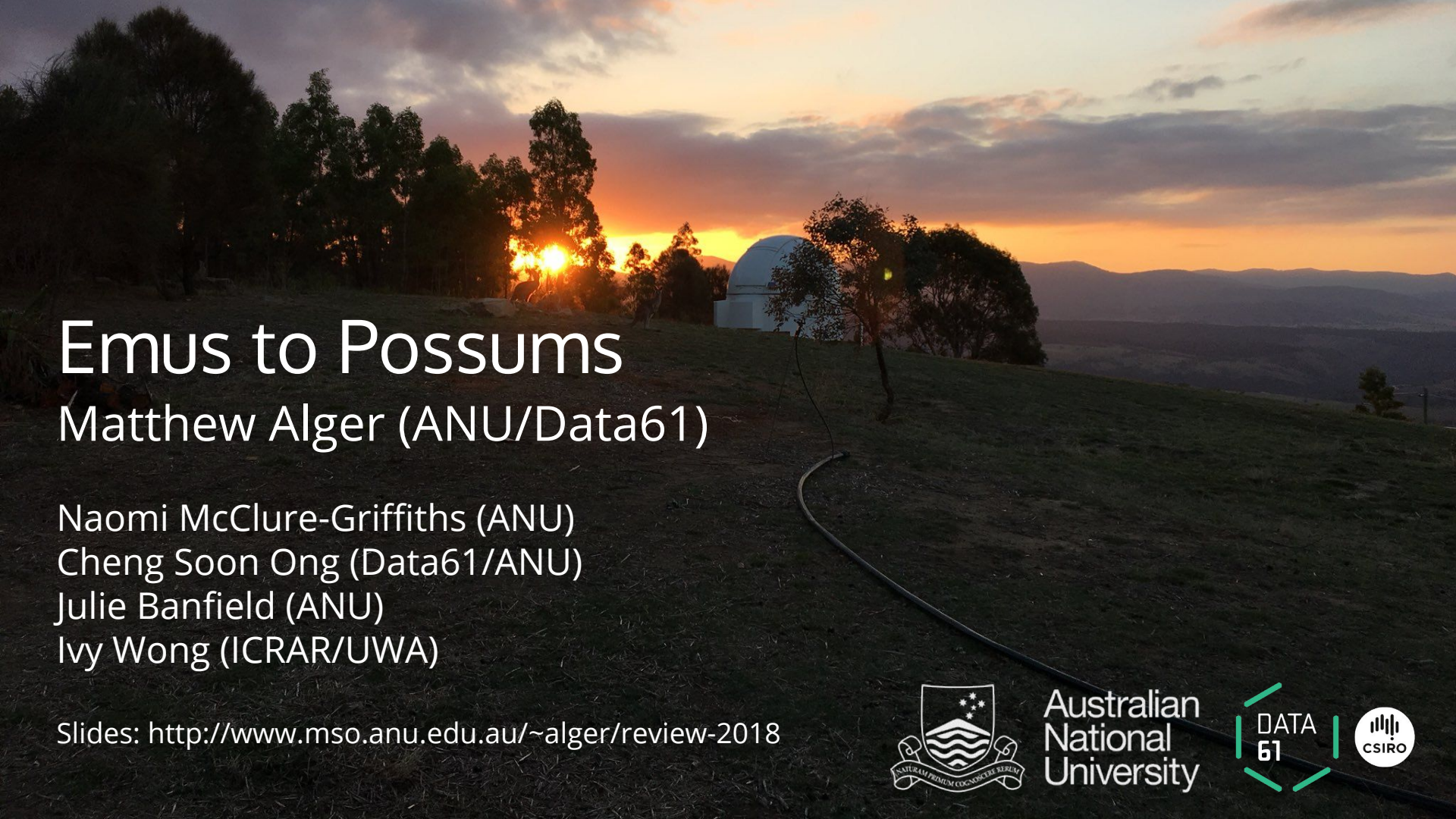


Emus to Possums

Matthew Alger (ANU/Data61)

Naomi McClure-Griffiths (ANU)
Cheng Soon Ong (Data61/ANU)
Julie Banfield (ANU)
Ivy Wong (ICRAR/UWA)

Slides: http://www.mso.anu.edu.au/~alger/review-2018

Australian National University

DATA 61

CSIRO

Matthew Alger
@IndecisiveMatt

nobody ever told me that big data would be so large

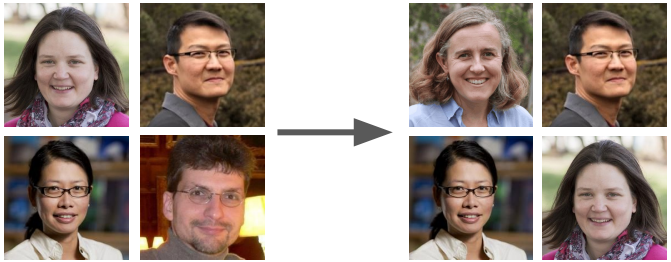
3:48 PM - 26 Sep 2017

4 Likes

1 4

The Year in Review

Administration



POSSUM
POLARISATION SKY SURVEY OF THE UNIVERSE'S MAGNETISM

DATA 61

CSIRO

$$\operatorname{sign}\left(\sigma\left(\sum_{i=1}^{n}\alpha_i k(x_i,x)\right)\right)$$

