Translational Research Computer Science and its Application to Supercomputing

David Abramson
Research Computing Centre
University of Queensland, Brisbane, Australia

Adjunct Professor,
Monash University, Melbourne, Australia

Introduction

- Translational Research
  - Medicine
  - Computer Science
- Exemplars
  - Nimrod
  - Guard
  - Others …
- Important issues for TCS
- Laboratory Scale Matters
- The role of students
- What could be done locally

The role of a French chateaux and red wine … with Manish Parashar

Workshop on CCDSC, Lyon France

Jack Dongarra, Bernard Tourancheau

Background: Translational Medicine

- An “interdisciplinary branch of the biomedical field supported by three main pillars:
  - Benchside, Bedside and Community.
  - Combine disciplines, resources, expertise, and techniques within these pillars to promote enhancements in prevention, diagnosis, and therapies.
- Differs subtly from applied biomedical research, in which a research problem has a potential real-world application (driver).
  - Findings are applied as a specific phase of the research plan.
  - This not only demonstrates applicability and practicality, but also generates tangible outcomes.
- Now well understood and has become a de-facto standard for much of biomedical research.
- Intrinsically helps generate outcomes because the research is applied as part of the original plan, as opposed to being an afterthought once the project has completed

Translational Computer Science

- Research that bridges
  - foundational,
  - use-inspired
  - applied research
    - with the delivery and deployment of its outcomes to a target community.
- Research that supports essential bi-direction interplays where delivery and deployment processes inform the research.

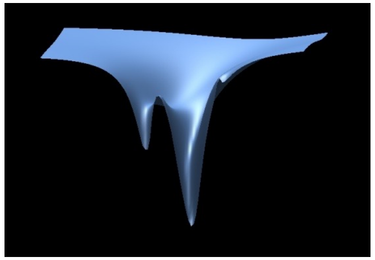
Translational Computer Science

- In TM, translation relies on
  - Taking research from the laboratory Bench to the Bedside
  - More recent refinements involve Community
    - healthy populations, patients and medical practitioners.
- In TCS, translation relies on
  - Taking research from the laboratory Laboratory to the Locale
    - might be physical or virtual
  - Community
    - users and early adopters who work with the technology, and can include public bodies that would help in the evaluation

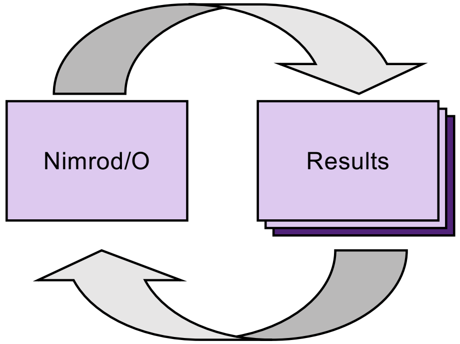
Accidental Translationists …

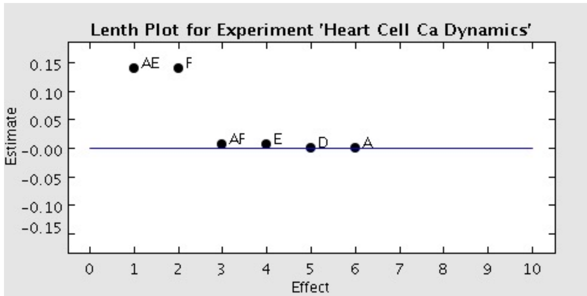
- The term Translational Medicine didn't come into widespread use until the mid 2000s
- The model for translational computer science presented was born from our experiences but we had no road map to follow.
- The projects were arguably successful in performing some degree of translation, but the path was difficult.
- In formalising the model and addressing the roadblocks, we believe that a translational research process in computing should be better defined and supported.
- Once the choice is made that a piece of research is to be translational, the steps to achieve success should be clearer

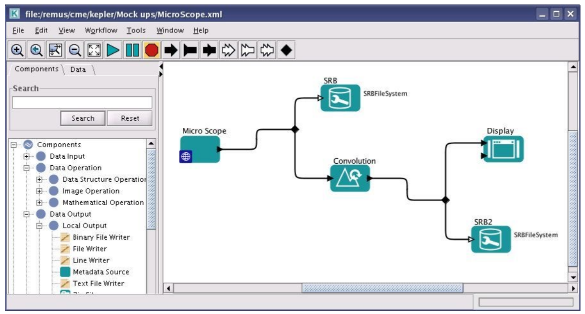
Nimrod supporting “real” science



- Niche distributed programming environment
  - A full parameter sweep is the cross product of all the parameters (Nimrod/G)
  - An optimization run minimizes some output metric and returns parameter combinations that do this (Nimrod/O)
  - Design of experiments limits number of combinations (Nimrod/E)
  - Workflows (Nimrod/K)
- Has survived many distributed computing platforms
  - Workstations, Clusters, Grids, Clouds
- Has contributed to the understanding of HPC and distributed computing



