

Title: Open is the New Black: Really?

Date: Mar 27, 2018 02:45 PM

URL: <http://pirsa.org/18030114>

Abstract: Through the last 20-25 years we've won many battles in the evolution of FLOSS into mainstream. No one can ignore today the role of open source in software, hardware, high-tech and even business development. However everything seems to be open today: Open Data, Open Innovation, Open Government, Open Research...what do we mean by that? Has "open" the same meaning in all of them? How reliable are the results from such openness? What about policies and science and technologies designed on top of them?

This talk will share Open Parallel's five years journey through the pre-construction challenges of the largest scientific instrument of the next decade -the Square Kilometre Array radio-telescope (SKA). Will present how a non-central country as New Zealand has a say on its design plus how open source software will be core to its success. Being involved in the OS side of the SKA, will also share some concerns around black swans and ask a few questions around cybersecurity. Open Science? Yeah, right.



Open Parallel
Innovation Partners

I

“Open is the New Black”

Presentation at

“Open Research: Rethinking Scientific Collaboration”
Conference at the Perimeter Institute

Waterloo, Canada, 27th March 2018

“Open is the New Black”

Presentation at

“Open Research: Rethinking Scientific Collaboration”
Conference at the Perimeter Institute

Waterloo, Canada, 27th March 2018

Nicolás Erdödy
Founder-CEO, Open Parallel Ltd



Open Parallel
Innovation Partners

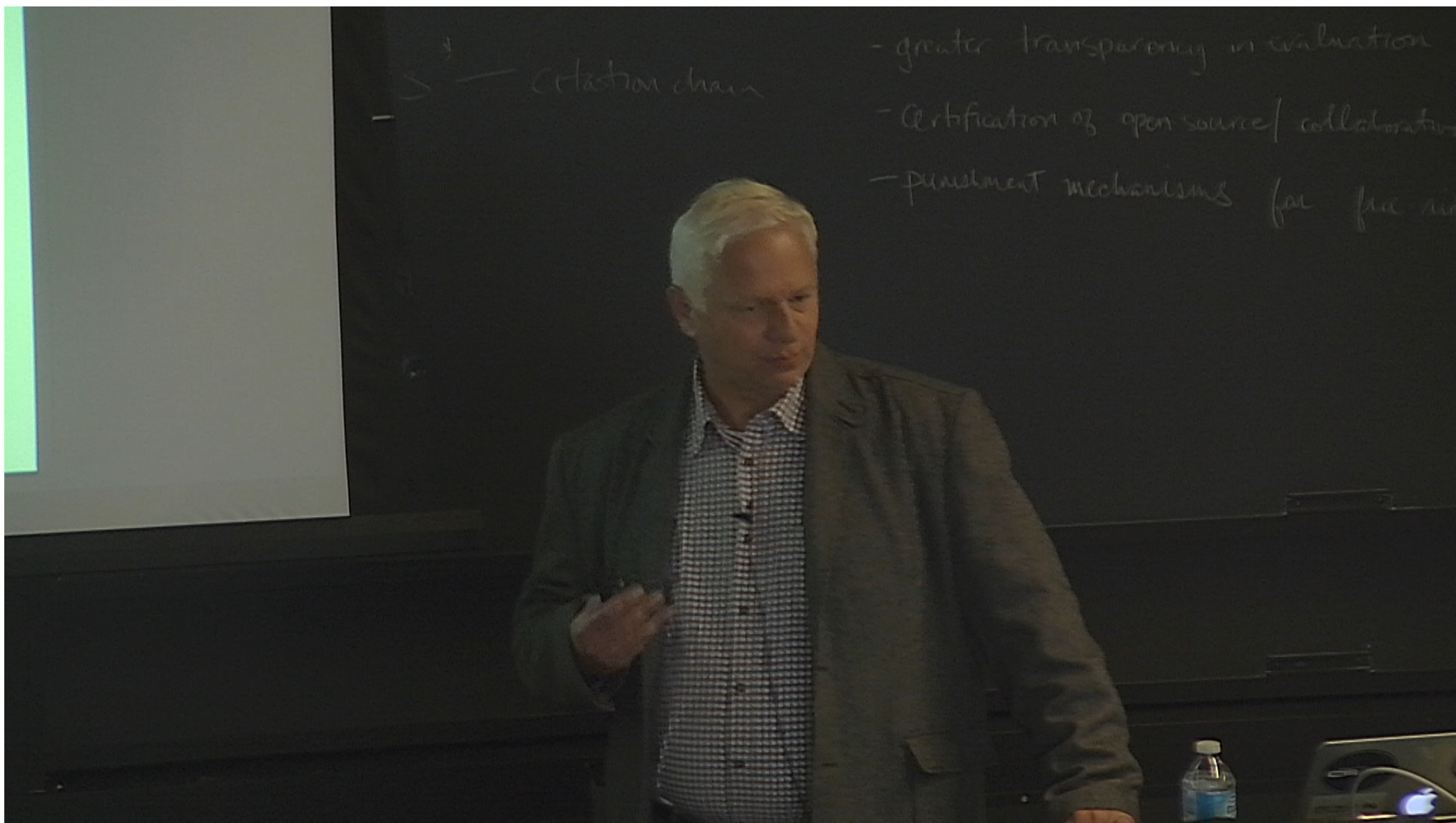
“Open is the New Black”