POSSUM
POLARISATION SKY SURVEY OF THE UNIVERSE'S MAGNETISM

Papers authored

Radio Galaxy Zoo: Machine learning for radio source host galaxy cross-identification

M. J. Alger[1,2⋆], J. K. Banfield[1,3], C. S. Ong[2,4], L. Rudnick[5], O. I. Wong[6,3], C. Wolf[1,3], H. Andernach[7], R. P. Norris[8,9], S. S. Shabala[10]

[1]*Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 2611, Australia*
[2]*Data61, CSIRO, Canberra, ACT 2601, Australia*
[3]*ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO)*
[4]*Research School of Computer Science, The Australian National University, Canberra, ACT 2601, Australia*
[5]*Minnesota Institute for Astrophysics, University of Minnesota, 116 Church St. SE, Minneapolis, MN 55455*
[6]*International Centre for Radio Astronomy Research-M468, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia*
[7]*Departamento de Astronomía, DCNE, Universidad de Guanajuato, Apdo. Postal 144, CP 36000, Guanajuato, Gto., Mexico*
[8]*Western Sydney University, Locked Bag 1797, Penrith South, NSW 1797, Australia*
[9]*CSIRO Astronomy & Space Science, PO Box 76, Epping, NSW 1710, Australia*
[10]*School of Natural Sciences, University of Tasmania, Private Bag 37, Hobart, Tasmania 7001, Australia*

Radio Galaxy Zoo: ClaRAN — a deep learning classifier for radio morphologies

Chen Wu[1⋆], O. Ivy Wong[1†], Lawrence Rudnick[2], Stanislav S. Shabala[3]
Matthew J. Alger[4,5], Julie K. Banfield[4,6], Cheng Soon Ong[5], Sarah V. White[7]
Avery F. Garon[2], Ray P. Norris[8,9], Heinz Andernach[10], Jean Tate[11], Vesna Lukic[12]
Hongming Tang[13], Kevin Schawinski[14] and Foivos I. Diakogiannis[15,1]

Talks presented

Learning to Cross-identify Wide-area Radio Surveys with Radio Galaxy Zoo: Data challenges in citizen science

Matthew Alger (ANU/Data61)
Julie Banfield (ANU)
Cheng Soon Ong (Data61/ANU)

Slides: http://www.mso.anu.edu.au/~alger/c3dis

C3DIS

Finding Radio Host Galaxies with Machine Learning and Radio Galaxy Zoo

Matthew Alger (ANU),
Julie Banfield (ANU/WSU),
Cheng Soon Ong (Data61/ANU),
Ivy Wong (ICRAR/UWA)

Slides: http://www.mso.anu.edu.au/~alger/ml-projects-xid

WSU

Machine Learning Methods for Radio Host Cross-Identification with Crowdsourced Labels

Matthew Alger (ANU),
Julie Banfield (ANU/WSU),
Cheng Soon Ong (Data61/ANU),
Ivy Wong (ICRAR/UWA)

Slides: http://www.mso.anu.edu.au/~alger/sparcs-vii

SPARCS VII

Finding Radio Host Galaxies with Machine Learning and Radio Galaxy Zoo

Matthew Alger (ANU),
Julie Banfield (ANU/WSU),
Cheng Soon Ong (Data61/ANU),
Ivy Wong (ICRAR/UWA)