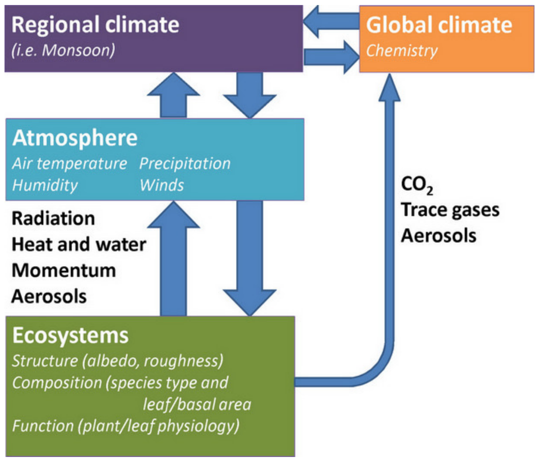


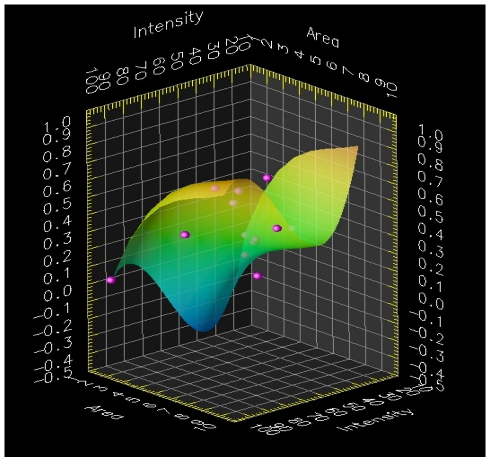


Fire in Australian savannas: from leaf to landscape

Lynch, Beringer, Uotila Monash U, AU

Science outcomes





Global Change Biology

Global Change Biology (2015) **21**, 62–81, doi: 10.1111/gcb.12686

RESEARCH REVIEW

Fire in Australian savannas: from leaf to landscape

JASON BERINGER[1,2], LINDSAY B. HUTLEY[3], DAVID ABRAMSON[1], STEFAN K. ARNDT[4], PETER BRIGGS[5], MILA BRISTOW[3], JOSEP G. CANADELL[5], LUCAS A. CERNUSAK[6], DEREK EAMUS[7], ANDREW C. EDWARDS[3], BRADLEY J. EVANS[8], BENEDIKT FEST[4], KLAUS GOERGEN[9,10,11], SAMANTHA P. GROVER[1,3], JORG HACKER[12], VANESSA HAVERD[5], KASTURI KANNIAH[1,13], STEPHEN J. LIVESLEY[14], AMANDA LYNCH[1,15], STEFAN MAIER[3], CAITLIN MOORE[1], MICHAEL RAUPACH[5], JEREMY RUSSELL-SMITH[3], SIMON SCHEITER[16], NIGEL J. TAPPER[1] and PETTERI UOTILA[17]

[1]School of Earth and Environment, The University of Western Australia, Crawley, WA 6009, Australia, [2]School of Geography and Environmental Science, Monash University, Melbourne, Vic. 3800, Australia, [3]School of Environment, Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT 0909, Australia, [4]Department of Forest and Ecosystem Science, The University of Melbourne, Melbourne, Vic. 3121, Australia, [5]CSIRO Marine and Atmospheric Research, GPO Box 3023, Canberra, ACT 2601, Australia, [6]School of Marine and Tropical Biology, James Cook University, Cairns, Qld 4878, Australia, [7]School of the Environment, University of Technology, Sydney, NSW 2007, Australia, [8]Department of Biological Sciences, Macquarie University, North Ryde, NSW 2113, Australia, [9]Meteorological Institute, University of Bonn, Bonn, D-53121, Germany, [10]Juelich Supercomputing Centre, Research Centre Juelich, Juelich 52425, Germany, [11]Centre for High Performance Scientific Computing in Terrestrial Systems, Research Centre Juelich, Juelich 52425, Germany, [12]Airborne Research Australia/Flinders University, Salisbury South, SA 5106, Australia, [13]Faculty of Geoinformation & Real Estate, Department of Geoinformation, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia, [14]Department of Resource Management and Geography, The University of Melbourne, Melbourne, Vic. 3121, Australia, [15]Department of Geological Sciences, Brown University, Providence, RI 02912, USA, [16]Biodiversity and Climate Research Centre (LOEWE BiK-F), Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325, Frankfurt am Main, Germany, [17]Finnish Meteorological Institute, Helsinki FI-00101, Finland

Abstract

Savanna ecosystems comprise 22% of the global terrestrial surface and 25% of Australia (almost 1.9 million km^2) and provide significant ecosystem services through carbon and water cycles and the maintenance of biodiversity. The current structure, composition and distribution of Australian savannas have coevolved with fire, yet remain driven by the dynamic constraints of their bioclimatic niche. Fire in Australian savannas influences both the biophysical and biogeochemical processes at multiple scales from leaf to landscape. Here, we present the latest emission estimates from Australian savanna biomass burning and their contribution to global greenhouse gas budgets. We then review our understanding of the impacts of fire on ecosystem function and local surface water and heat balances, which in turn influence regional climate. We show how savanna fires are coupled to the global climate through the carbon cycle and fire regimes. We present new research that climate change is likely to alter the structure and function of savannas through shifts in moisture availability and increases in atmospheric carbon dioxide, in turn altering fire regimes with further feedbacks to climate. We explore opportunities to reduce net greenhouse gas emissions from savanna ecosystems through changes in savanna fire management.