Presentation at

“Open Research: Rethinking Scientific Collaboration”
Conference at the Perimeter Institute

Waterloo, Canada, 27th March 2018

Nicolás Erdödy
Founder-CEO, Open Parallel Ltd



- \supset — citation chain

- greater transparency in evaluation
- certification of open source / collaborative
- punishment mechanisms for free riders

Agenda

- Take Aways
- Introduction (NZ, NE, OP, MW)
- The SKA - science, milestones
- Open Parallel & the SKA
- Videos (2)
- OSS ethos and bikesheds
- Black Swans and cybersecurity
- Some considerations and questions
- Architecture for Participation

Agenda

- Take Aways
- Introduction (NZ, NE, OP, MW)
- The SKA - science, milestones
- Open Parallel & the SKA
- Videos (2)
- OSS ethos and bikesheds
- Black Swans and cybersecurity
- Some considerations and questions
- Architecture for Participation

Open Parallel
Innovation Partners



Take aways

- Big Data just started: SKA is one specific case
- Exascale will really change things in the 2020s -tide or tsunami?
- Will your algorithm scale? What about your performance per watt?
- FOSS is here to stay -successful computing is modular, and open
- Heterogeneity could lead to vertical integration (are we going back 50 years?)
- Do you need 22+ million lines of code? (Linux 4.13). Really?
- How reliable is your data? And your Science?
- There are rules about security that would avoid Black Swans in your project
- Less is more (less code, less problems!)
- “Culture eats strategy for breakfast” (Peter Drucker)
- “Talent needs purpose” (NE) – Think about OSS ethos
- Architecture for Participation – OpenStack (Catalyst – New Zealand)

Open Parallel
Innovation Partners

A bit of history and perspective



Perception matters



Why New Zealand?

- *Note: Factoids*
- **#1 Prosperity** (<http://www.prosperity.com/rankings>)
- **#1 Ease to do business**
(<http://www.doingbusiness.org/data/exploreeconomies/new-zealand/>)
- **#1 Non-corruption** (<https://www.transparency.org/country/NZL>)
- **All 8 NZ universities are in the Top450** (from 26.000 universities ww)
- **97% of businesses are SME** (less than 20 employees)
(<http://www.mbie.govt.nz/info-services/business/business-growth-and-internationalisation/documents-image-library/sbdg-2016-report.pdf>)
- **Wellington #1 best city in the world** (Deutsche Bank)
- We are far from everywhere...and only **4.7million** of us
- Rugby? Hobbits? Outdoors? Lifestyle? Community?
- Make a difference? Or just to see things ...with a different perspective?

