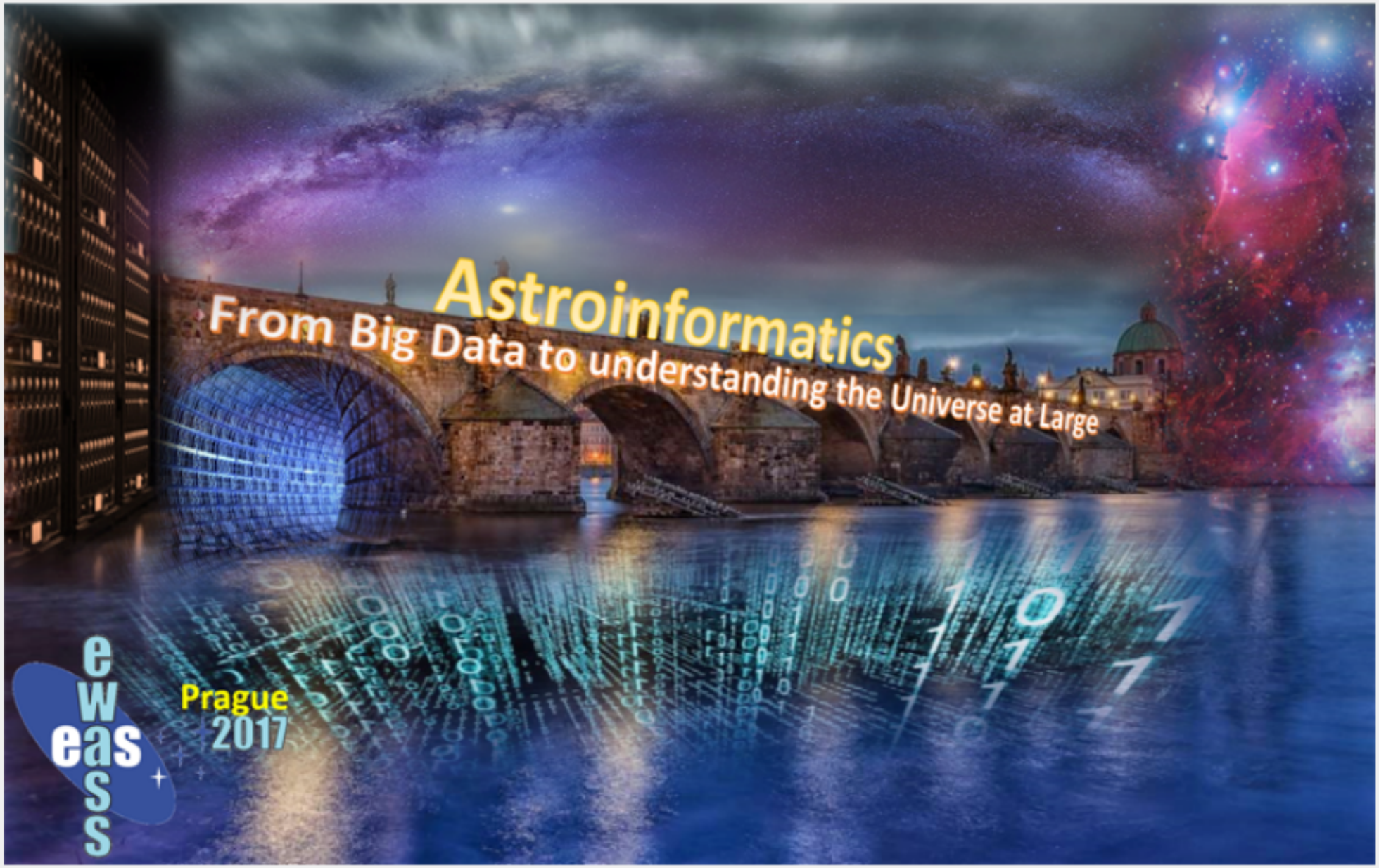
Symposium S14

29 – 30 June 2017



Artwork by Massimo Brescia (THANKS !)

Welcome

- Astroinformatics – first time at EWASS !

  Milestones :
- Astroinformatics (2009) Position Report NSF
- „Lodge of Freemasons“ like meetings (2010)
  - Great effort of Prof. G. Longo and Ashish Mahabal (here)
- 2015 – IAU GA – FM8 Statistics and Exoplanets
- 2016 – IAU S325 Sorrento – Astroinformatics
- IAU Commission B3 – 207members,
- International Astrostatistics Association, COIN (Rafael de Souza))
- 2017 EWASS S14 ;-) AI named for 1200 people

Logistics

- Be understandable to WIDE audience !
- Publishing – Zenodo + extended abstract (1-2p)

- Friday is COST BSE day of S14 (sign !)
- join COST Big Sky Earth - Transdisciplinary !
- Preparing book about Big Data and Knowledge Discovery in SKY and EARTH observation

  Who wants to collaborate ?