Keywords: biomass burning, climate feedbacks, greenhouse gas exchange, net ecosystem carbon balance, savanna

Received 19 December 2013; revised version received 16 April 2014 and accepted 8 June 2014

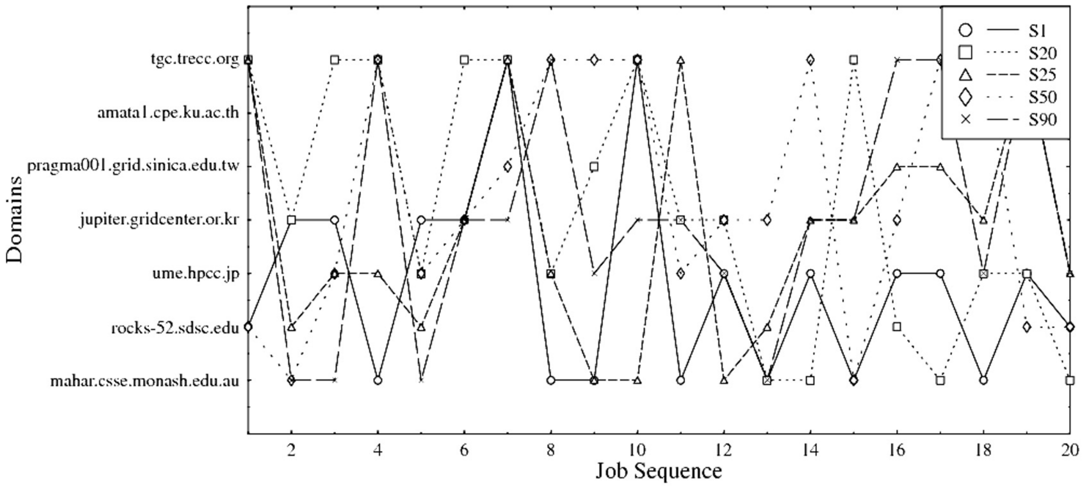
Introduction

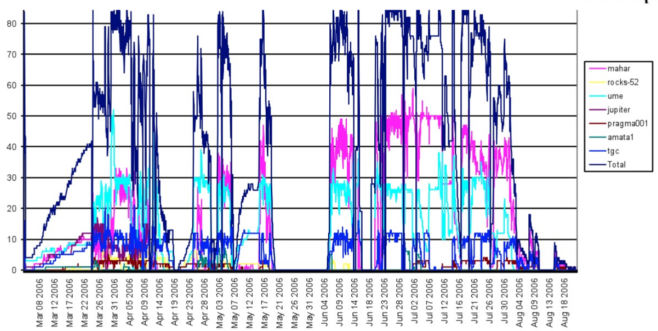
Tropical savanna ecosystems account for around 22% of the global land surface (Ramankutty & Foley, 1999). Annually, up to 75% of global tropical savanna landscapes are burned either by natural or anthropogenic fires (Hao *et al.*, 1990) and accordingly, 50% of the total annual amount of biomass burned globally takes place in the savanna region (Hao & Liu, 1994). The wet-dry tropics of northern Australia include extensive areas of savanna vegetation, which occupy approximately 1.9 million km^2. This area accounts for 12% of the world's tropical savanna ecosystems, making this

Correspondence: Jason Beringer, tel. +61 409355496, fax + 61 86488103, e-mail: jason.beringer@uwa.edu.au

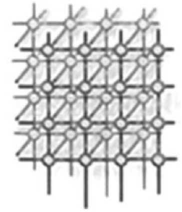
THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

CS outcomes





CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. (2008)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cpe.1353



Fault-tolerant execution of large parameter sweep applications across multiple VOs with storage constraints

Shahaan Ayyub*, †, David Abramson, Colin Enticott, Slavisa Garic and Jefferson Tan

Faculty of Information Technology, Monash University, Melbourne, Vic. 3145, Australia

SUMMARY

Applications that span multiple virtual organizations (VOs) are of great interest to the e-science community. However, our recent attempts to execute large-scale parameter sweep applications (PSAs) for real-world climate studies with the Nimrod/G tool have exposed problems in the areas of fault tolerance, data storage and trust management. In response, we have implemented a task-splitting approach that facilitates breaking up large PSAs into a sequence of dependent subtasks, improving fault tolerance; provides a garbage collection technique that deletes unnecessary data; and employs a trust delegation technique that facilitates flexible third party data transfers across different VOs. Copyright © 2008 John Wiley & Sons, Ltd.

