al. (In prep.)

Collaborative workspace

Display ecology, visual analytics



Meade et al. (2017)

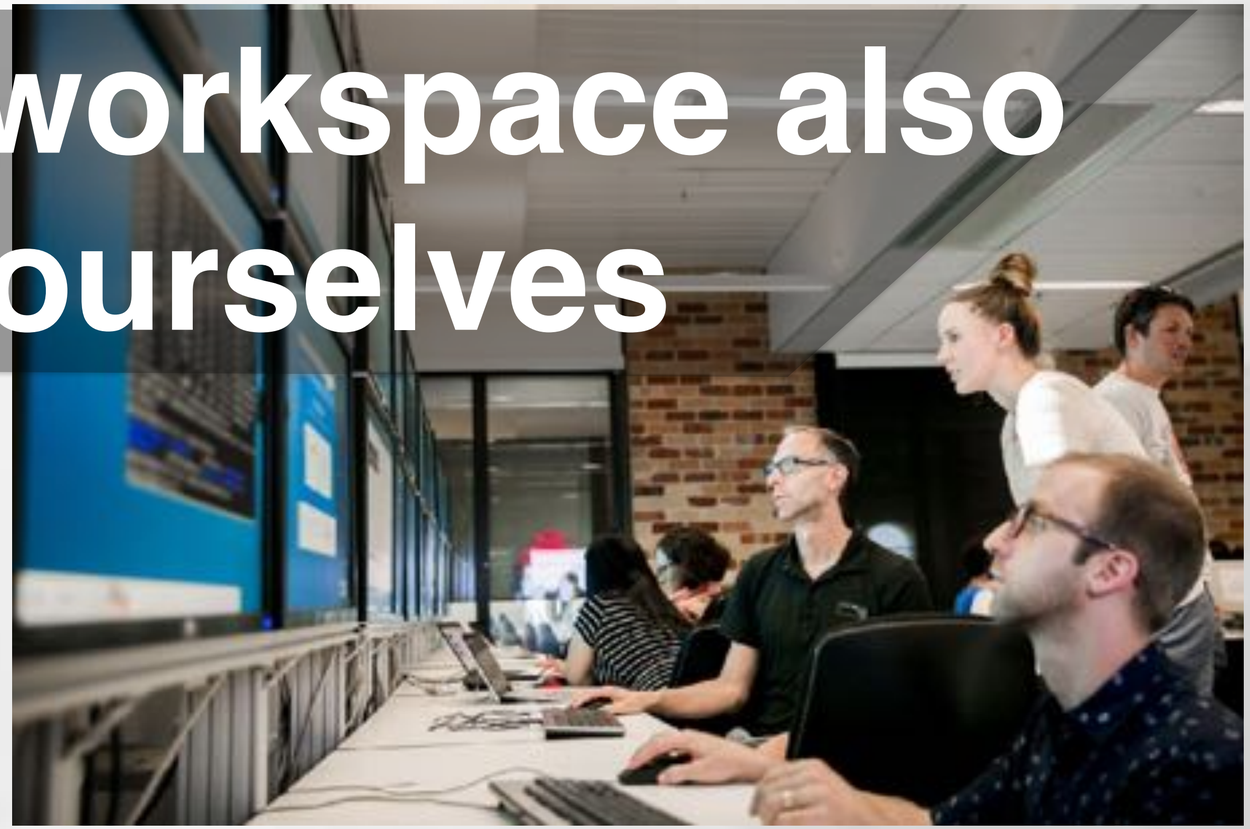
Remote Telescope Operations Centre, Swinburne University of Technology