Astroinformatics: The Methodology of Knowledge Discovery in Petabyte-Scaled Archives

Petr Škoda

Astronomical Institute of the Czech Academy of Sciences

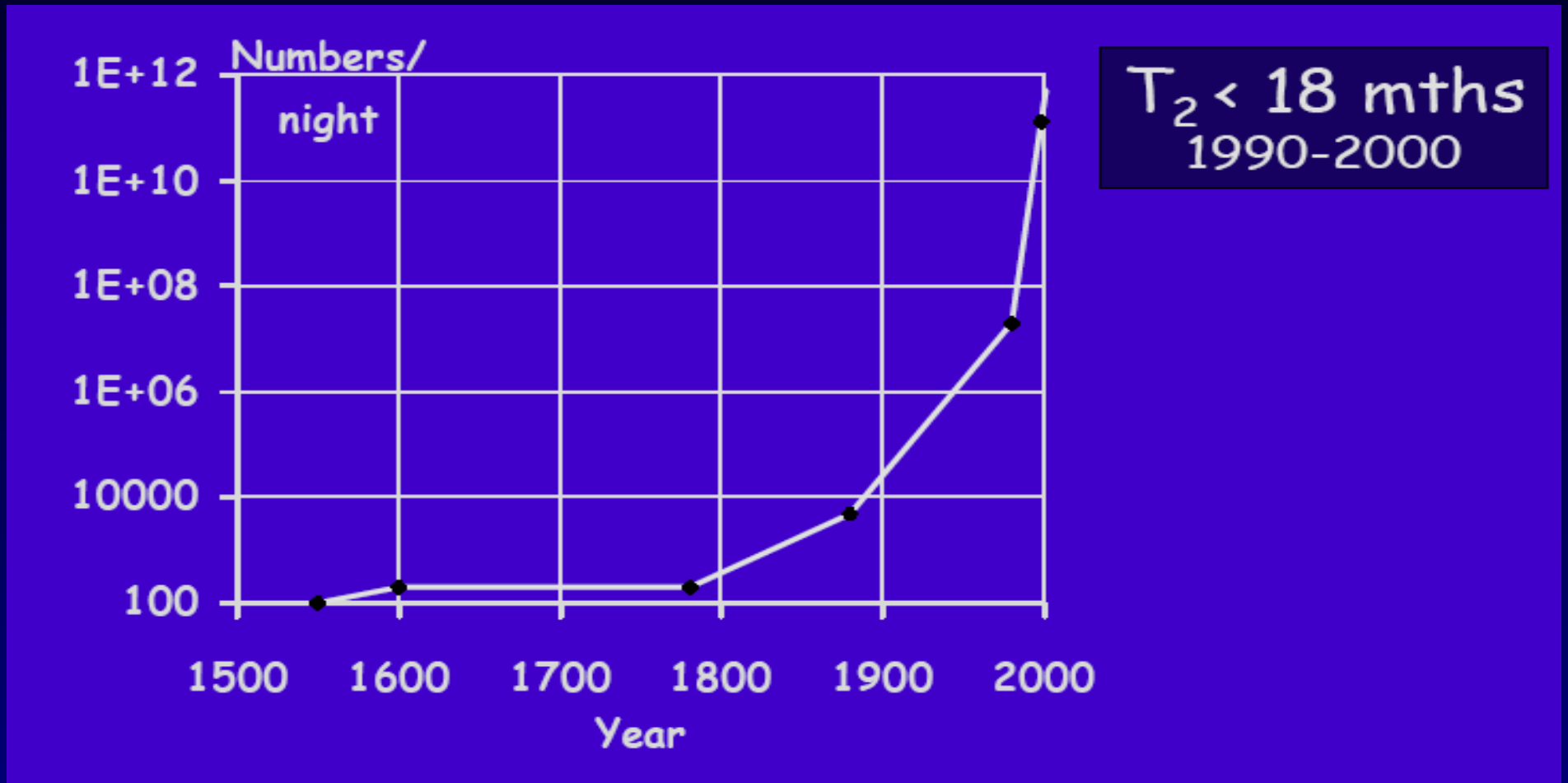
Supported by grant COST LD-15113 of the
Czech Ministry of Education Youth and Sports