Received 11 December 2007; Revised 30 March 2008; Accepted 7 May 2008

KEY WORDS: e-science; parameter sweep applications; Grid

1. INTRODUCTION

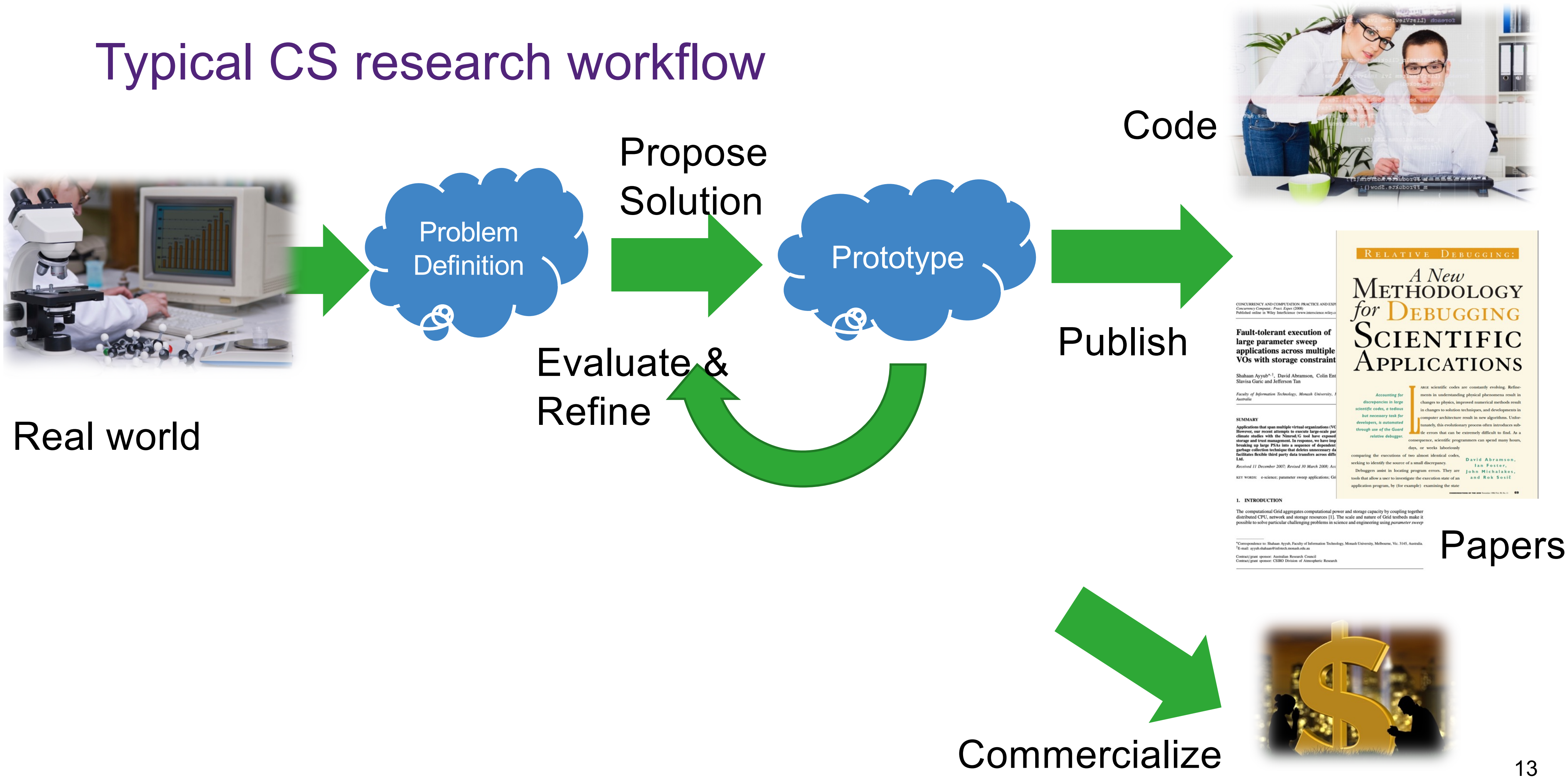
The computational Grid aggregates computational power and storage capacity by coupling together distributed CPU, network and storage resources [1]. The scale and nature of Grid testbeds make it possible to solve particular challenging problems in science and engineering using *parameter sweep*

*Correspondence to: Shahaan Ayyub, Faculty of Information Technology, Monash University, Melbourne, Vic. 3145, Australia.
†E-mail: ayyub.shahaan@infotech.monash.edu.au