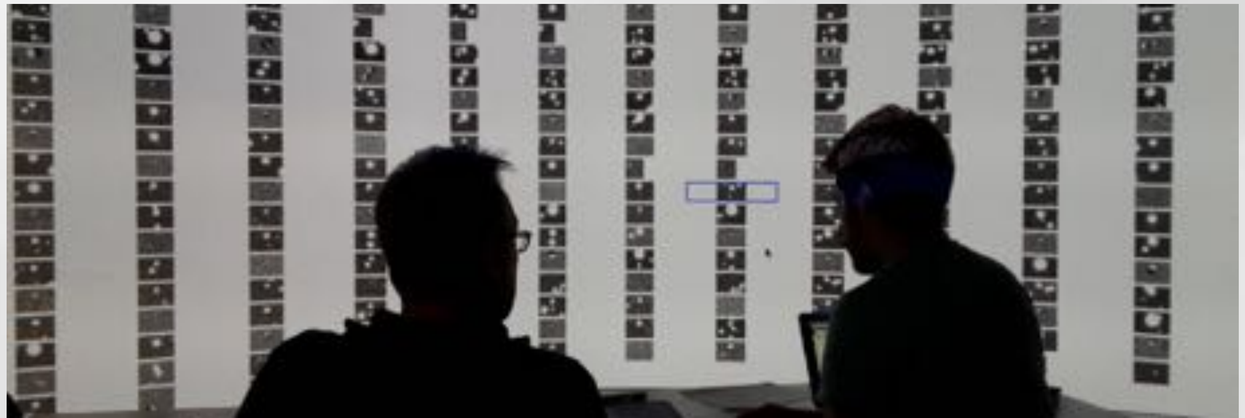
On site or remotely: PerSieve (web portal)

Hegarty et al. (In prep.)

How do we do?

A collaborative workspace also allows to study ourselves

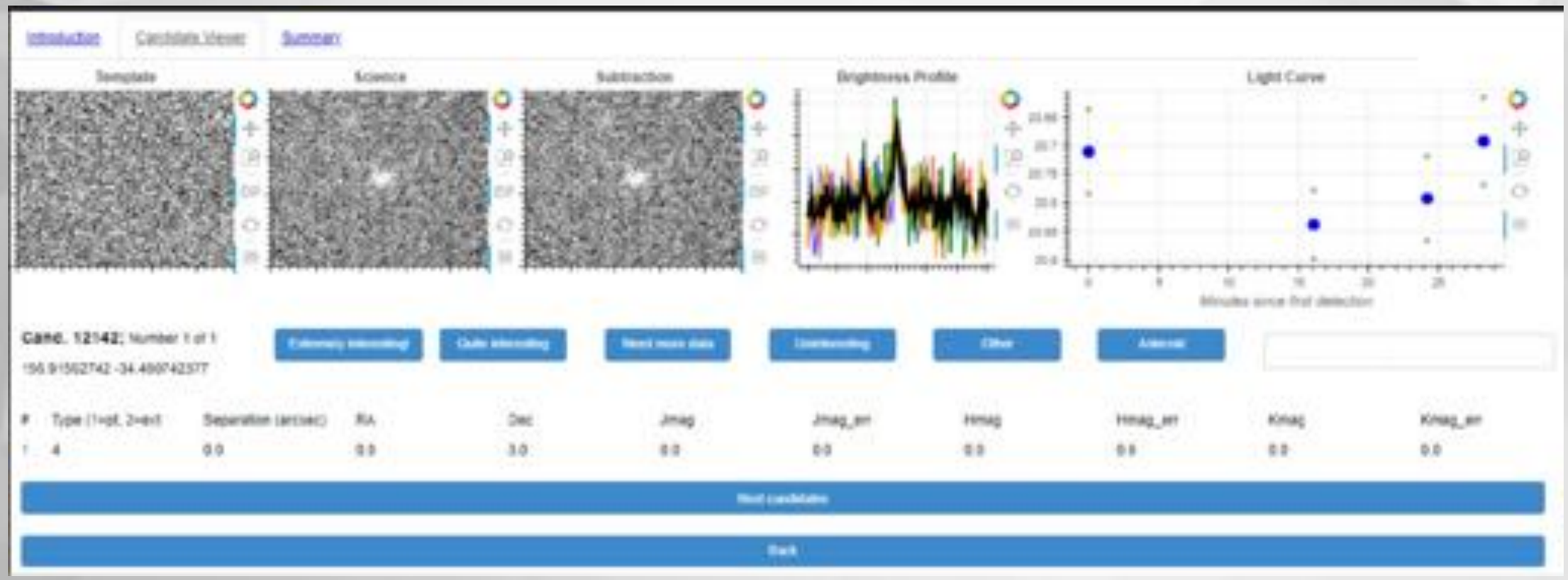




How do we do?