EWASS 2017 – S14 Intro
Prague , Czech Republic, 29-30.6.2017

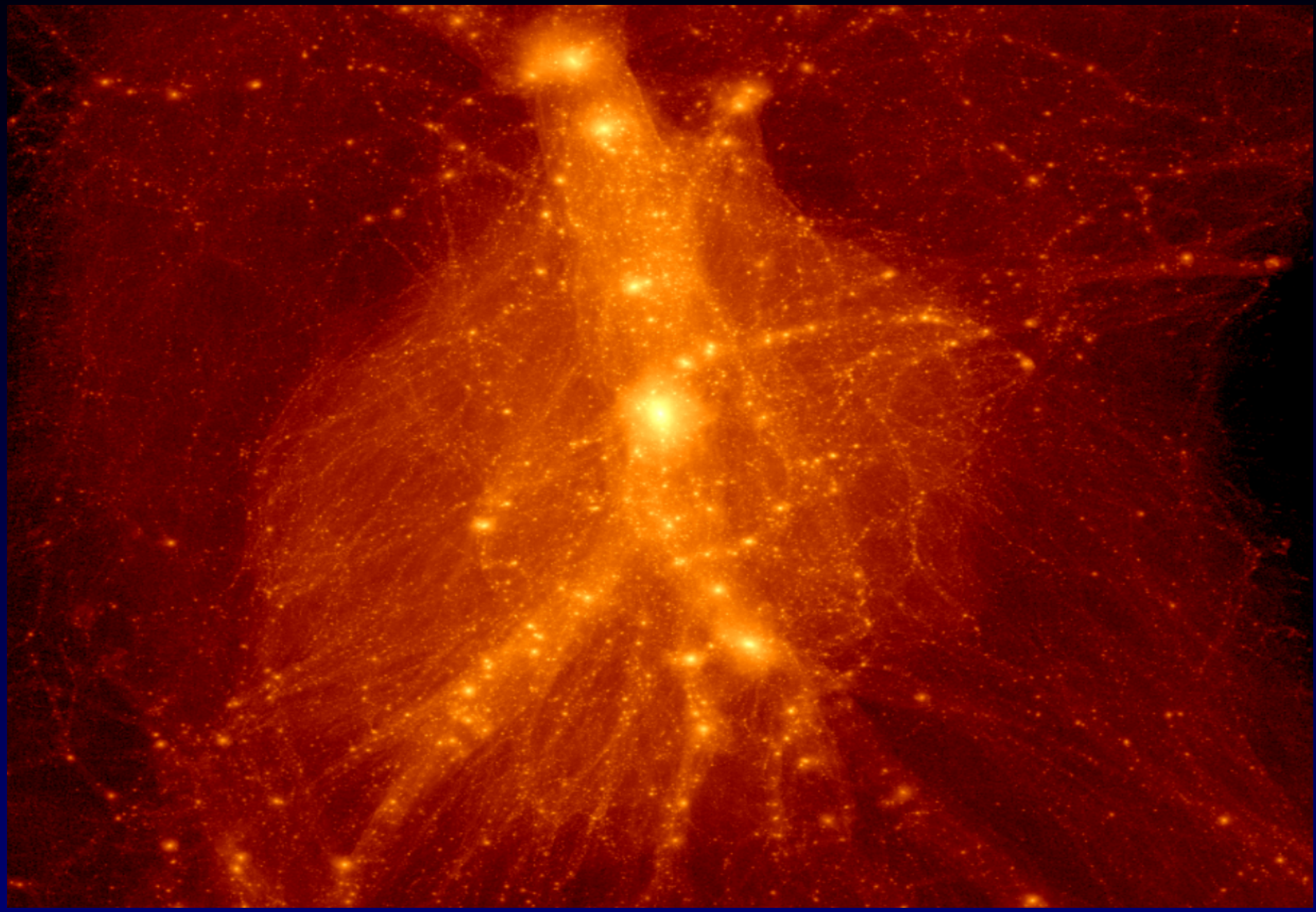
Data Avalanche

Moore law for chips –doubling 1.5 year

Data in astronomy – doubling < 1 yr ! (1000/10 yr)