Open Parallel
Innovation Partners





Finns & Hungarians chat about common origins



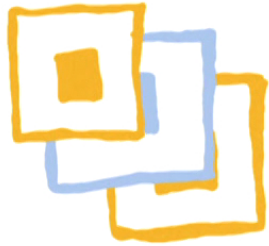
Introduction

- Born in Uruguay, living in New Zealand since 2003
- Maths, Engineering (UdelaR-Uruguay), MEntr (Otago-NZ)
- Academia Universal (e-learning, Latin America)
- Venture Capital (NZVCA), Angel Investment (AANZ)
- 20+ knowledge based start-ups
- Open Parallel (2010)
- Involved with the SKA project since 2011 - CSP / SDP since 2013
- Multicore World conferences since 2012
- Linux Conference Australia (LCA) 2006-2015
- NZOSS Councillor 2013-2016

Open Parallel
Innovation Partners







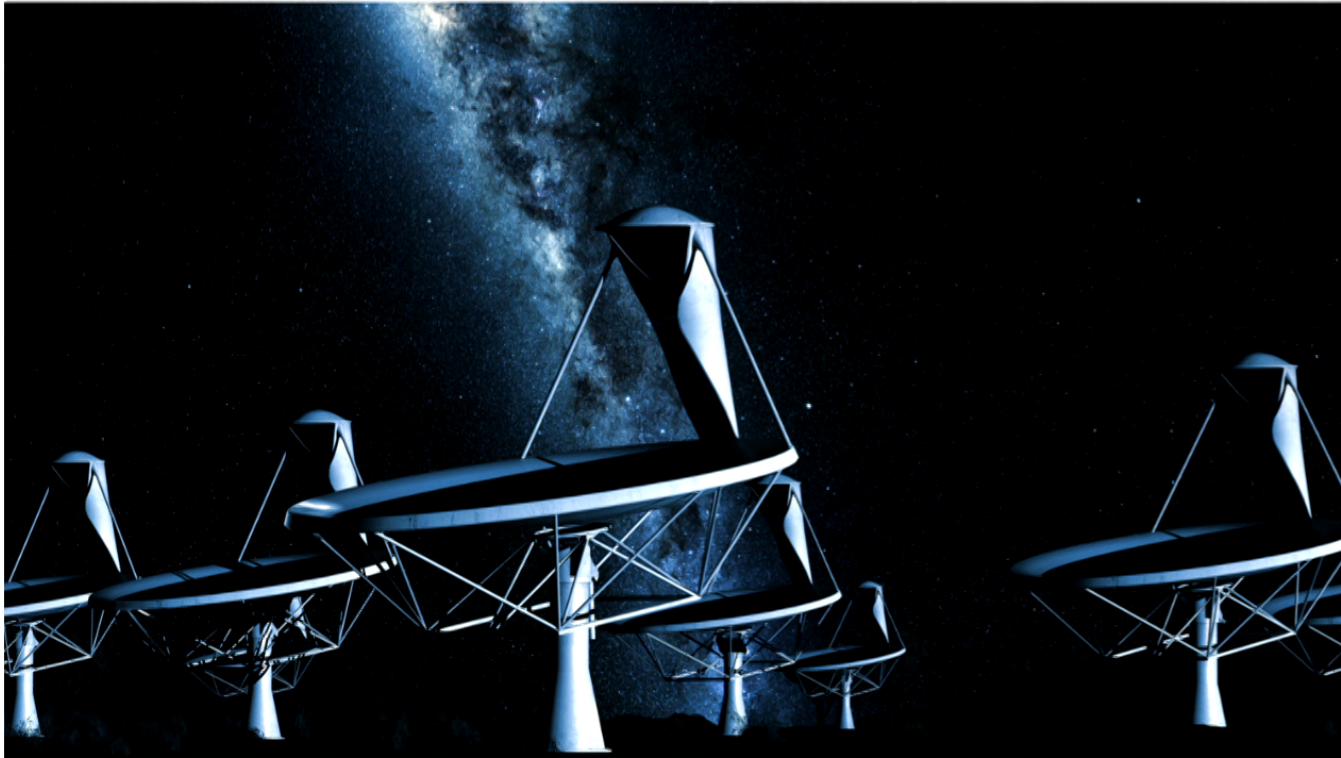
MULTICORE WORLD

- 8th edition: **12-13-14 February 2019**
- Wellington, New Zealand
- **Prof Satoshi Matsuoka**, Tokio Institute of Technology, Japan
- **Drs Ruud van der Pas**, DE Oracle / SPARC / OpenMP, Linux Performance, US
- Plus 20+ distinguished speakers
- Multicore.world

Open Parallel
Innovation Partners

What is the SKA project?