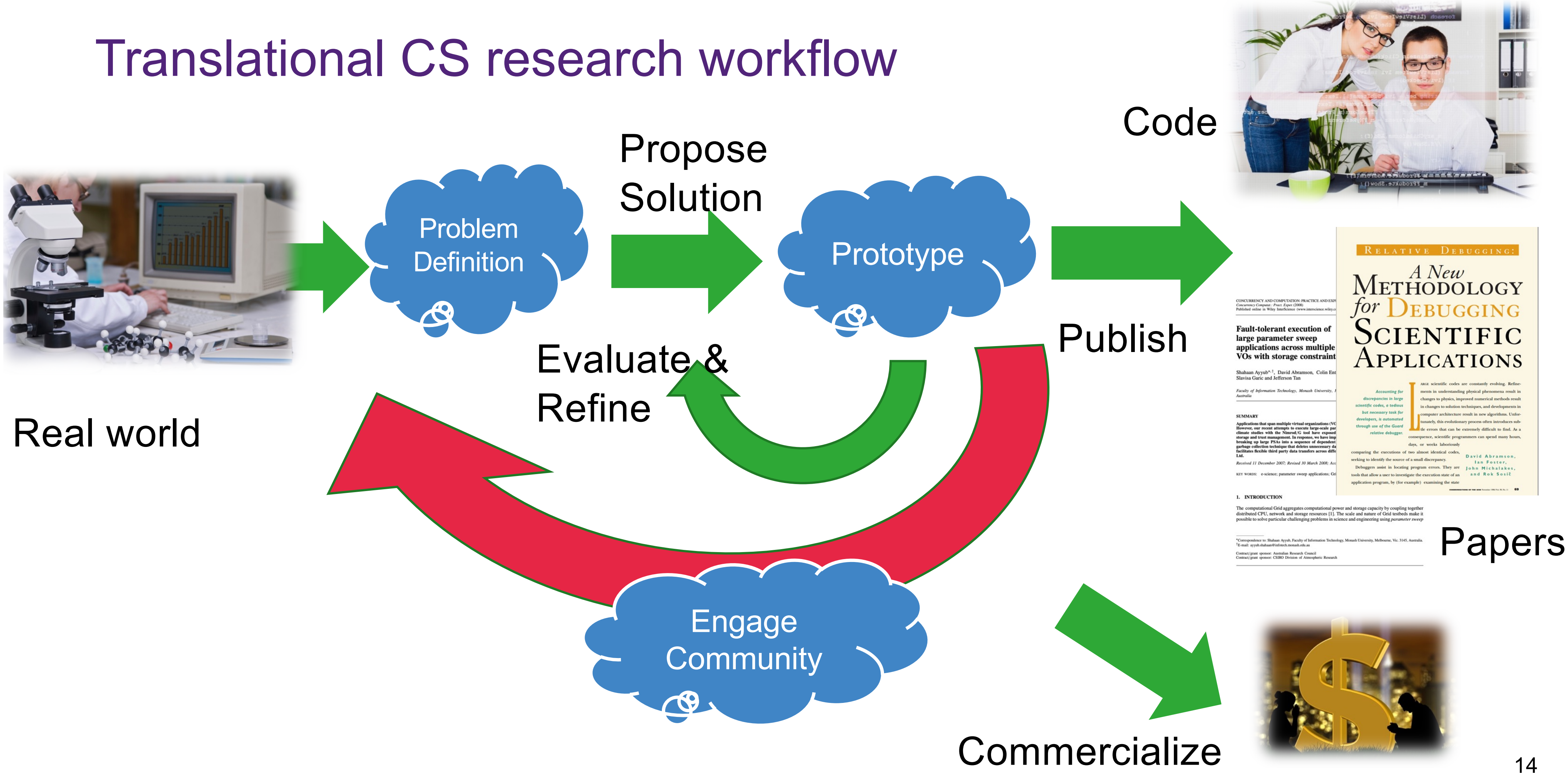
Contract/grant sponsor: Australian Research Council
Contract/grant sponsor: CSIRO Division of Atmospheric Research

So how does this differ from traditional research pipeline?

Typical CS research workflow



Translational CS research workflow



Roadblocks

Roadblocks

1. In computer science, translation is often confused with commercialization
2. Open source techniques are often confused for translation
3. Funding agencies typically don’t provide support for translation
4. PhD programs don’t allocate time and resources to translation
5. Traditional publication venues don’t value translation
6. There are a lack of exemplars

Translation is not commercialization

- Commercialisation almost always occurs after the research has been completed,
  - almost never funded as part of the original research proposal.
- Commercialisation implies a financial angle that has little to do with the research per-se.

Use of Open Source

- Helps with distribution of a software system, but doesn’t intrinsically drive impact
- No direct link between the way the software is used, and the research program. Thus, there is no explicit feedback from lessons learned in the adoption into the research itself.
- More focussed on producing software that is maintained in a sustainable way, by building a distributed workforce.

Funding bodies don’t typically support translation

- Evaluation criteria typically focus on the quality of the investigator team, the project quality and innovation, the feasibility and the benefit.
- Translation is not usually highlighted as a desirable property, thus a proposal might be marked down for including translational activities.
- A budget that allocates resources to items such as a community trial, software distribution, software maintenance, may be pruned back to the basic research program.

PhD timelines don't support translation

- Typical PhD projects in computer science follow a very standard and often rigid template.
  - Students engage in a project of interest to them
  - Execute a plan much like any other research project.
  - Milestones and deliverables include software prototypes, experiments and tests, producing publication outputs along with possibly software and data artefacts.
  - At the very least, a PhD student needs to produce a thesis.
- TR adds complexity by requiring a translation phase,
  - might extend the timeline beyond that of current PhD programs.

Traditional publication venues don’t value translation

- Many editorial boards would argue translation is secondary to their scope,
- More focussed on primary research outcomes in computer science
- Many translational research projects are interdisciplinary,
  - Outcomes might not align well with the journal’s primary focus.
- Most journals do not publish failures.

The University of Queensland
AUSTRALIA

Lack of exemplars

- Numerous examples of computer science research being commercialised and adopted
- Few examples of successful translational research projects
- Changing the culture in an organisation is difficult because people don’t know what a good TCS project looks like.

Funding

- Currently ad hoc funding.
- Sustained funding programs and mechanisms focussed on fostering and nurturing TCS
- More money alone not solution
- Need to *build* translation into the research plan
- Funding must be used to experiment with solutions and prototypes.
- TCS typically involves substantial interaction with end users.
- Additional travel, user engagement, and provisioning of computing resources
- Translation process feeds back into the research,
  - may be a loop of research and translation rather than a linear waterfall style of workflow
- Should be free to report on both research successes, but also translation **success** or **failures**.