PerSieve

- **Task?** Rate events between 0 (least interesting) to 5 (most interesting)
- **Inspectors?** Novice, Intermediate, Expert
- **Is the workflow effective?**

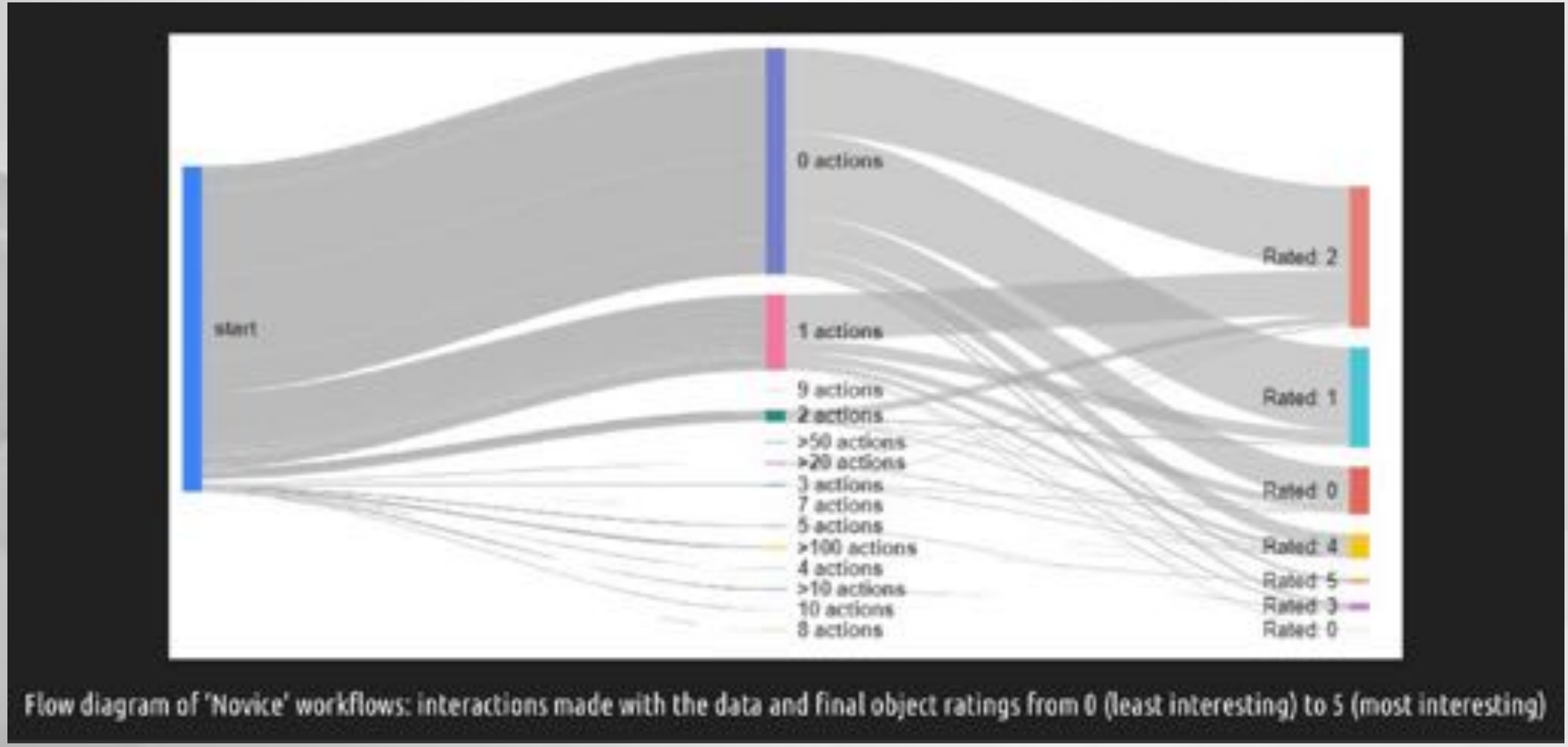


*Swinburne Ethics approval, with participants consent.