Visualization of Big Data

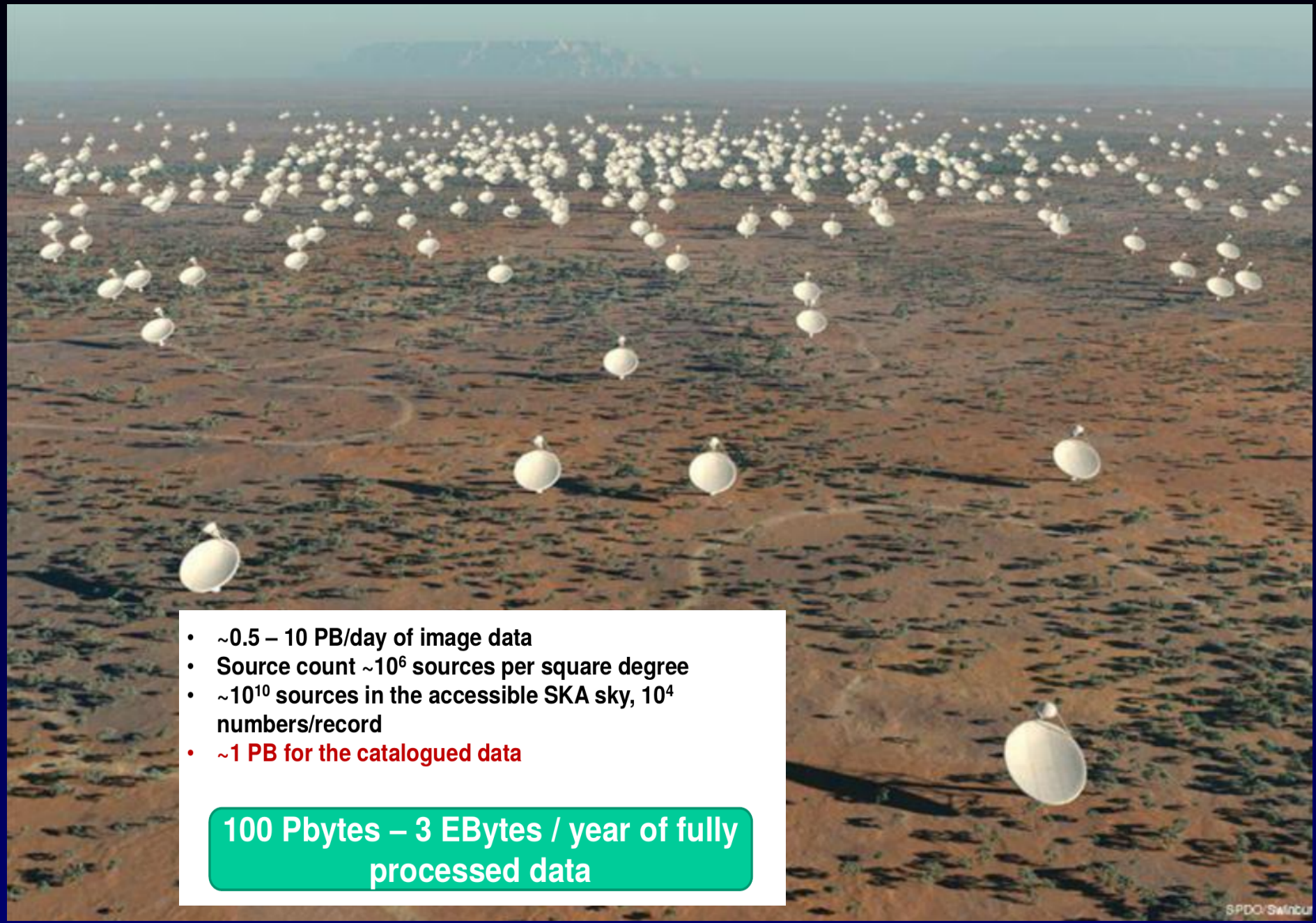


Breddels 2016

LOFAR network

	LOFAR	SKA
Raw Telescope	112 PB/yr	60 EB/yr
Archive Rate	6 PB/yr	100 PB/yr

SKA



- ~0.5 – 10 PB/day of image data
- Source count ~10^6 sources per square degree
- ~10^{10} sources in the accessible SKA sky, 10^4 numbers/record
- ~1 PB for the catalogued data

