Australian Government
Department of Innovation, Industry, Science and Research

eResearch

from research by the knowledge poor
to research by the knowledge rich

International Association of Technological University Libraries
29th Annual Conference: Digital Discovery: Strategies & Solutions
Auckland New Zealand, 21-24 April 2008

Rhys Francis
Executive Director
Australian eResearch Infrastructure Council

Outline

- eResearch Development
  - The Australian experience
  - The trajectory we are on
- Global Trends
  - Communication, Information, Technology, Population
- The human limitations
  - And the various ways out
- The eResearch agenda

NCRIS: a National Collaborative Research Infrastructure Strategy

- NCRIS brings a strategic approach to Australia's investment in research infrastructure
- It has developed priority areas for support with the research community - published as the "NCRIS Roadmap"
- Each area aims to support research excellence, promote collaboration, provide national benefit
- NCRIS is a response to
  - Recognition of the need for a 'small' country to be strategic
  - Increasing importance of major research infrastructure to research and innovation
  - Increasing need for collaborative research
  - Increasing cost and complexity of research infrastructure

Overall Investment Pattern

$540M over the five years: 2007-2011

- Evolving bio-molecular platforms and informatics
- Integrated biological systems
- Characterisation
- Fabrication
- Biotechnology products
- Optical and radio astronomy
- Integrated marine capability
- Structure and evolution of the Australian continent

- Networked biosecurity framework
- Population health and clinical data linkage
- Terrestrial ecosystem research network

Platforms for Collaboration (allocated $75M -> $82M)

- Data access and discovery, storage and management
- Grid enabled technologies and infrastructures
- Technical expertise
- High performance computing
- High capacity communication networks

- Controlled sharing of researcher identities
- Researchers and resource owners should control who can do what to their resources
- Expanding the e-Research community to “non builder” users
- Engaging the broader NCRIS community

Building on a decade of investment

1997: High Performance Computing Committee

- Established the Australian Partnership for Advanced Computing to provide access to high performance computing capability

2000: Advanced Networks Programme

- Established advanced demonstrator networks

2002: Higher Education Bandwidth Advisory Committee

- Established the Australian Research and Education Network Advisory Committee, and the Australian Research and Education Network

2004: Research Infrastructure Taskforce Report

- Established the National Collaborative Research Infrastructure Strategy Committee to implement a program of strategic investment in research infrastructure

2006: eResearch Coordinating Committee Report

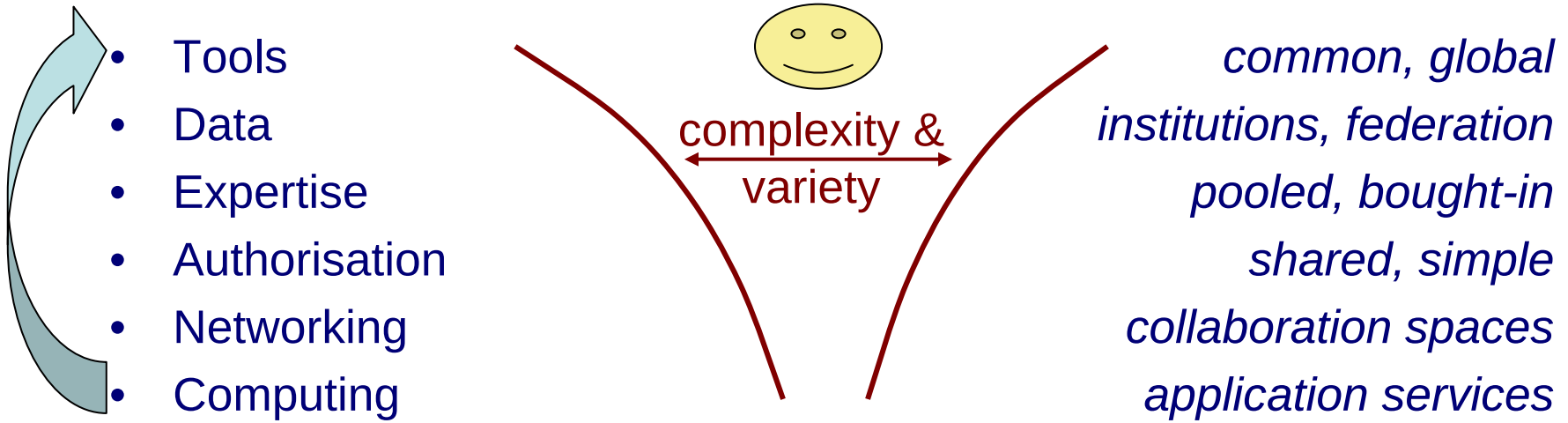
- Outlined an integrated program of skills development and of middleware and computer science research