Ongoing work by Sarah Hegarty (Swinburne)

How do we do?

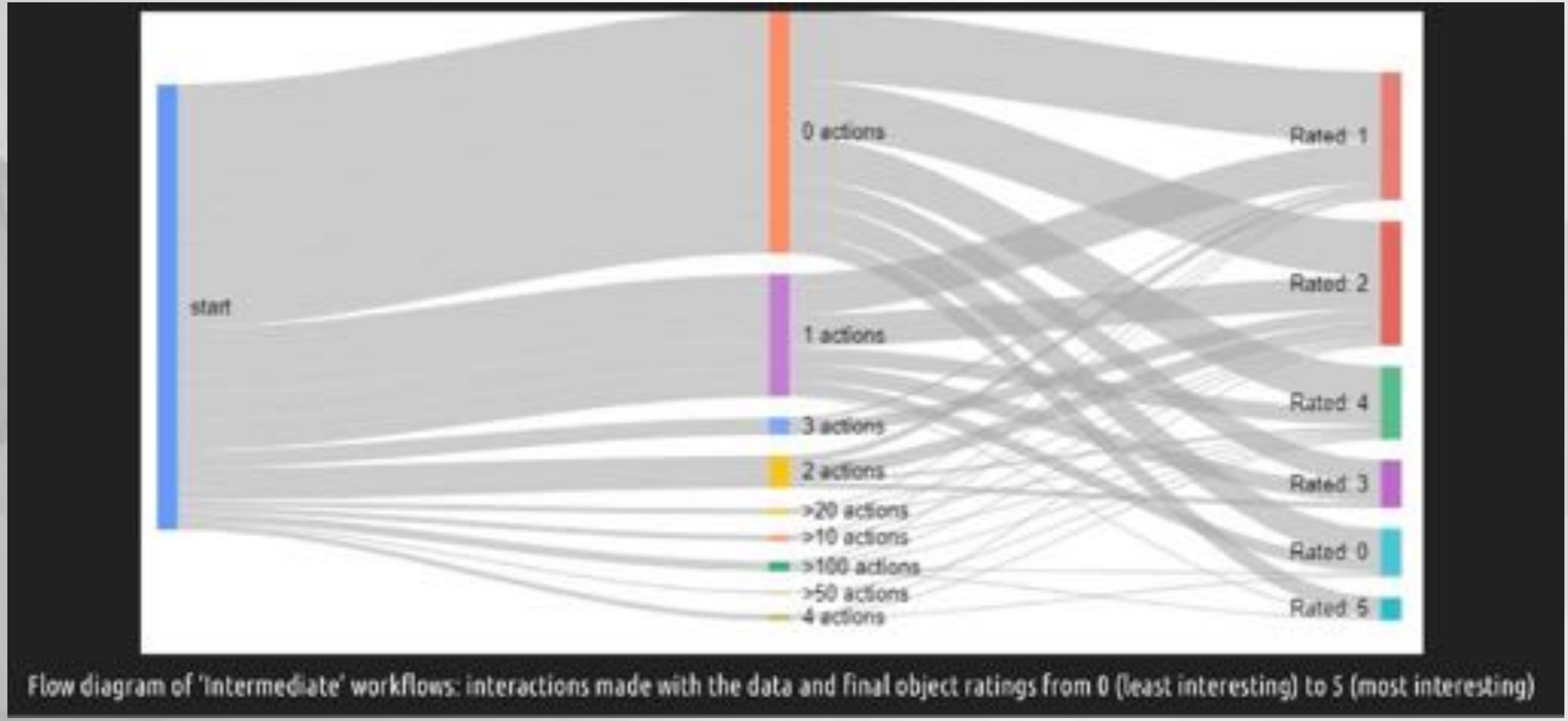
Novice inspectors



Flow diagram of 'Novice' workflows: interactions made with the data and final object ratings from 0 (least interesting) to 5 (most interesting)