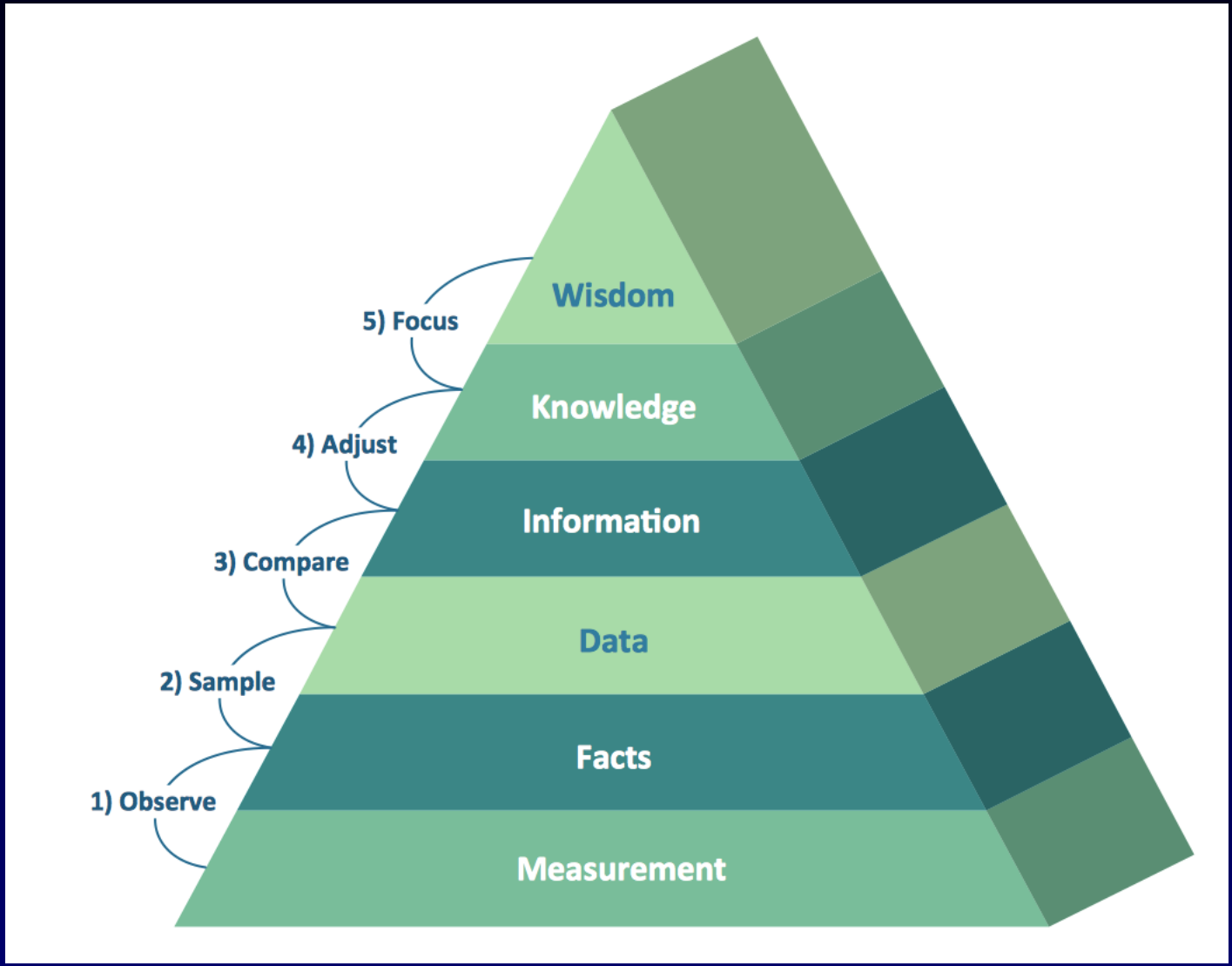
100 Pbytes – 3 EBytes / year of fully processed data

Large Synoptic Survey Telescope

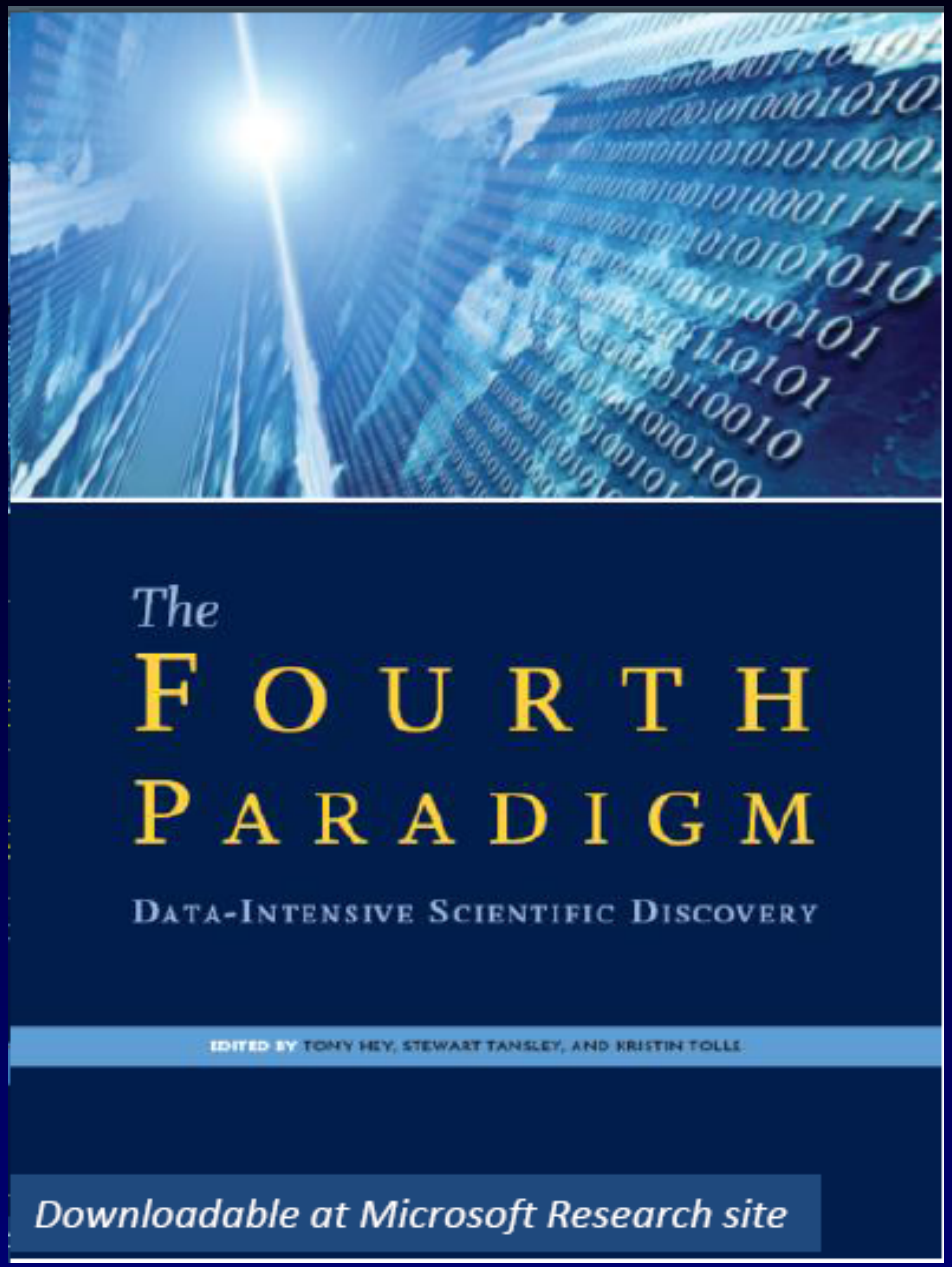
201 CCD 4kx4k,
3.2 Gpix every 20 sec
3.5 deg FOV (64cm)
20 TB/day=6 PB/yr RAW
1.5 PB catalogue !!!
detection of changes 60s!

