From Domain to Data: a case study or two…

Lesley Wyborn

National Computational Infrastructure ANU

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

Australian Government Bureau of Meteorology
Australian Government Geoscience Australia
Australian Government Australian Research Council

CSIRO
Australian National University

nci.org.au
@NCInews

Who is actually doing this type of work

- Review a survey done in GA in 2013 to evaluate 2 teams involved in eResearch
  - https://eresearchau.files.wordpress.com/2012/08/eresau2012_submission_129.pdf
  - https://eresearchau.files.wordpress.com/2012/11/01_lesley_wyborn.pdf

- Short review of the current NCI team
- General Comments on what determines 'who' is best suited for this work

© National Computational Infrastructure 2017

GA/CSIRO capability analysis of a successful eResearch team

Questions were designed by GA's Human Resources

- Part 1 - Each participant was asked to define what they felt were the core elements of eResearch

- Part 2 - Each participant was asked questions relating to:
  1. Their **Qualifications** and how they have applied them at work (knowledge application)
  2. Their **Job Experiences** (that are directly relevant to eResearch functions/achieving results)
  3. Their **Skills/knowledge** (that they apply in eResearch functions)
  4. The **Behavioral attributes** identified as key for eResearch
  5. The **Organisational Support** they felt essential for eResearch
     - » Not a statistical sample

© National Computational Infrastructure 2017

Both teams defined the core elements of eResearch as:

1. Being directly connected to electronic data and enabling scientists to work directly on the observational data, rather than on syntheses of sub-sampled data
2. Enabling probabilistic analysis with multiple scenarios being investigated and uncertainties being quantified
3. Being characterized by technical innovation and undertaken by those willing to work on cutting edge of problem solving
4. **Requiring computers to enable it and humans to drive it**

© National Computational Infrastructure 2017

GA/CSIRO Survey Part 2: Academic Qualifications

GA Science Team

- **Ranged from BSc to PhD**
- With qualifications in mathematics, applied mathematics, and geophysics with some computational science, modelling and numerical analysis

CSIRO Technical Team

- **Ranged from BSc to PhD**
- With qualifications in computer science, programming, mathematics, physics and geophysics
- That is, most had some scientific qualifications, but a much stronger focus on computer science/software engineering than in the GA science teams

© National Computational Infrastructure 2017

GA Science Team

- Computational programming skills
- Mathematical skills, statistics
- Spatial skills
- Data analysis, data curation and stewardship
- **Transdisciplinary science, thinking big scale**

CSIRO Technical Team

- Information systems design and engineering
- **Geophysics: understanding the underlying problems GA is addressing**
- Ability to weave multiple disparate web services into a coherent application
- Spatial database design and development of spatial information systems
- Engineering, both process and project management
- **Extensive experience in developing eResearch tools, portals and technologies**

© National Computational Infrastructure 2017

GA Science Team

- Intuitive
- Logical
- Non-linear thinker
- Risk taker/willingness to try new things
- **Early adopter**

CSIRO Technical Team

- Analytical skills
- Logical: problem solving
- Emotional intelligence
- Ability to build teams/teamwork
- **All participants listed the ability to communicate and actively listen was the most important attribute for an eResearch expert**

© National Computational Infrastructure 2017

- The identified key enablers by the teams for institutionalising eResearch within the workforce were:
  1. Organizational agility
  2. An approach to fostering early adopters (aka CSIRO)
  3. A specific eResearch enabling team to support the scientists
  4. Recognition of the high skills levels of software developers
  5. **Recognition of the 'hybrids'**
  6. CSIRO quote *"Our team is built on ex-scientists or reformed software engineers (now scientists) who try to bridge that gap each day")*