2007: NCRIS Platforms for Collaboration

- Commitment to an infrastructure program covering computing, data and inter-operation components, and supporting the development of the Australian Access Federation

What were the priority issues…

- Data - just how to cope — *Who owns the issue*
- Tools - the wild west of eResearch — *How does oz play*
- Authorisation - somebody should do it — *Go it alone mentality*
- Computing - isn't that free — *Strategy for moore's law*
- Networking - security problem — *Expected to be a commodity*
- Expertise - just make-do — *Everyone wants it in-house*



- Tools — *common, global*
- Data — *institutions, federation*
- Expertise — *pooled, bought-in*
- Authorisation — *shared, simple*
- Networking — *collaboration spaces*
- Computing — *application services*

eResearch Strategy (circa 2006)

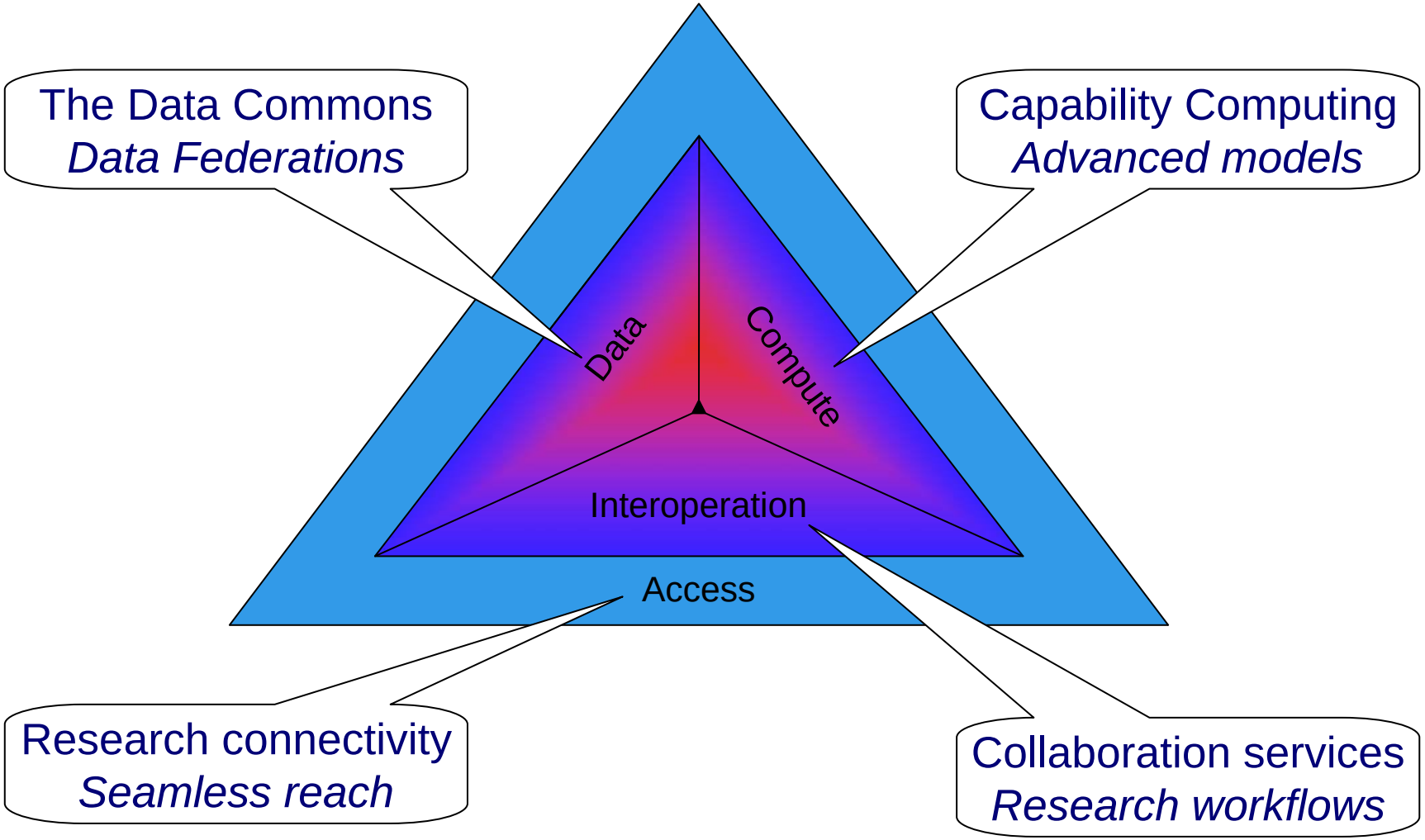
Thematic Issues

- Continuing Need for a Focus
  - through national coordination
- Human Capabilities
  - People, skills and understanding
- Linkage of eResearch Resources
  - seamless access to resources
- Access to Data
  - best practice data management and curation
- Structural and Cultural Change
  - evolution of organisational structures and cultures
- Awareness and Support
  - develop researchers' ability to adopt eResearch

Service Clusters

- Data
  - outreach, curation, data management
  - meta-services, location, access, movement
  - practice, providers and users
- Computing
  - capability computing facilities
  - national computing environment
- Interoperation
  - discipline services (tools ((software))
  - user and operations support
  - collaboration services support
- Access
  - the Australian access federation
  - the Australian research and education network

Platforms for Collaboration (circa 2007)



Investments

Australia National Data Service — *Data Federations*

- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers share data through the data commons

National Computational Infrastructure — *Advanced Models*

- Develop and operate a shared national computational facility
- Develop domain oriented advanced modelling capabilities

Australian Research Collaboration Service — *Research Workflows*

- Develop and operate services linking systems and resources nation wide
- Develop and operate collaboration and workflow tools for researchers