Slides: http://www.mso.anu.edu.au/~alger/icrar-amt

ICRAR

I trained a classifier and now I don't know what to do with it

Matthew Alger

Slides: http://www.mso.anu.edu.au/~alger/mso2017

MSS317

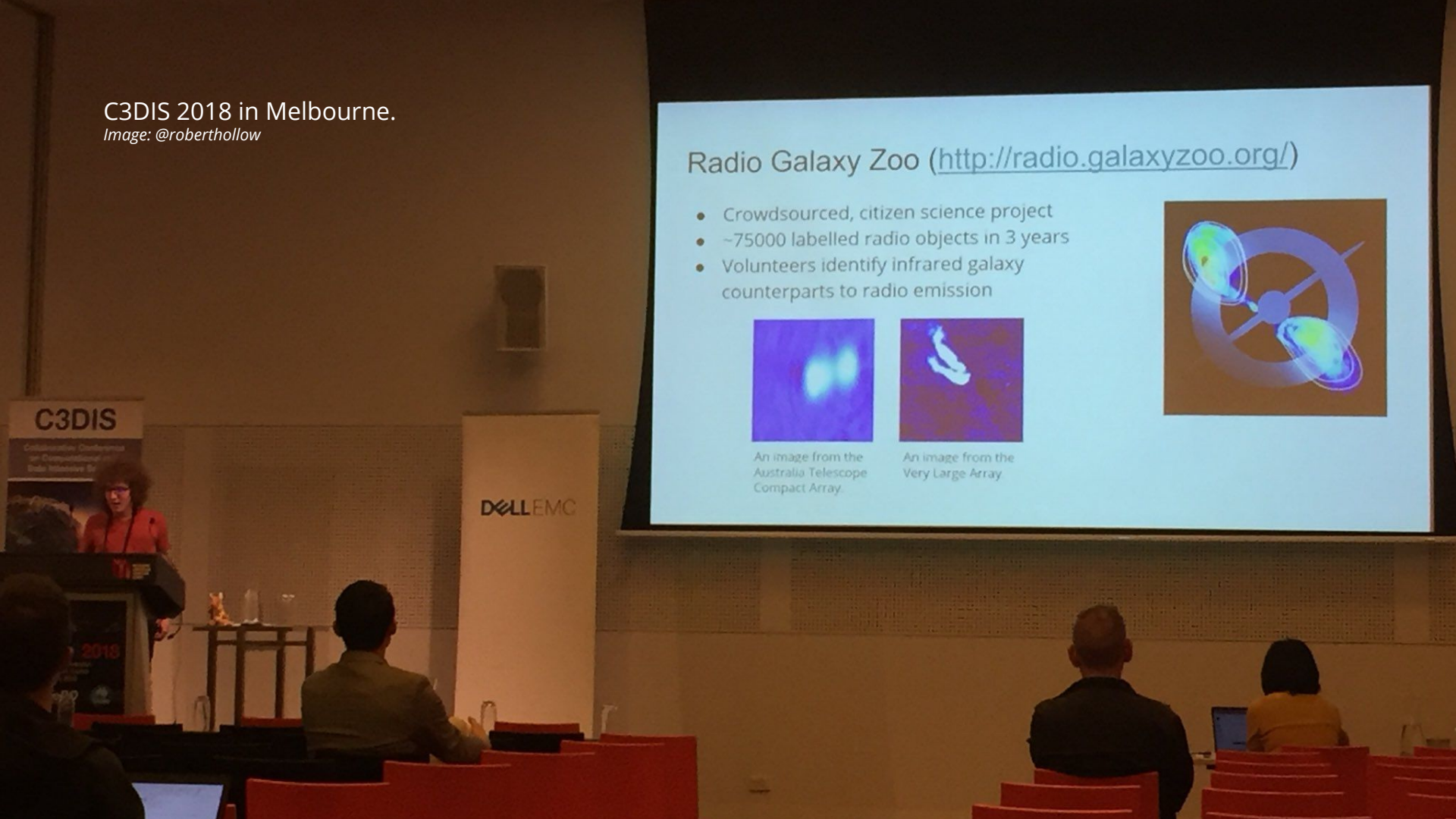
Machine Learning for Radio Astronomy: Everything is binary classification if you phrase it right