© National Computational Infrastructure 2017

The team at NCI

Data Service Innovation

Dr Joseph Antony
Senior Data HPC Specialist

Computational Chemistry

Mr Pablo Rozas Larraondo
High Performance Data Analyst

Meteorology

Dr Adam Steer
Earth Systems Data Specialist

LiDAR data

Data Collection Management

Dr Jingbo Wang
Data Collections Manager

Seismology

Dr Kelsey Druken
Earth Systems Data Service Specialist

Geodyanmics

Claire Trenham
Senior Data HPC Specialist

Astronomy

HPC Applications and Data Optimisation

Dr Dale Roberts
Senior HPC Specialist

High Energy Physics

Dr Marshall Ward
Senior HPC Specialist

Computational Oceanography

Dr Rui Yang
Senior HPC Specialist

Computational Chemistry

© National Computational Infrastructure 2017

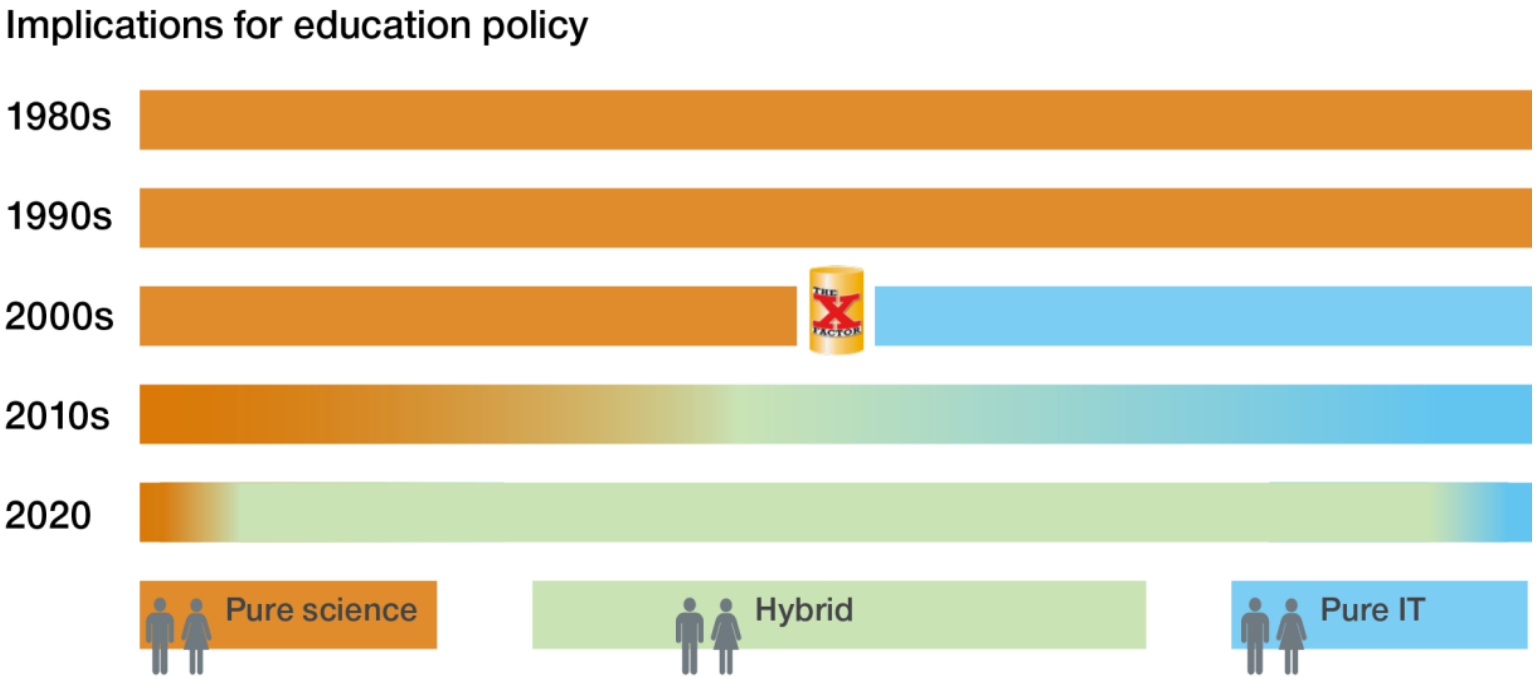
2020 SCIENCE

Microsoft Research: Towards 2020 Science (2005)—External Benchmarking

Summary

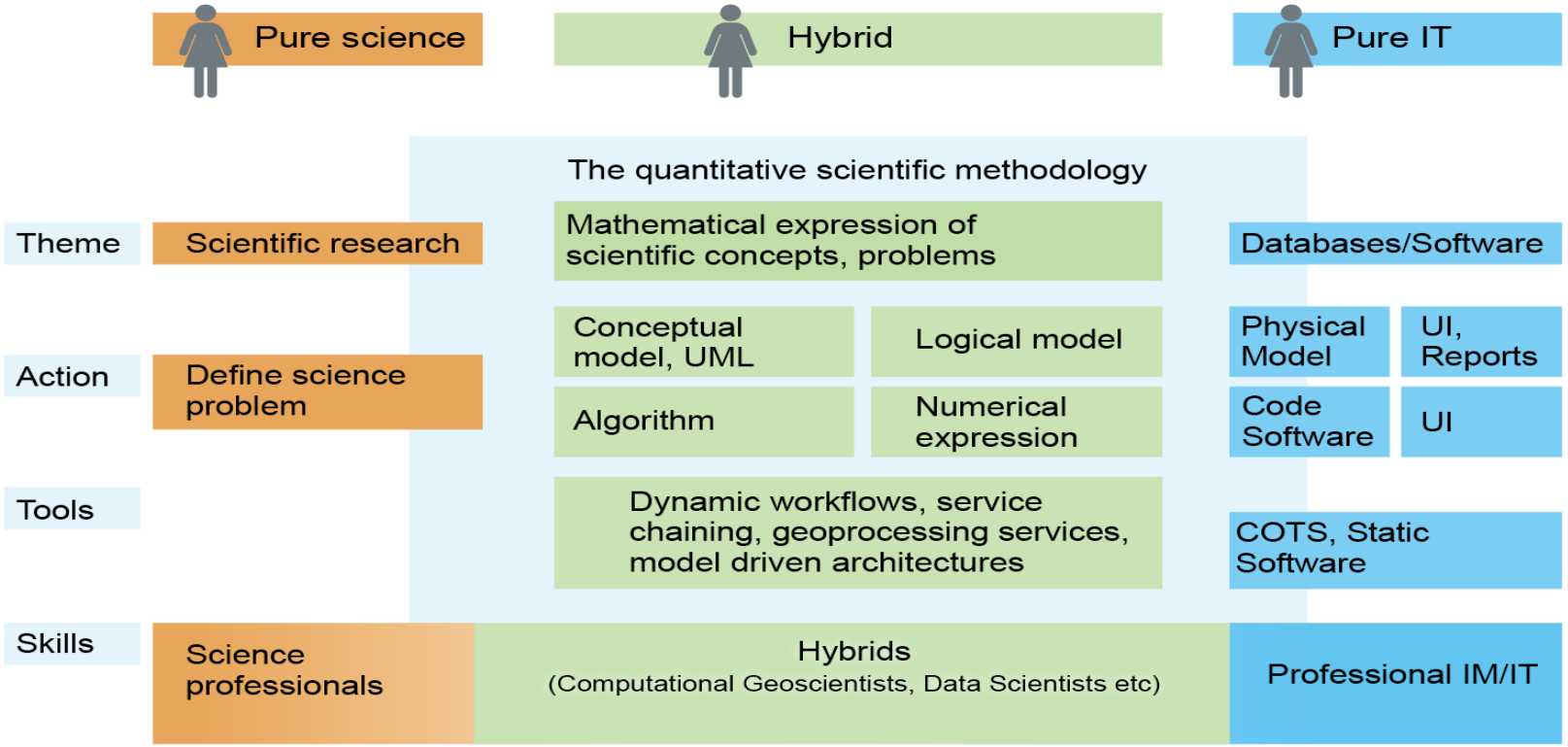
7. Finally, our findings have significant implications for the education of tomorrow's scientists and science policy funding. Scientists will need to be completely computationally and mathematically literate, and by 2020 it will simply not be possible to do science without such literacy.

© National Computational Infrastructure 2017



© National Computational Infrastructure 2017

Graphical summary of it as it is today



© National Computational Infrastructure 2017