Venues, metrics and reward structures

- Traditional publications are not well suited to TCS.
  - Drawing on TM, new journals have been created that explicitly target translational medicine.
  - New set of similarly scoped journals and conferences.
  - Metrics, recognitions and rewards structures, especially in the academic community.
    - software and data and to track their use, citations and impact are a step in the right direction
  - Metrics that report uptake of their work, and measure how many of these have resulted in successful translation
  - Integrate metrics into promotion processes

Education and Training

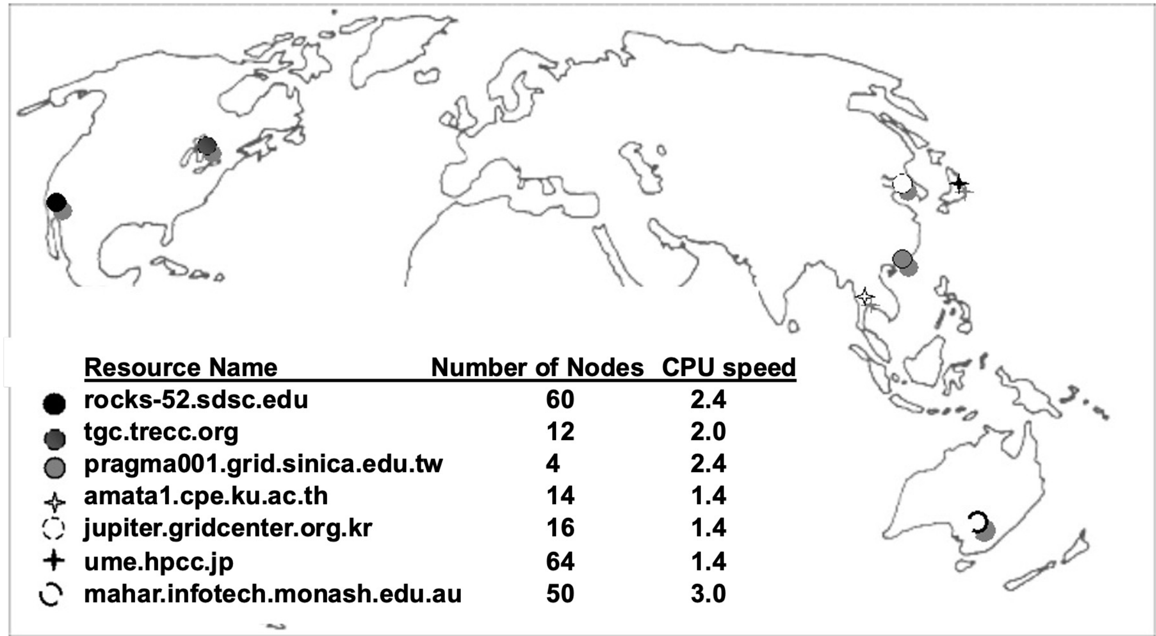
- Integration of translational approaches and methodologies into more formal computer science curricula
- New materials and mechanisms for providing translational skills to practitioners, in both computer and other disciplines.
  - doctoral training centers have been established that encourage and enable trans-disciplinary research
- Extreme example, a PhD could be entirely devoted to the translation of work performed by another researcher, with no original research on the background IP per se