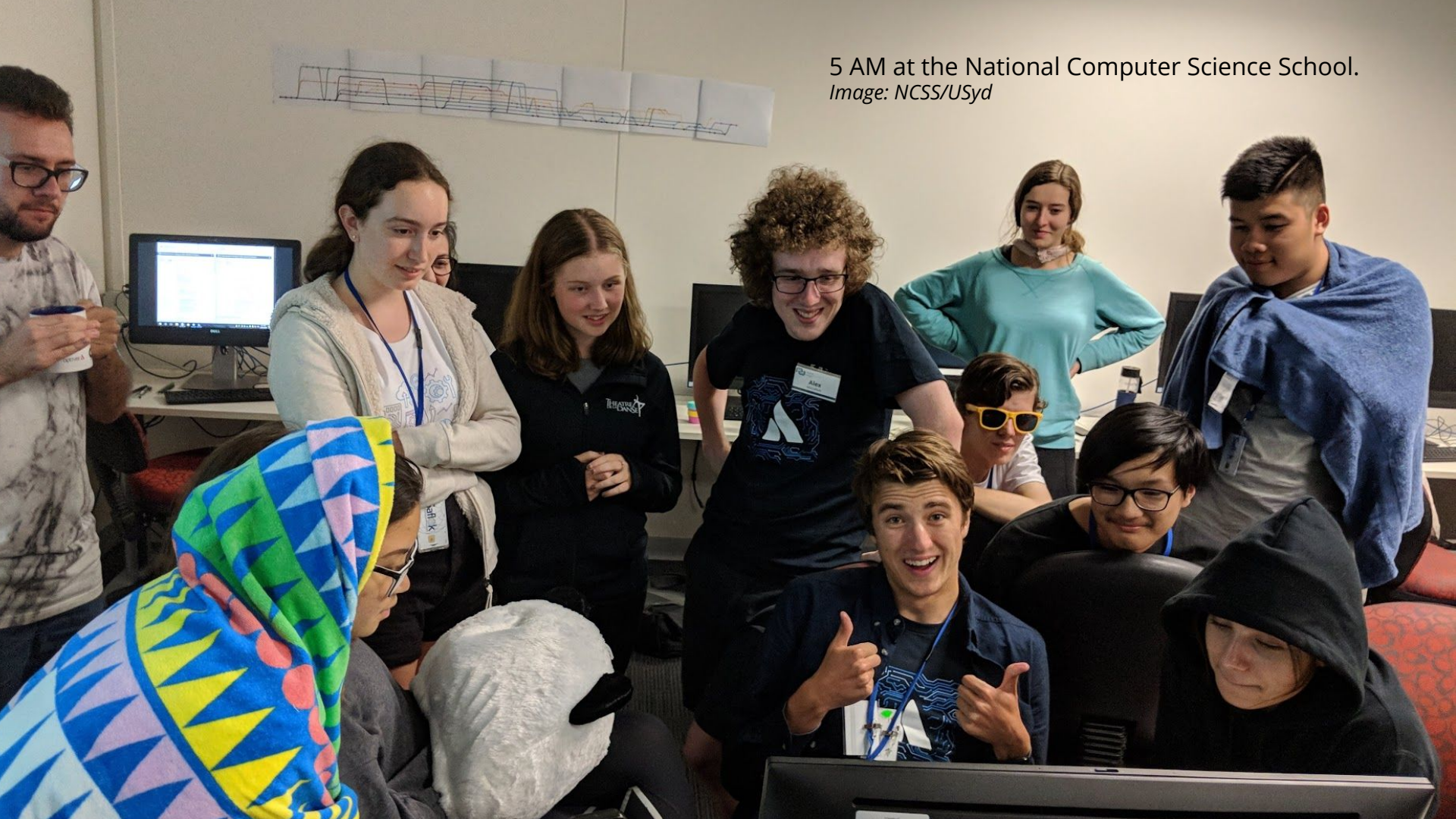
Matthew Alger

Slides: http://www.mso.anu.edu.au/~alger/radio-lunch-may

MSO

C3DIS 2018 in Melbourne.
Image: @roberthollow

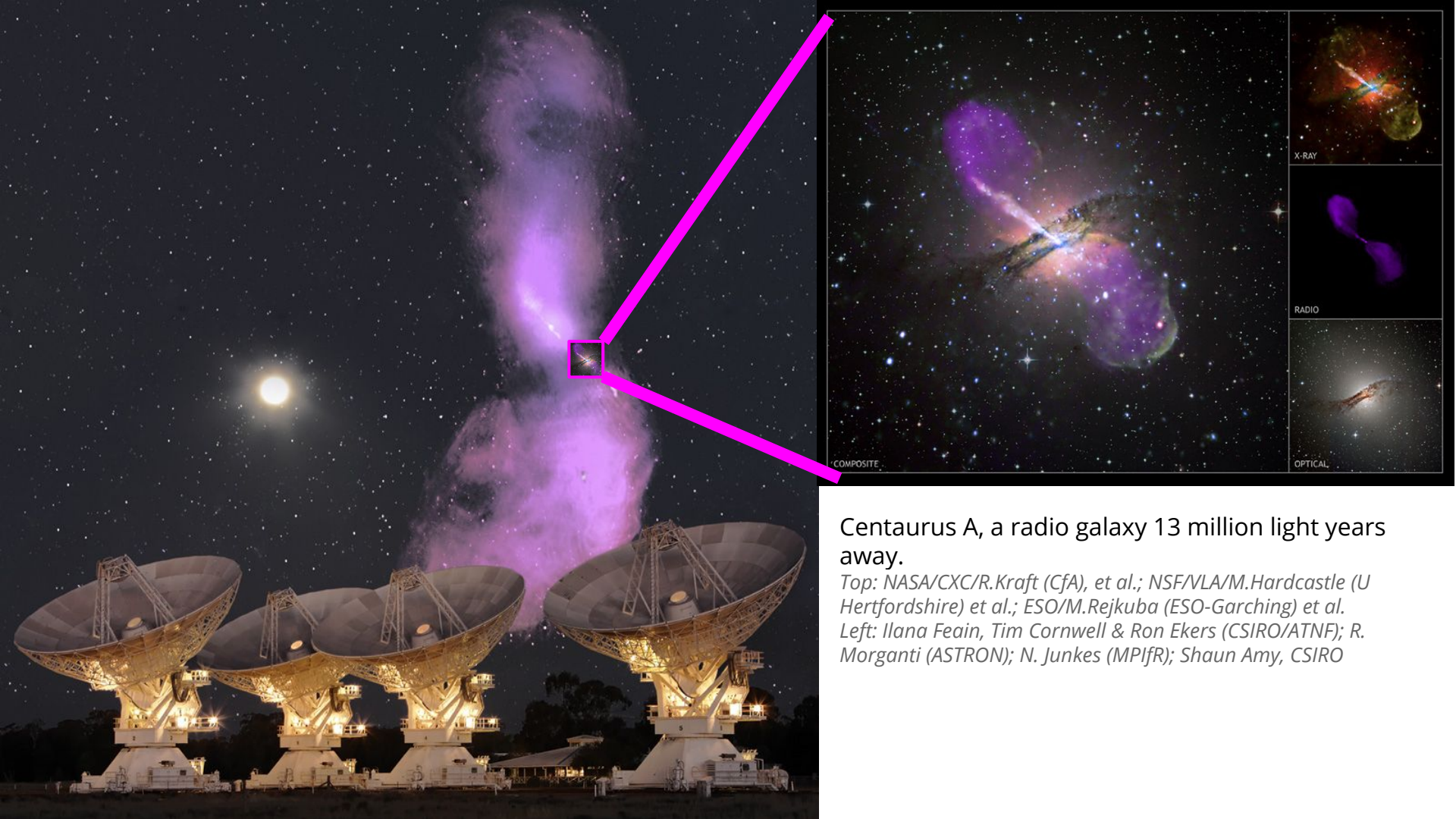




5 AM at the National Computer Science School.
Image: NCSS/USyd

Machine learning and cross-identification

Image: ESO

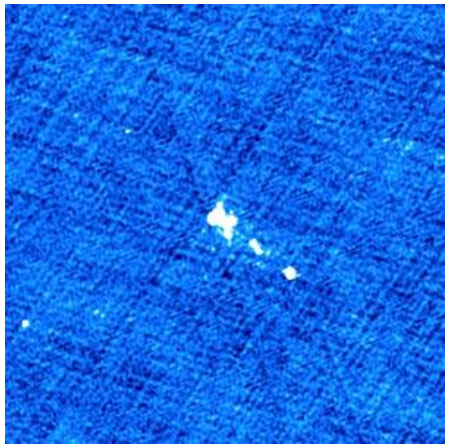


Centaurus A, a radio galaxy 13 million light years away.

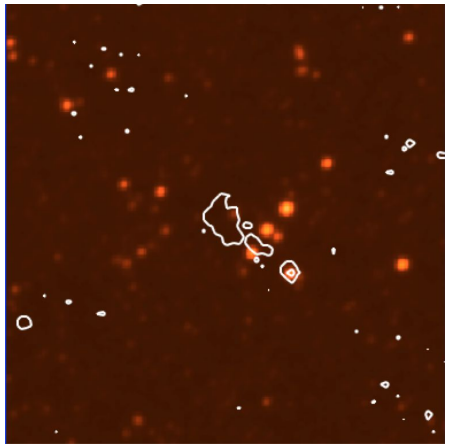
Top: NASA/CXC/R.Kraft (CfA), et al.; NSF/VLA/M.Hardcastle (U Hertfordshire) et al.; ESO/M.Rejkuba (ESO-Garching) et al.
Left: Ilana Feain, Tim Cornwell & Ron Ekers (CSIRO/ATNF); R. Morganti (ASTRON); N. Junkes (MPIfR); Shaun Amy, CSIRO

Radio/infrared cross-identification

- Problem:
  - Match radio emission to the corresponding galaxy in infrared
  - Important for understanding galaxies throughout cosmic time
- Hard:
  - Radio emission can be very extended across the sky