AREN and AAF — *Seamless Reach*

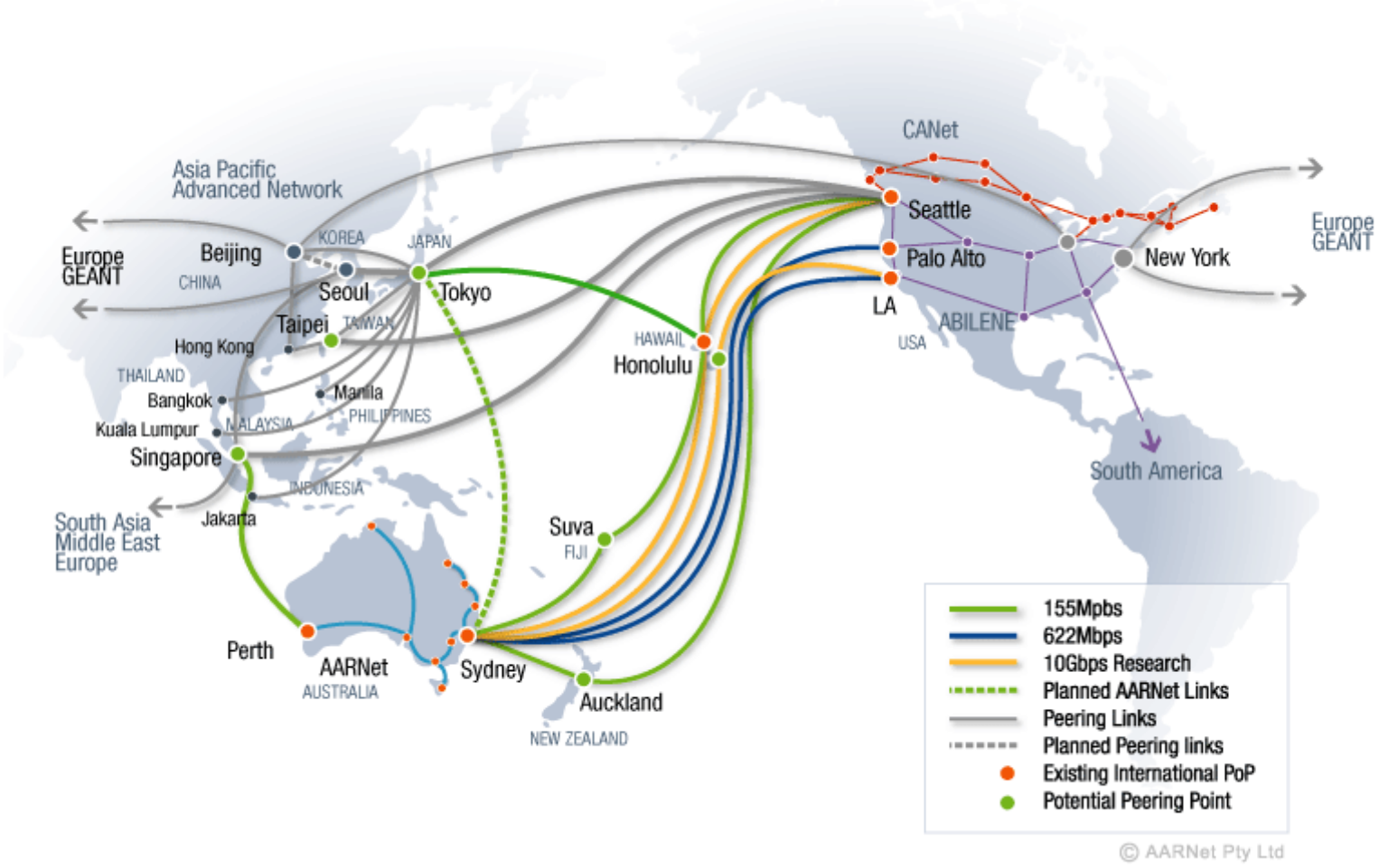
- Connect researchers and research resources at required bandwidth
- Develop a shared authorisation framework for access to research resources

Revising Platforms (circa 2008)

- Access frameworks and access enabling services
  - National rules and agreements to simplify sharing
- Collaboration support services
  - Applying ‘social technologies’ to research
- Data capture, management and curation services
  - Managing research data output / providing data access as a research input
- Modelling and computational analysis services
  - Modelling capability where we need it, massed where the data is
- Connectivity services, backbone and end-to-end
  - Person to person / person to resource / resource to resource
- Discipline development and support services
  - Resources to build new services for research