Courtesy of Sarah Hegarty (Swinburne)

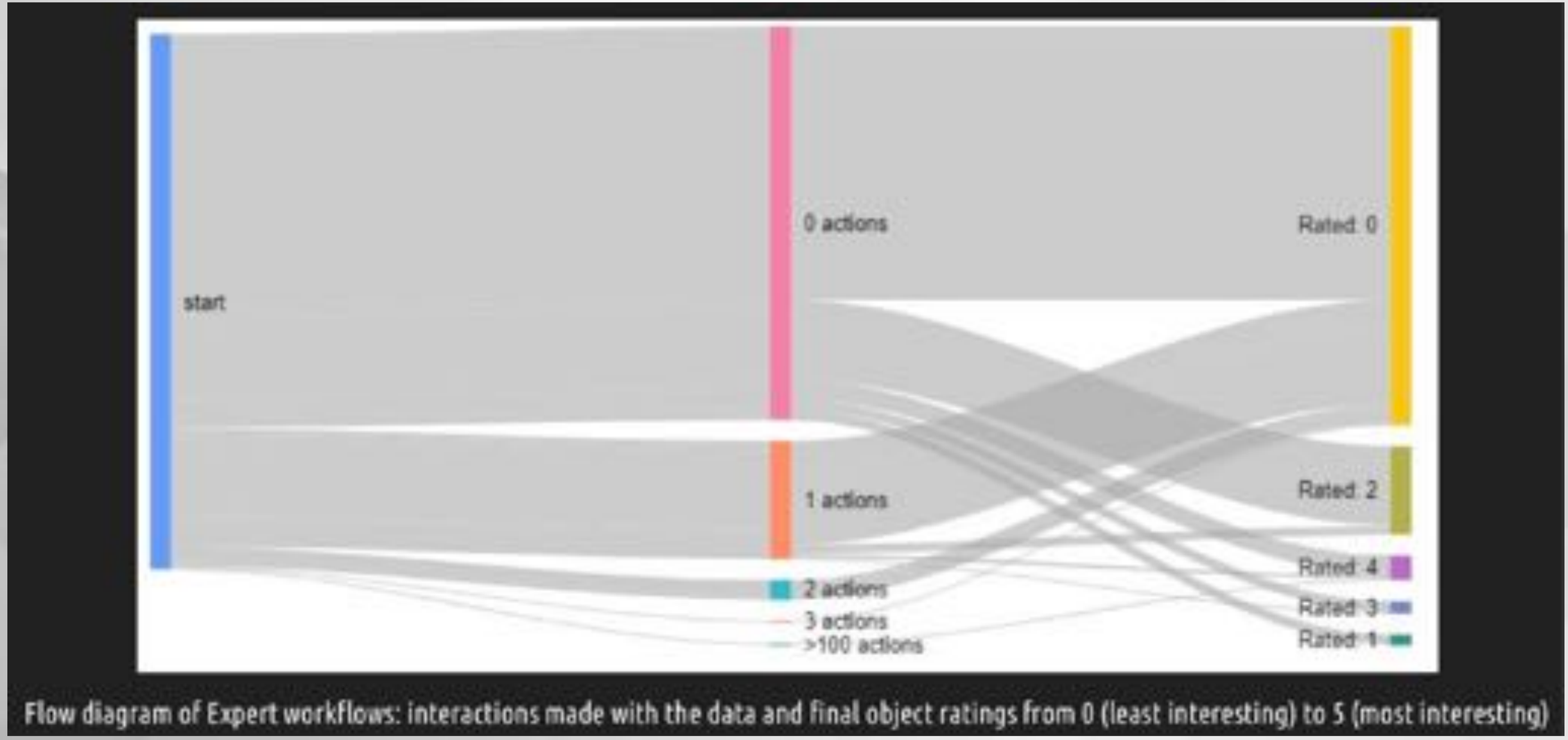
How do we do? Intermediate inspectors



Flow diagram of 'intermediate' workflows: interactions made with the data and final object ratings from 0 (least interesting) to 5 (most interesting)