  - Often no clear relationship between radio emission and the emitting galaxy



Radio image from FIRST.
Image: FIRST



Infrared image from *WISE*. White contours show the radio image on the left.
Image: WISE

Cross-identification as binary classification

- We can represent cross-identification as binary classification
- A good framework:
  - Binary classification is well-understood
  - Lots of readily applicable off-the-shelf classification models
  - Easy to train
- Problems:
  - Converting cross-identification catalogues to binary labels loses information
  - Unclear how uncertainties in this formulation are related to dataset or physical uncertainties

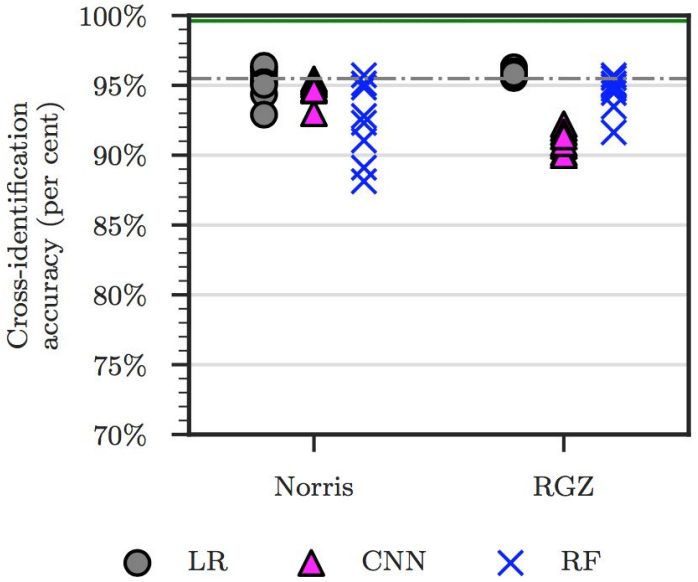
$$f : \mathbb{R}^d \rightarrow \mathbb{R}$$

$$\mathrm{xid} : \mathrm{Radio} \rightarrow \mathrm{IR}$$

$$\mathrm{xid}(r) = \underset{i \in \text{IR objects}}{\operatorname{argmax}}\, f(i)$$