(Square Kilometre Array)



Open Parallel
Innovation Partners

What is the SKA project?

- A complex scientific, political and engineering effort to build the world's largest, most powerful radio-telescope.
- Image resolution: **50x** Hubble Space Telescope.
- Distributed between Australia & South Africa.
- www.skatelescope.org/project/
- Also “**The Largest Supercomputer of the World**”.
- 1+ ExaFlop to process and reduce the massive amount of data generated by the sensors (that's **10^{18}**).
- All dishes together will produce **~160Tb/s**.
- Extreme data throughput requirements (**8x** of today's largest supercomputers).

Open Parallel
Innovation Partners

Where?

- Western Australia and Southern Africa
- South Africa, Botswana, Namibia, Ghana, Mozambique, Mauritius, Madagascar, Zambia
- **Ten Full SKA Member Countries** (today): Australia, South Africa, UK (SKAO HQ), Canada, China, India, Italy, Netherlands, New Zealand, Sweden
- Site selection decided in 2012 between **competing proposals** from South Africa vs Australia & NZ
- **Potential new members:** France, Germany, South Korea, Japan, Malta, Portugal, Spain, Switzerland

Open Parallel
Innovation Partners

What for?

The SKA will aim to address **five fundamental questions about the Universe:**

- 1) How do galaxies evolve and what role does dark energy play?
- 2) What generates the huge magnetic fields in the space?
- 3) Was Einstein right? Test relativity theory
- 4) How were the first black holes formed?
- 5) Are we alone in the Universe?

Open Parallel
Innovation Partners

Did you know?

Estimated distribution of energy and matter in the Universe:

- Dark Energy ~ 70%
- Dark Matter ~ 25%
- **Ordinary Matter** (what we actually know / understand) ~**5%**

(source: Wikipedia / European Space Agency)

Open Parallel
Innovation Partners

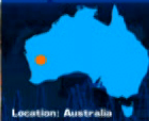
Science Data Processor (SDP)

- SDP designs compute hardware platforms, software and algorithms needed to process data from the correlator into data science products
- <https://www.skatelescope.org/sdp/>
- Since 2013 Open Parallel contributes to the SDP Compute Platform - Common Software (non-domain software)
- Currently (partially) funded by Ministry of Business, Innovation and Employment - MBIE (New Zealand Government) from Aug 2016 – March/June 2018, possibly till mid 2019

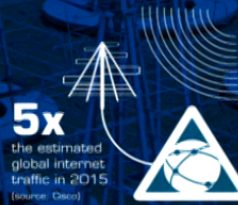
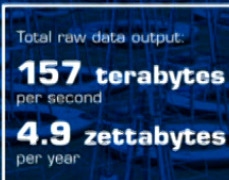
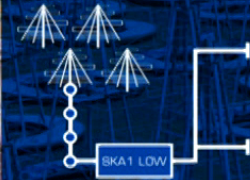
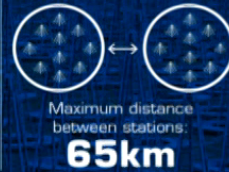
Open Parallel
Innovation Partners

SKA1 LOW - the SKA's low-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.