Seamless Reach: AREN

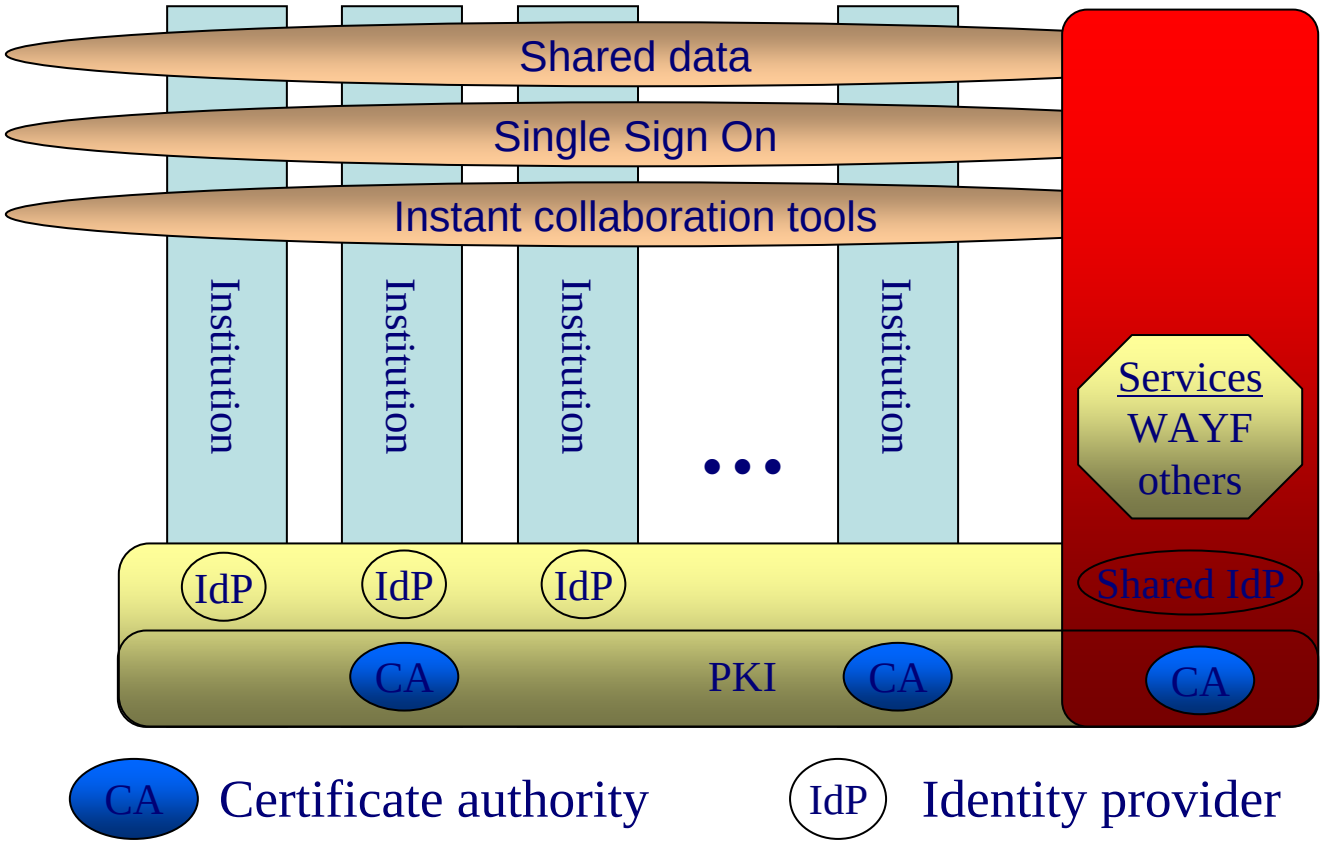
- Connect researchers and research resources at required bandwidth
- Develop a shared authorisation framework for access to research resources



Seamless Reach: Australian Access Federation

- Connect researchers and research resources at required bandwidth
- Develop a shared authorisation framework for access to research resources