38 billion objects x 1000
32 tril. Meas. - 5 PB table

Data-Knowledge-Wisdom Pyramid



X-informatics



Jim Gray heritage (SDSS)

E-Science

Data intensive science

Data Driven Science

What is Fundamentally New Here?

- The ***information volumes and rates*** grow exponentially

➡ ***Most data will never be seen by humans***

- A great increase in the data ***information content***

➡ ***Data driven vs. hypothesis driven science***

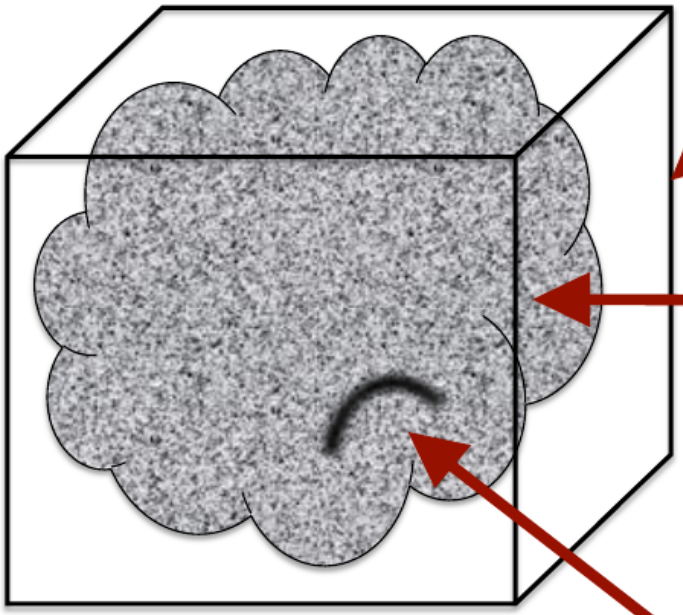
- A great increase in the ***information complexity***

➡ ***There are patterns in the data that cannot be comprehended by humans directly***

Djorgovski

Hidden Patterns in Data

Pattern or structure (Correlations, Clustering, Outliers, etc.) Discovery in High-Dimensional Parameter Spaces



D >> 3 parameter space hypercube

High-D data cloud: mostly noise, of an arbitrary distribution

But in some corner of some sub-D projection of this data space, there is ***something ≠ noise***

BIGSKYEARTH

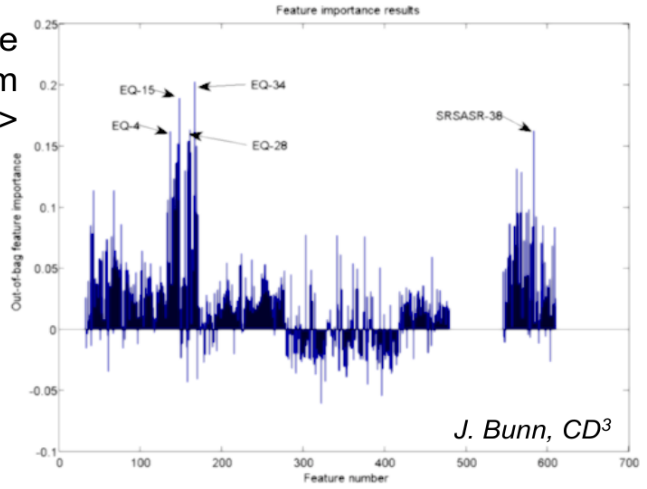
Big Data Era in Sky and Earth Observation

TD COST Action TD 1403

Astro-Neurology

From Sky Surveys to Neurobiology

- Using the data analytics tools based on ML, developed for the analysis of sky surveys, to design a better diagnostics for autism
- Feature importance using random forests =>
- Next: correlate with MRI scans



(with R. Adolphs et al.)