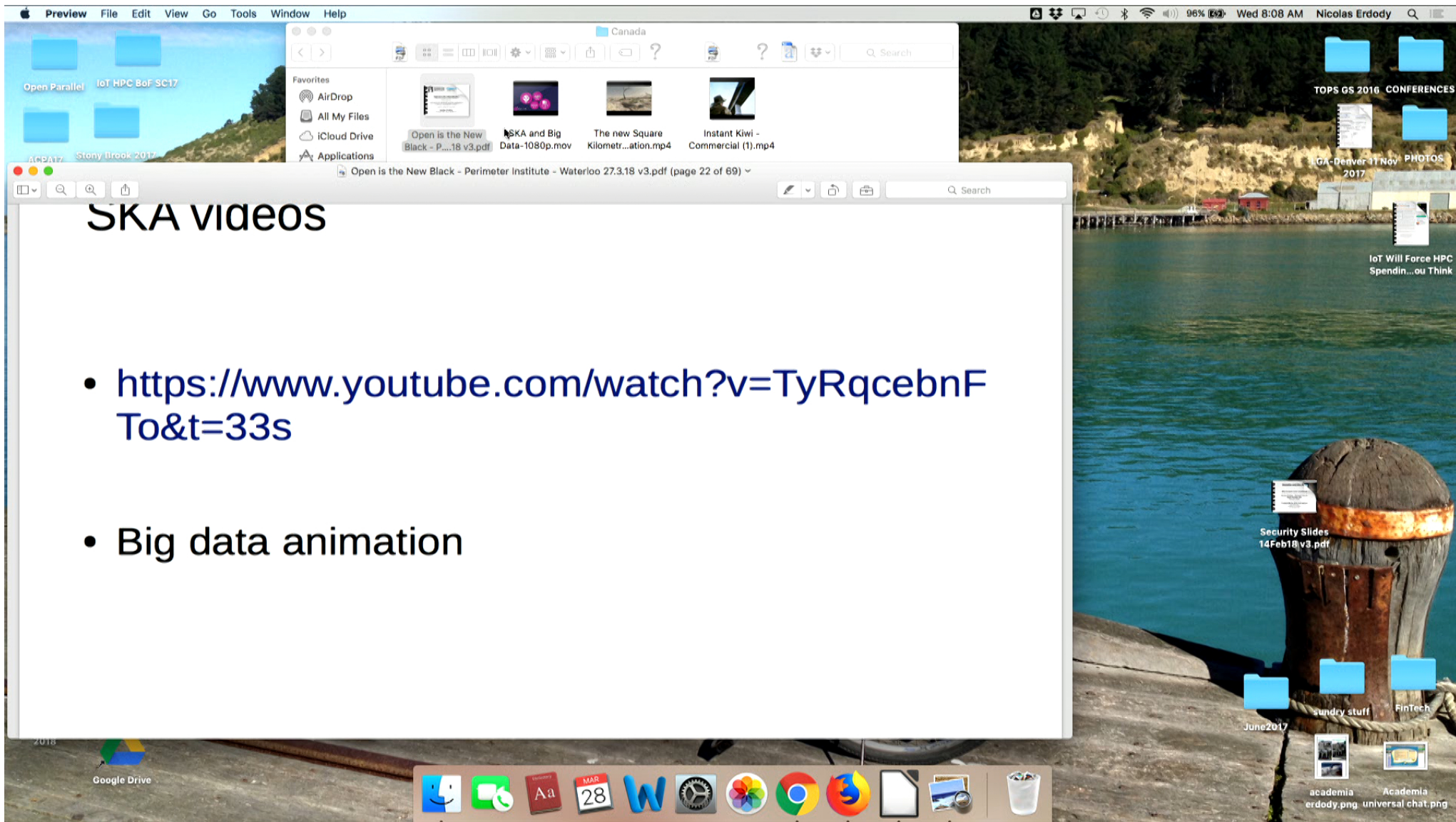


Location: Australia



Compared to LOFAR Netherlands, the current best similar instrument in the world





ARCHIVE SIZE
1 Petabyte (PB)