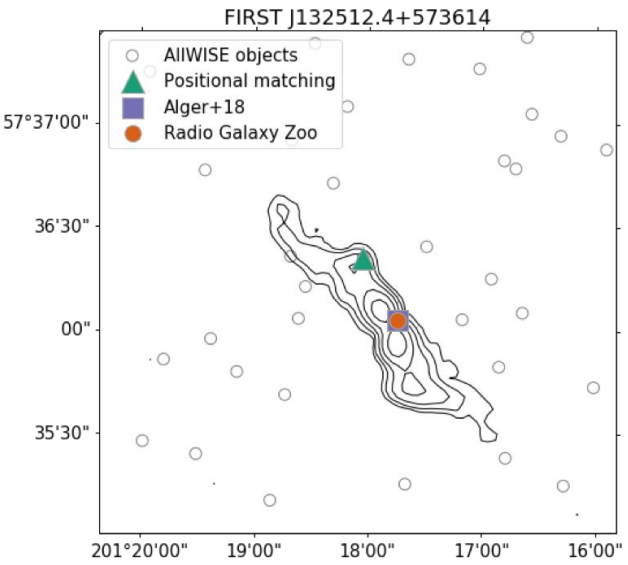
Results from ATLAS

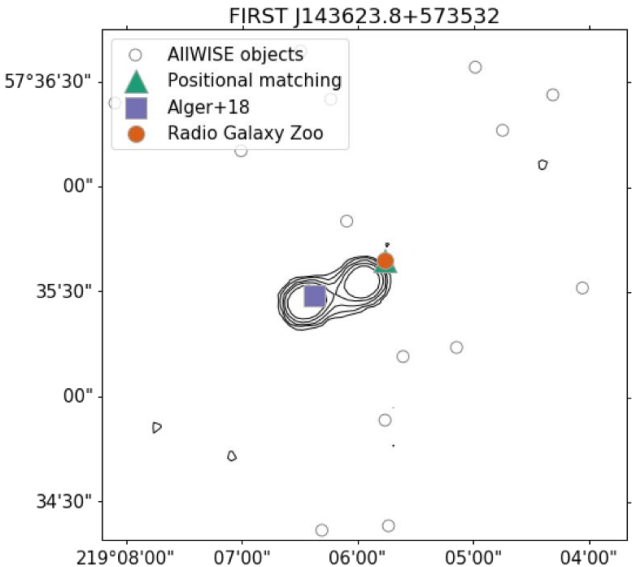
- Nothing works as well as simple positional matching
- Results generalise across different fields not in the original training set
- RGZ-trained methods performed comparably to expert-trained methods