J. Bunn, CD³

Djorgovski

SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM

JBEMi JOURNAL OF BIOMEDICAL ENGINEERING AND MEDICAL IMAGING

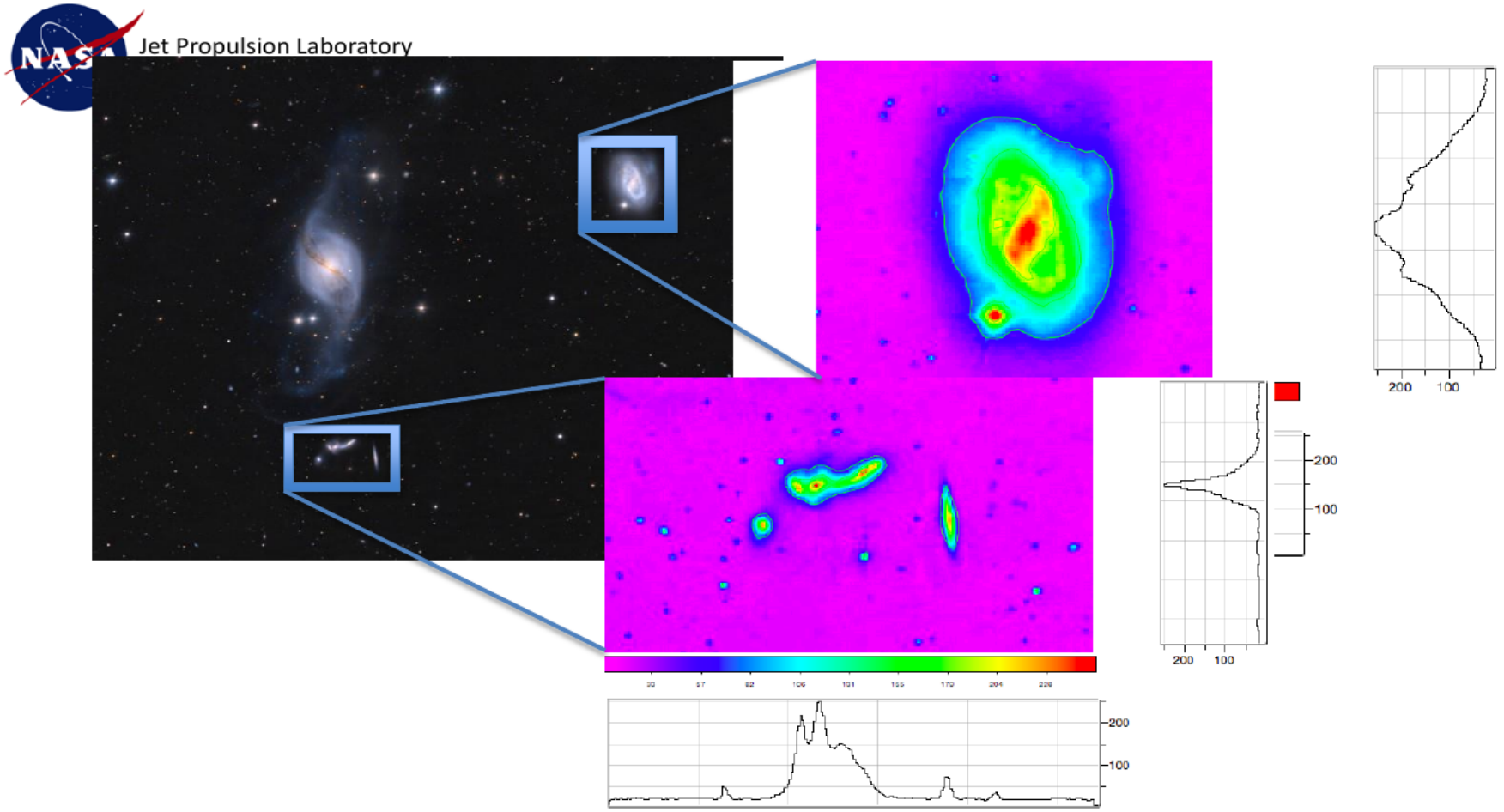
ISSN: 2055-1266
VOLUME 3 ISSUE 3

Visualization of Decision Tree State for the Classification of Parkinson's Disease