United States
Census
Bureau

4PB

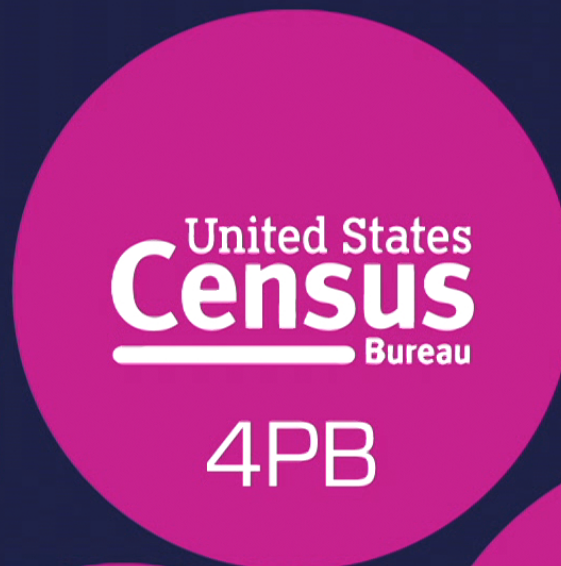
NASDAQ

3PB





6PB



4PB



3PB



LIBRARY OF
CONGRESS

5PB

ARCHIVE SIZE
1 Petabyte (PB)

searches on
Google
98PB

uploads to
facebook.
180PB

You Tube
15PB

CERN
15PB

NORR
6PB

**United States
Census**
4PB

NASDAQ
3PB

**LIBRARY OF
CONGRESS**
5PB

PER YEAR

● 1 Petabyte (PB)

searches on
Google
98PB

uploads to
facebook[®]
180PB

YouTube
15PB

CERN
15PB

NOAA
6PB

**United States
Census**
4PB

NASDAQ
3PB

**LIBRARY OF
CONGRESS**
5PB

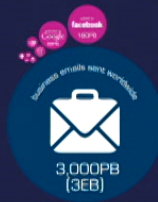
PER YEAR

● 1 Petabyte (PB)



1 Exabyte (EB)

62EB
SKA1 MID raw data



PER YEAR

1 Exabyte (EB)



PER YEAR

- 1 Exabyte (EB)