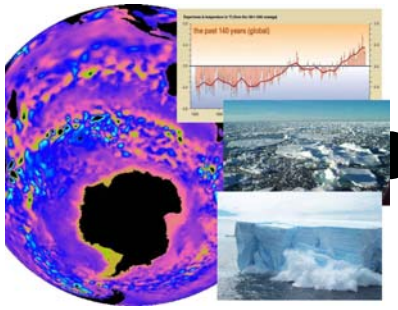
Trust Federation



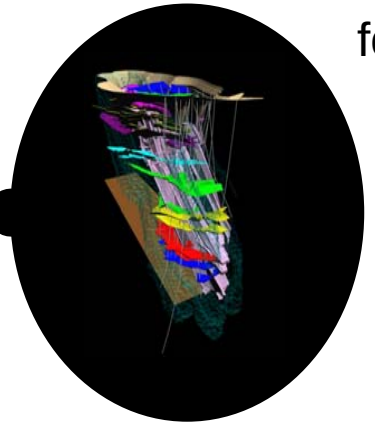
Advanced Models: National Computational Infrastructure

- Develop and operate a shared national computational facility
- Develop domain oriented advanced modelling capabilities

Ecosystem focus



Geoscience focus



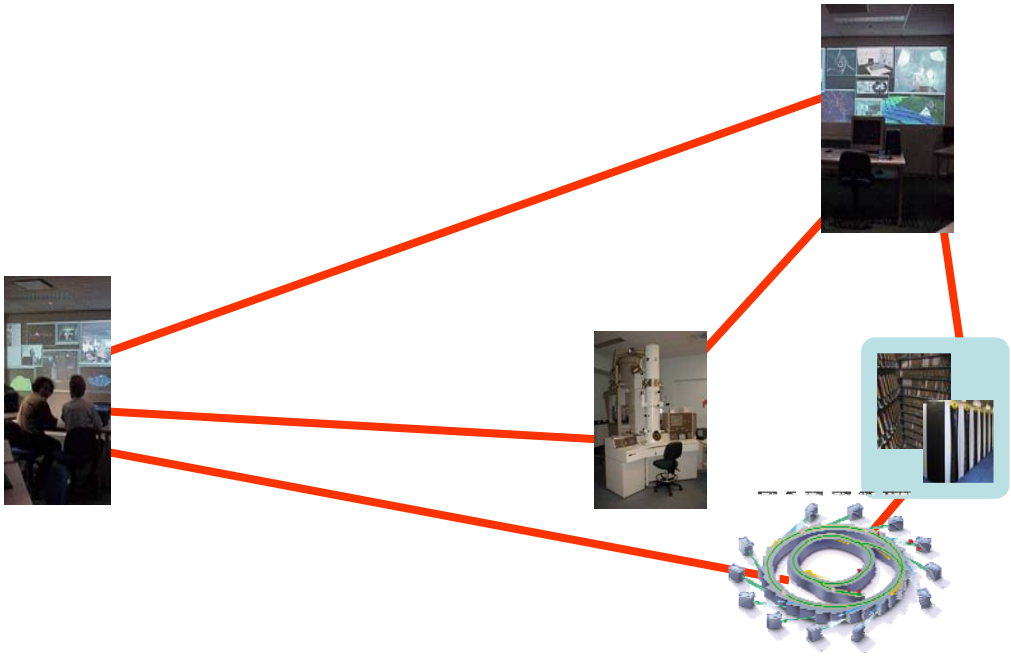
NCI Peak Facility
- managed by ANU-SF
- national merit access

Biotech focus

?