Overpasses

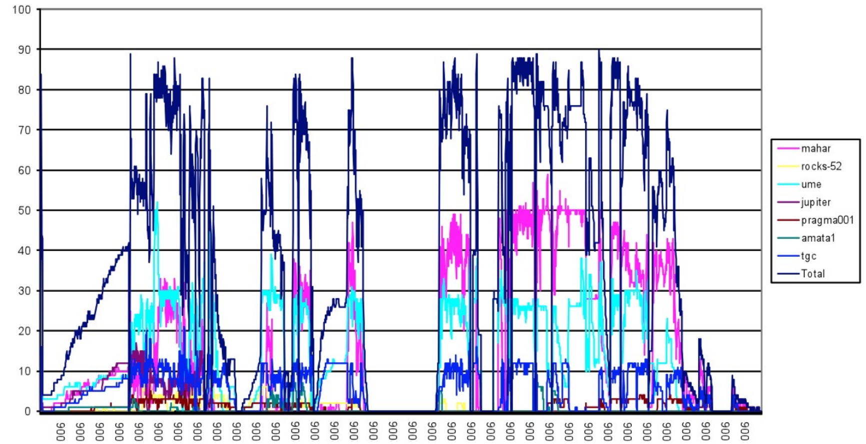
Laboratory Scale Matters

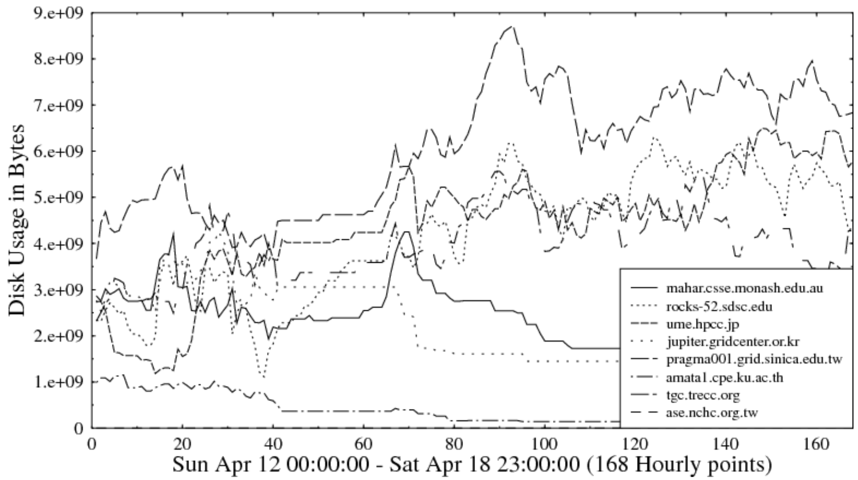
PRAGMA

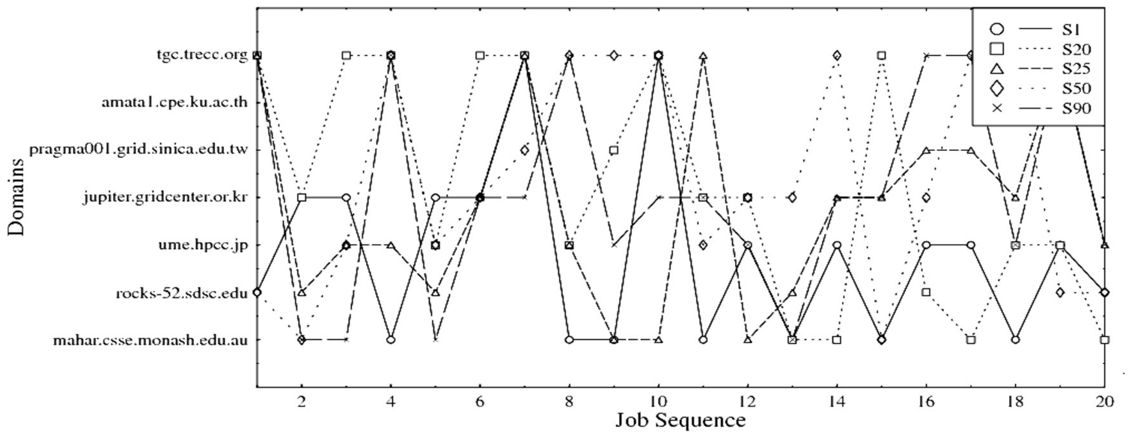
- Strengthen Existing and Establish New Collaborations
- Work with Science Teams to Advance Grid Technologies and Improve the Underlying Infrastructure
- In the Pacific Rim and Globally



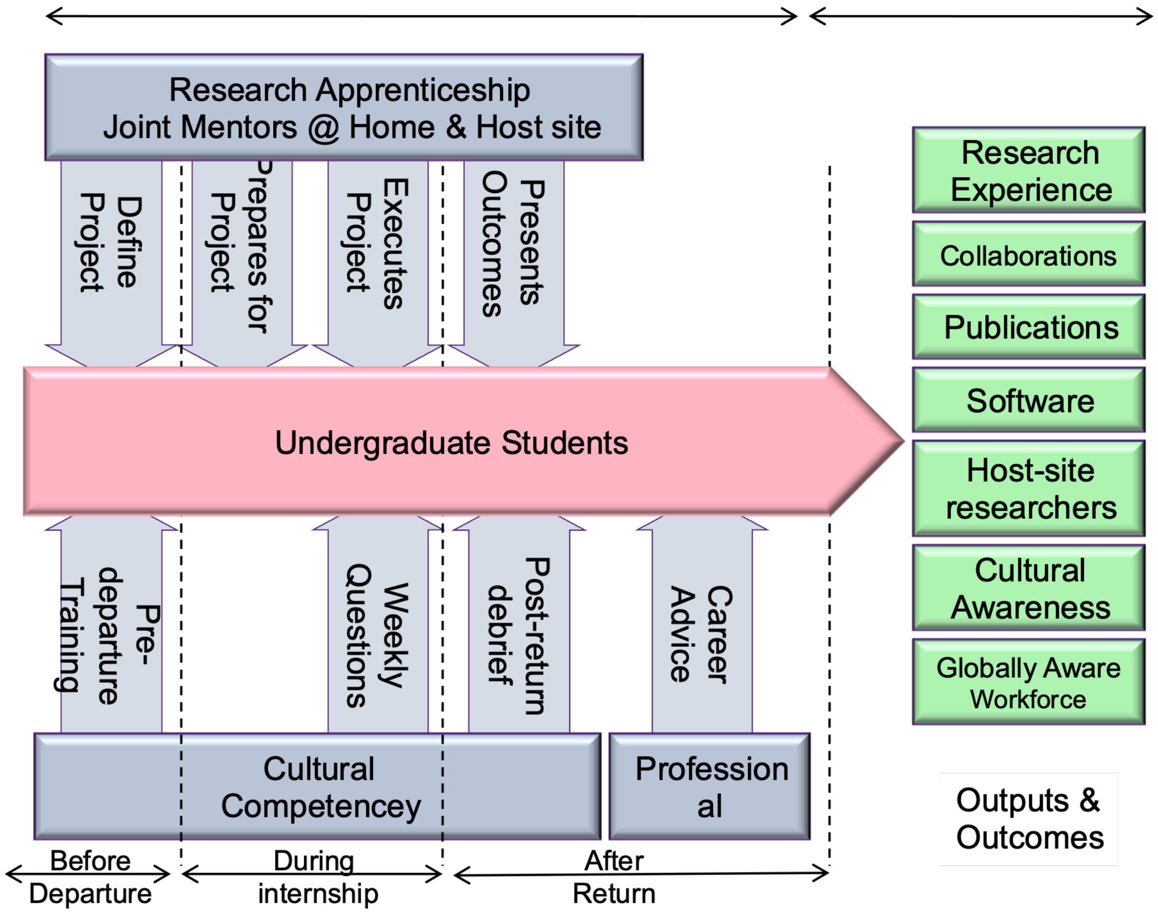
Leveraging the PRAGMA testbed: Technical