[1]David P. Williams, [1]Deborah Mudali , [2]Hugo Buddelmeijer , [2,5]Parisa Noorishad, [3]Sanne Meles , [3]Remco J. Renken , [4]Klaus L. Leenders, [2]Edwin A. Valentijn, [1]Jos B.T.M. Roerdink

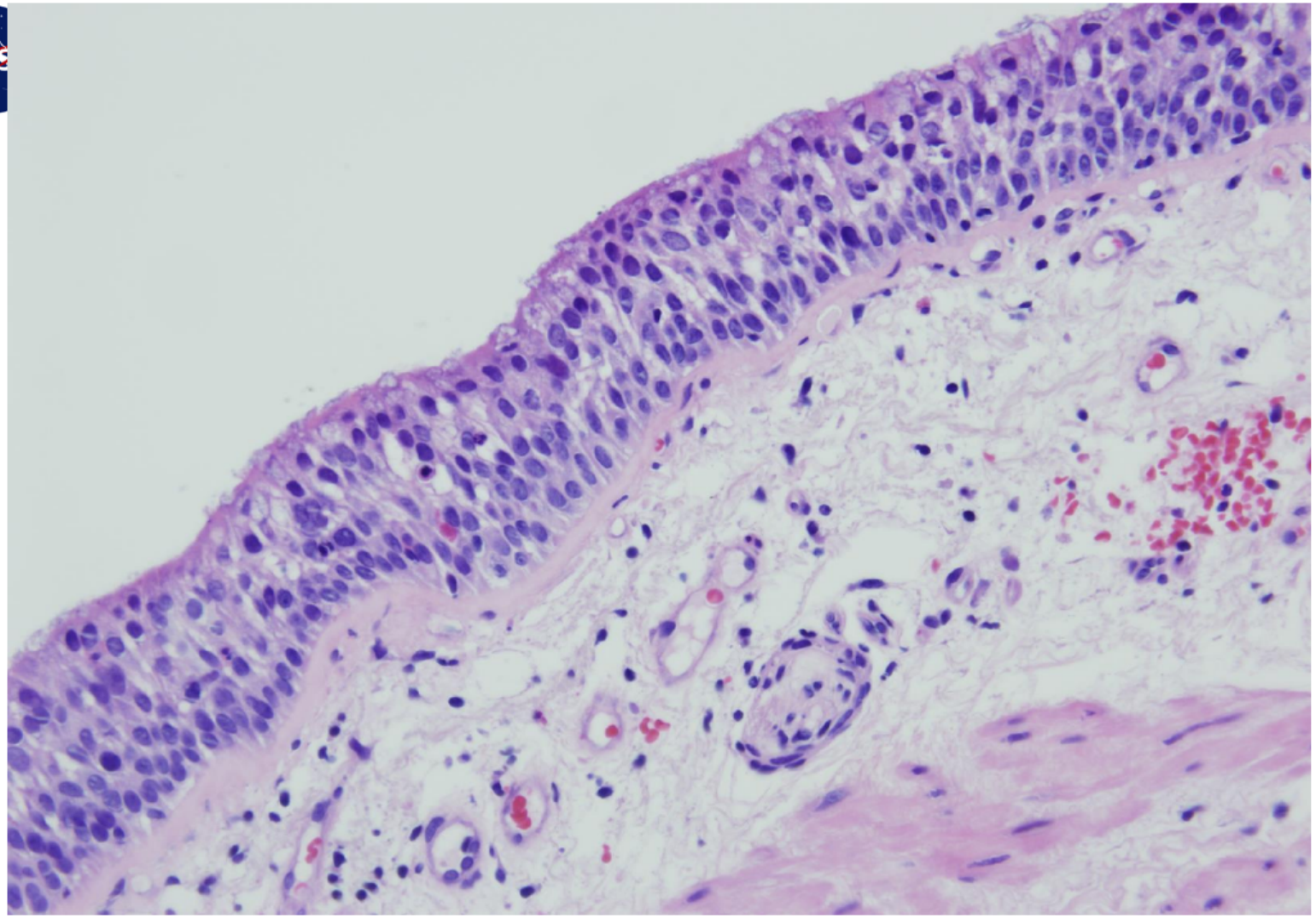
[1]Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen

Finding Galaxies by Shape NASA



Description: Detecting objects from astronomical measurements by evaluating light measurements in pixels using intelligent software algorithms.
Image Credit: Catalina Sky Survey (CSS), of the Lunar and Planetary Laboratory, University of Arizona, and Catalina Realtime Transient Survey (CRTS), Center for Data-Driven Discovery, Caltech.

Finding Cancer Signatures NASA



Description: Detecting objects from oncology images using intelligent software algorithms transferred to and from space science.
Image Credit: EDRN Lung Specimen Pathology image example, University of Colorado

Practical usage of AI

Astronomy - good to outreach but diificult to explain practical impact

Astroinformatics – brings „industrializable“ results - Funding agencies

AND IT IS GREAT FUN !