Courtesy of Sarah Hegarty (Swinburne)

How do we do? Expert inspectors



Flow diagram of Expert workflows: interactions made with the data and final object ratings from 0 (least interesting) to 5 (most interesting)

Courtesy of Sarah Hegarty (Swinburne)

Effect on DWF science

- During the 2016 observing campaign alone
  - **1000s of rising transient events** provided targets for follow-up spectroscopy
  - 2 triggers were sent to Gemini; 5 to SALT, ~1000 to the AAT for transient or host galaxy spectra
- **ATel #10072**: Discovery of 4 possible supernovae
- **ATel #10078**: Confirmation with the SALT telescope of a young Type Ia supernova at z=0.046

Final thoughts

- Working with Big Data is non-trivial
- Still, with a few tricks and human-centered methodology, we showed that we can bypass some of these issues
  - E.g. Simple compression enabled real-time team work
- Use hardware/display at the right time / for the right task

Acknowledgements

netherlands eScience center

Netherlands eScience Center grant AA-ALERT (027.015.G09)

References

- Andreoni, I., Jacobs, C., Hegarty, S., Pritchard, T., Cooke, J., & Ryder, S. (2017). Mary, a pipeline to aid discovery of optical transients. *Publications of the Astronomical Society of Australia, 34*.
- Andreoni, I., & Cooke, J. (2018). The Deeper Wider Faster program: chasing the fastest bursts in the Universe. *arXiv preprint* arXiv:1802.01100.
- Chung, H., North, C., Self, J. Z., Chu, S. , and Quek, F. (2014). "VisPorter: facilitating information sharing for collaborative sensemaking on multiple displays," *Personal and Ubiquitous Computing*, vol. 18, pp. 1169-1186.
- Chung, H. (2015). "Designing Display Ecologies for Visual Analysis," Ph.D. Dissertation, Dept. of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA.
- Fluke, C. J., Parrington, L., Hegarty, S., MacMahon, C., Morgan, S., Hassan, A. H., & Kilborn, V. A. (2017). Sports Stars: Analyzing the Performance of Astronomers at Visualization-based Discovery. *Publications of the Astronomical Society of the Pacific, 129*(975), 058009.
- Meade, B., Fluke, C., Cooke, J., Andreoni, I., Pritchard, T., Curtin, C., ... & Vohl, D. (2017). Collaborative Workspaces to Accelerate Discovery. *Publications of the Astronomical Society of Australia, 34*.
- Punzo, D., Van der Hulst, J. M., Roerdink, J. B. T. M., Oosterloo, T. A., Ramatsoku, M., & Verheijen, M. A. W. (2015). The role of 3-D interactive visualization in blind surveys of H i in galaxies. *Astronomy and Computing, 12*, 86-99.
- Verheijen, M. A. W., Oosterloo, T., Heald, G., & van Cappellen, W. (2009). *Proc. Sci.*, HI Surveys with APERTIF.
- Vohl, D., Barnes, D. G., Fluke, C. J., Poudel, G., Georgiou-Karistianis, N., Hassan, A. H., ... & Bonnington, C. P. (2016). Large-scale comparative visualisation of sets of multidimensional data. *PeerJ Computer Science, 2*, e88.
- Vohl, D., Pritchard, T., Andreoni, I., Cooke, J., & Meade, B. (2017). Enabling near real-time remote search for fast transient events with lossy data compression. *Publications of the Astronomical Society of Australia, 34*.
- Vohl, D., Fluke, C. J., Barnes, D. G., & Hassan, A. H. (2017). Real-time colouring and filtering with graphics shaders. *Monthly Notices of the Royal Astronomical Society, 471*(3), 3323-3346.