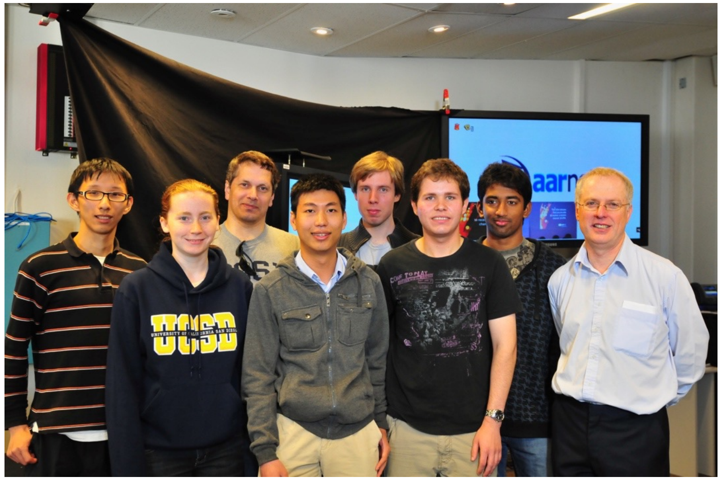


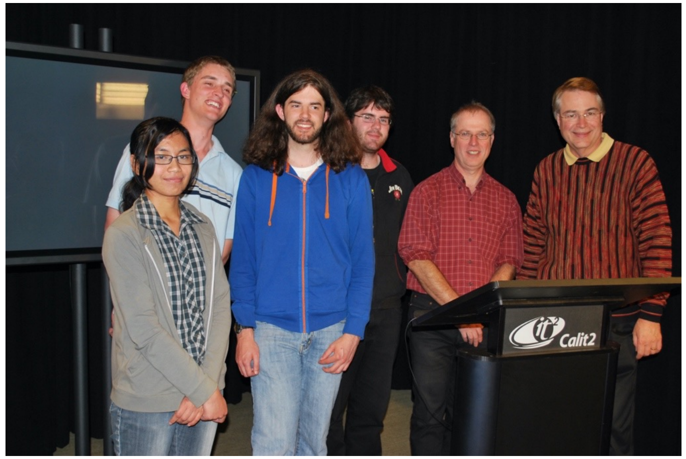




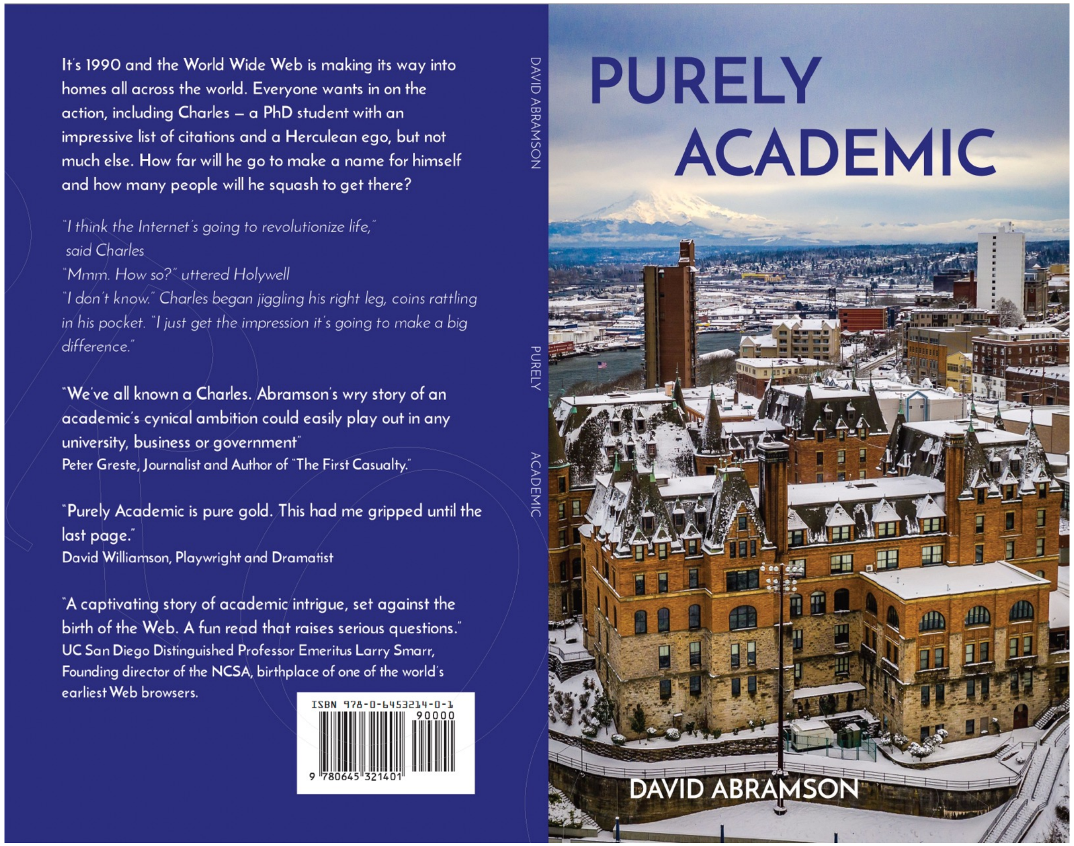
Leveraging the PRAGMA testbed: Social







There’s often a back story …



purely-academic.com

Thank you and Questions