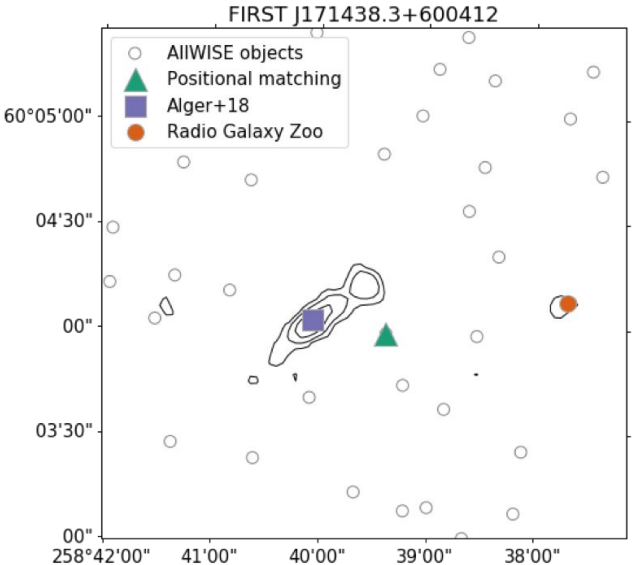


Alger et al. (2018), doi:10.1093/mnras/sty1308

Preliminary examples from FIRST

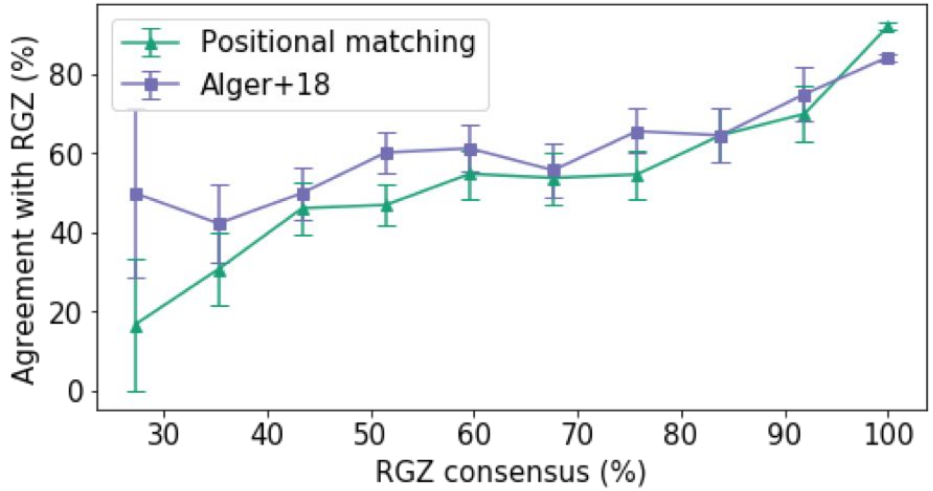


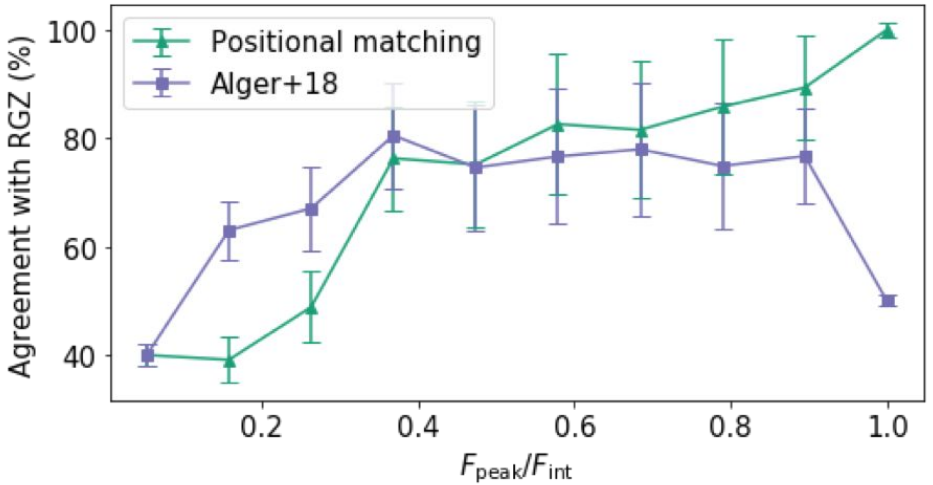




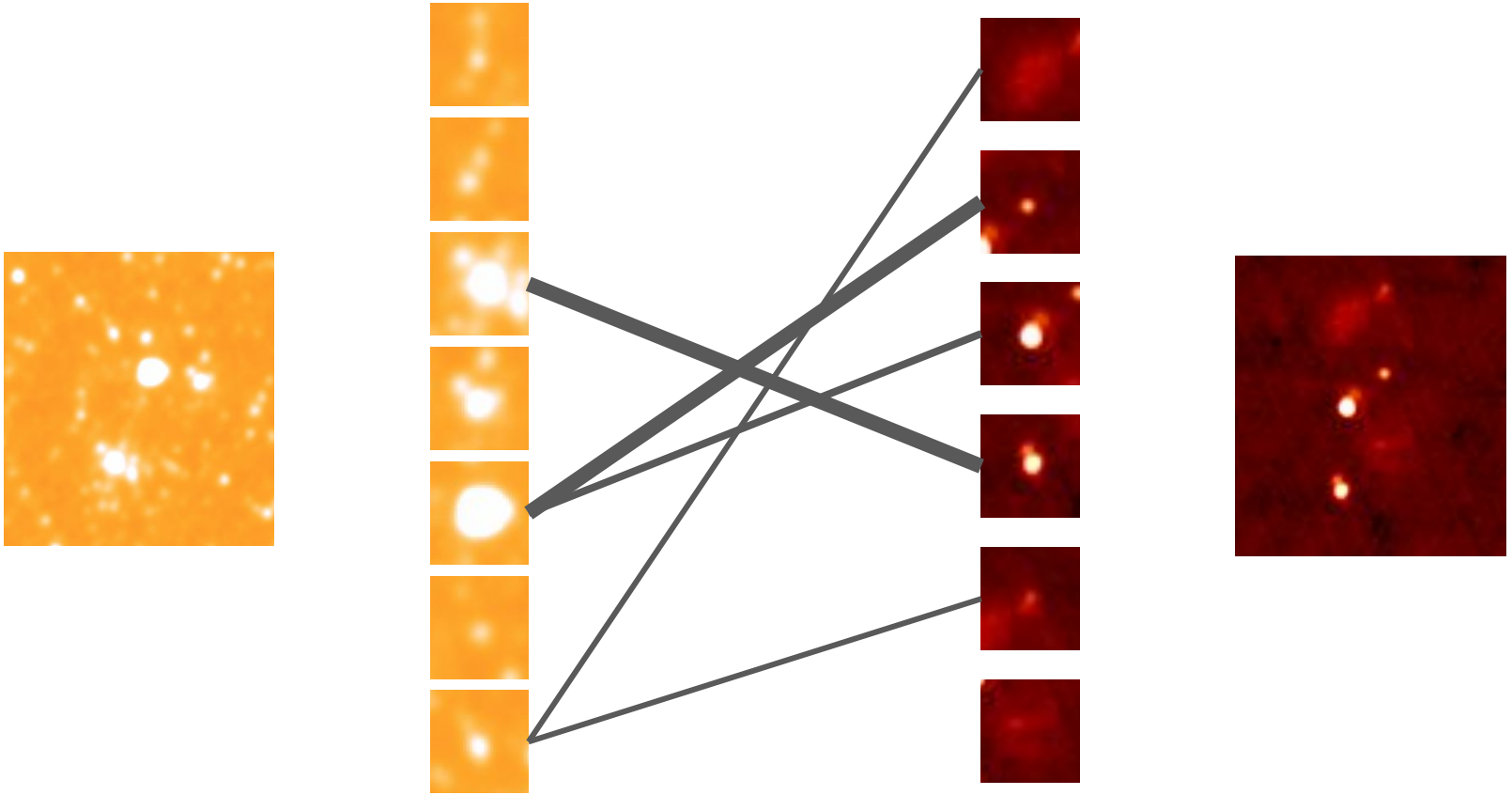
Alger et al. (in prep)