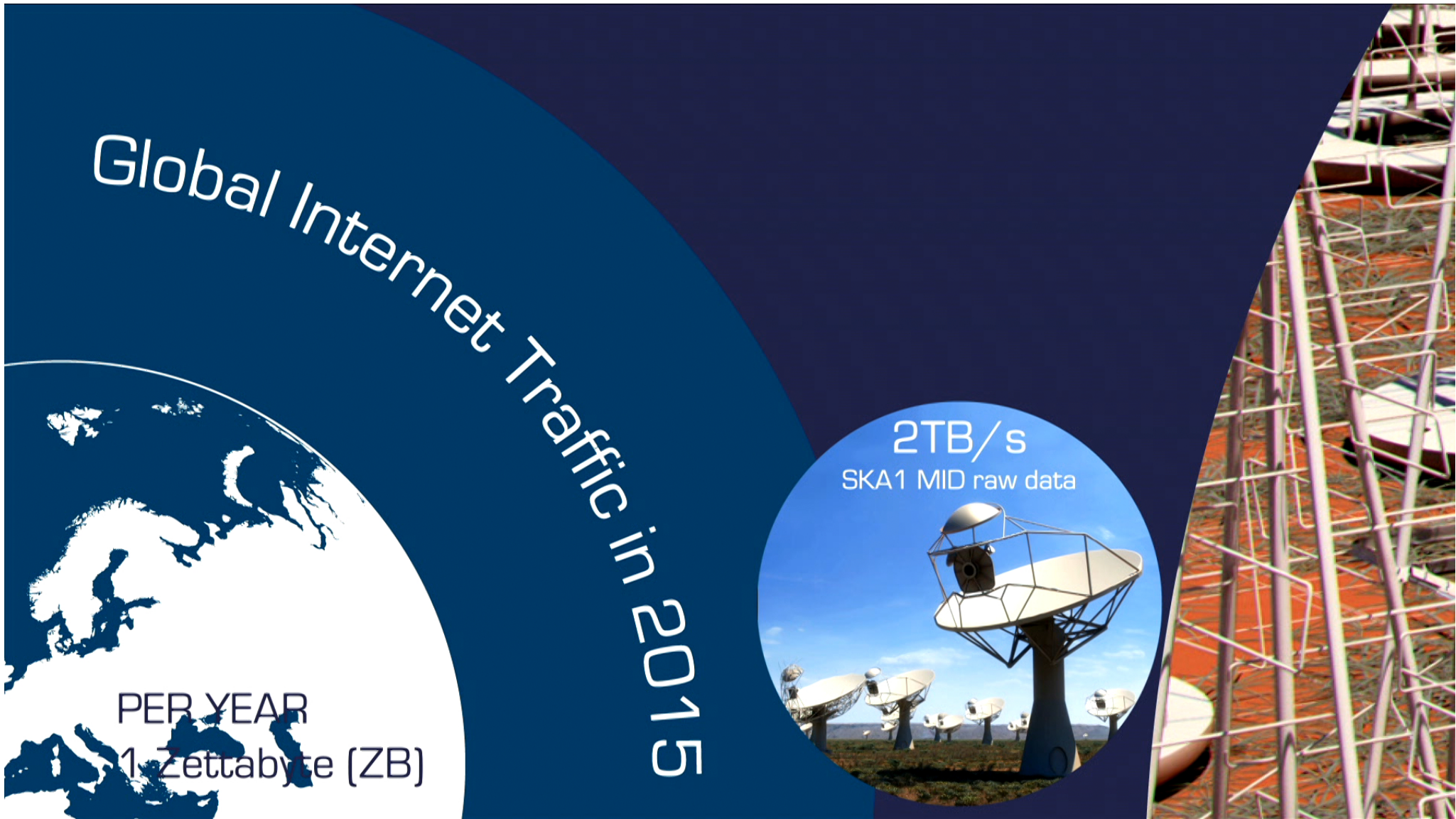


Enough data to fill
4
Laptops



2TB/s
SKA1 MID raw data

IN 0 SECOND





PER YEAR



Why the SKA?

- Besides being the biggest mega-science project in the world, **the SKA is the biggest BIG Data problem in History**
- SKA Phase 1 must grapple with BIG Data on a scale that ICT Industry will be facing in 5-10 years time
- i.e. ASKAP (Australian SKA Pathfinder) is generating 5.2 Tb/second -now
- https://www.theregister.co.uk/2017/01/18/murchison_radiotelescope_opens_the_science_firehose/
- SKA Phase 2 will be an order of magnitude BIGGER yet
- Mega-science projects attract clever people to innovate, give many spin-offs, help steer technology evolution
- High potential for producing technology disruption, new IP
- Large international collaboration, multinational companies keen to be involved

(from Dr. Andrew Ensor, AUT)

Open Parallel
Innovation Partners

Open Parallel's work at the SKA (2013-2018)

Relevant examples:

- **Software Development Environment** for the CSP (Central Signal Processor)
- **Minimalist Operating System (OS)** for the SDP (Science Data Processor)
- **Security at OS level** for the SDP -and the SKA



Is your data safe?

- Attacks on IoT and Big Data repositories are real, and present dangers.

- In general, information security is concerned with safeguarding:

Confidentiality -if data needs to be protected from unwanted viewing.

Integrity – Data can't be altered, or if it is then the data owner is immediately able to determine that it has been altered.

Availability – Data can't be deleted or moved somewhere inaccessible.

Non-repudiation – Parties cannot deny having received / sent a transaction.

- The protection of data in hostile environments – such as the Internet – requires **defence in depth**.
- **Defence in depth ultimately requires protecting the operating systems that applications run on.**
- Have you heard of “Meltdown” and “Spectre”?