Preliminary results from FIRST

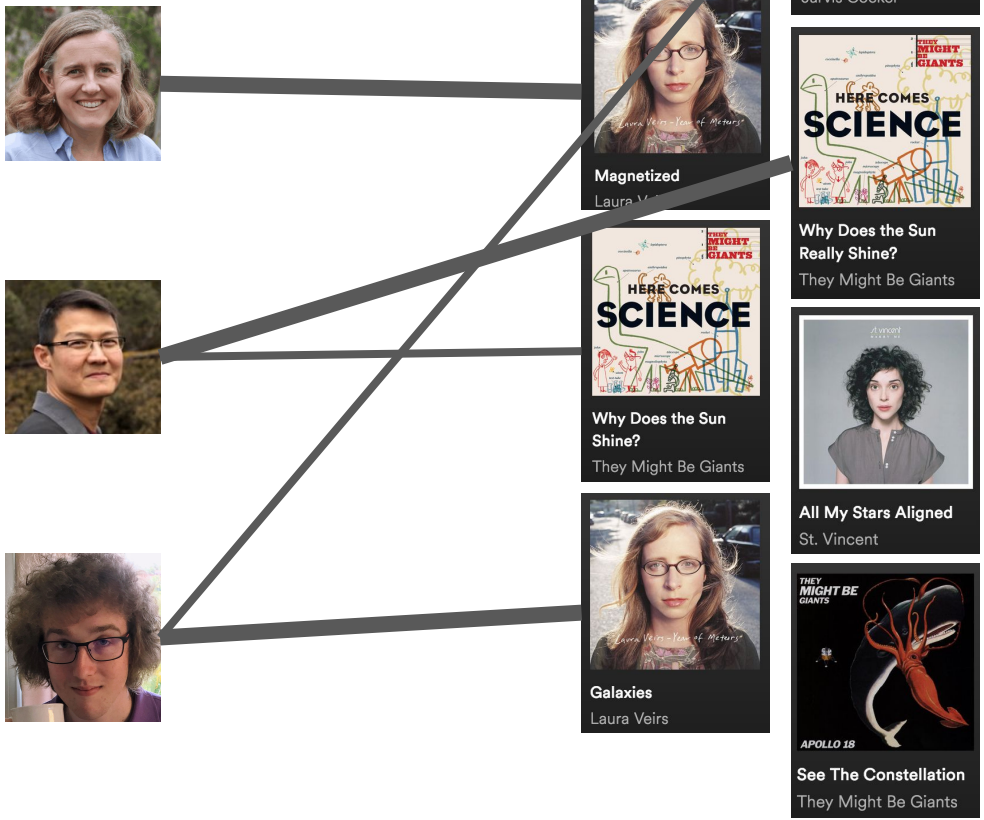




Bipartite graph edge finding



Bipartite graph edge finding



Working with Dawei Chen on the related problems of music playlist generation and radio-IR cross-identification.


Where to?

Plan for the next year

- Get involved with the POSSUM source finding research group
- Further analyse statistical problems in astronomical citizen science
- Apply recommendation techniques to cross-identification in POSSUM
- *Radio Galaxy Zoo: Automated radio cross-identification for FIRST*
  Alger et al. (in prep)
- *Radio Galaxy Zoo: Data Release 1 — morphological classifications of 75,000 FIRST radio sources*
  Wong et al. (in prep)
- Paper 3 on polarised cross-identification and population statistics