Open Parallel
Innovation Partners

Some Issues

- “Attacks” and “accidents” can be similar.
- Accident: you can lose data due to bugs in other people’s code.
- Software security vulnerabilities are often due to errors.
- Encryption on all external traffic is a good idea.
- Sensors are vulnerable to spoofing by transduction attacks.
- “Autonomous systems making safety-critical decisions should remain safe when an adversary can exploit physics to influence the output of sensors” (“Risks of Trusting the Physics of Sensors” – February 2018, CACM)

Open Parallel
Innovation Partners

Security in HPC environments (*)

- Security Policy is a statement of what is, and what is not, allowed.
- Systems are usually very open, used by scientists worldwide whose identities never been validated
- Issues and threats: data leakage and integrity, code or data alteration, misuse of computing cycles, disruption or denial of service
- The Science DMZ security framework: scientific computing systems are moved away from other types of systems (i.e. HR, financial, etc)
- Reduction of Complexity – key benefit to systems robustness, including security

(*) Sean Peisert, CACM, Sept 2017

Open Parallel
Innovation Partners

Towards a Transformational OS (I)

Open Parallel Tasks at the SDP (2017-18):

- Minimalist OS for SDP Compute Node
- Linux performance monitoring systems
- Survey OS candidates for SDP
- **Architecture considerations for security**
- **Security at OS level (security at the kernel)**
- **OS security in HPC / SDP environments**
- OS memory and latency challenges
- Possible algorithm research

What rights do we need?

Some useful rights for scientific software:

- **The Right to Use** -*for a very long time*
- **The Right to Support** -*who will maintain the code?*
- **The Right to Inspect** -*we should be able to monitor, trace and measure the system*
- **The Right to Modify** – *adding new features*
- **The Right to Distribute** -*i.e. modified versions*
- **The Right to Control Rights** -*software purpose-written for your science should have this right*

Open Parallel
Innovation Partners

The OSS ethos

- “Release early, release often”
- “Given enough eyeballs, all bugs are shallow”
(Linus’ law)
- “Why should I care what color the bikeshed is?”
(Law of Triviality) (PHK)

Multicore World 2013

(L-R Paul McKenney, Ian Foster, Poul-Henning Kamp (PHK), Mark Moir)



So...what can you do???

Page 35 of 69

Remember MMM?

- “In most projects, the first system is barely usable. It may be too slow, too big, awkward to use, or all three”
- **You need to start again** and address these problems.
- Should you plan to build a **“throwaway pilot system”** or would you promise to deliver it to customers?
- Delivering the throwaway **buys time**, at the cost of
 - agony for the user
 - distraction for the builders while redesign
 - bad reputation for the product

Conclusion:

“plan to throw one away; you will, anyhow”

Source: **Mythical Man Month**, pg 116, 1995 edition / Fred Brooks (“father of IBM System/360”)

“Bike-shedding”

- Technical disputes over minor, marginal issues conducted while serious ones are being overlooked.
- The implied image is of people arguing for hours (usually in a board decision-making meeting) over what color to paint the bicycle shed or the material of its roof
- Meanwhile the nuclear plant design (or IT project scope / research plan / science budget /...) is approved in 10 minutes. Part of “Parkinson's Law of Triviality”.
- Option: Van Halen's “no brown M&M's clause”
- If any brown M&M's were found backstage, the band could cancel the entire concert at the full expense of the promoter. Try it -as a “canary in the coal mine” technique.

Do try Van Halen's “no brown M&M clause”

Image Credit: <http://bit.ly/29CjUOe>



You need fresh and non-contaminated thinking to tackle next generation challenges:

*"What are those little drawers for? To store tools?"
(17 years old student visiting School of Engineering, Uruguay, June 2017)*



Open Parallel
Innovation Partners

Science and research communities would benefit
by exploring the OSS ethos (on top of actively
using OSS tools).

We don't have to follow Wall Street ideology.



Crowd-source it

- Governance and guidance structure
 - Clear architecture and interface design
 - Modular and distributed
 - Collaborative tool chain
 - High quality testing and integration support
 - Community driven
- ... in short, **open source**

How it works in practice

OpenStack

- 6.8k code submitters, 30k participants, 107k commits/year (Feb, 2018)
- Governance: Board of directors, technical committee, user committee, then sub-projects
- Collaboration: Tools, resources, user groups, conferences, sponsorship, jobs, hall of fame
- Automated integration testing gates

<http://activity.openstack.org/dash/browser/>