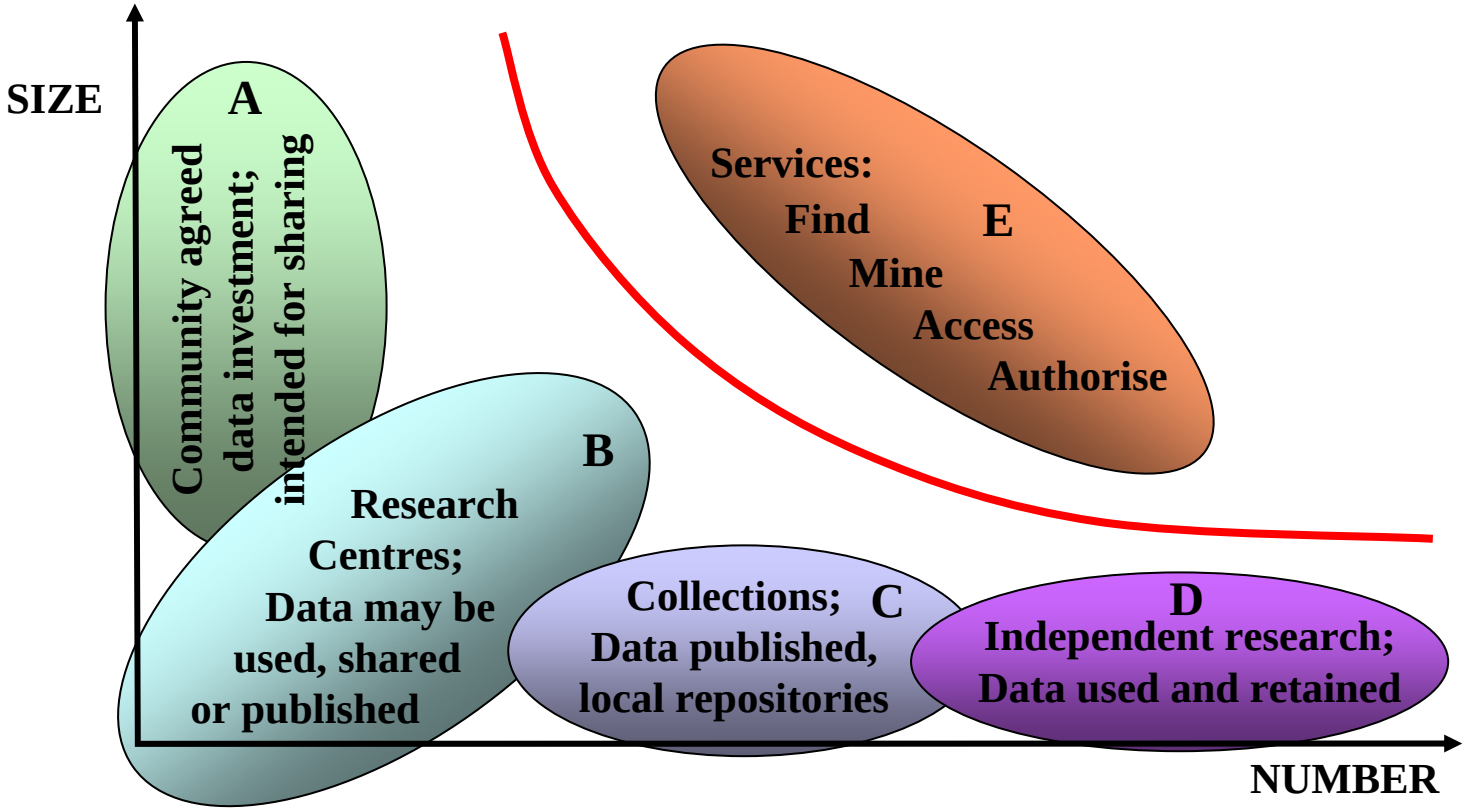
Research Workflows: Australian Research Collaboration Service

- Develop and operate services linking systems and resources nation wide
- Develop and operate collaboration and workflow tools for researchers



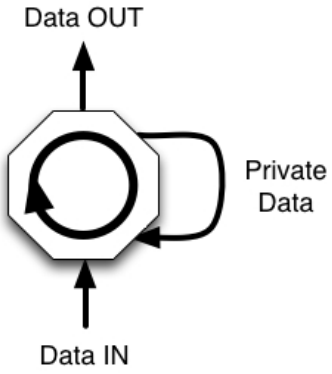
Data Federations: Australian National Data Service

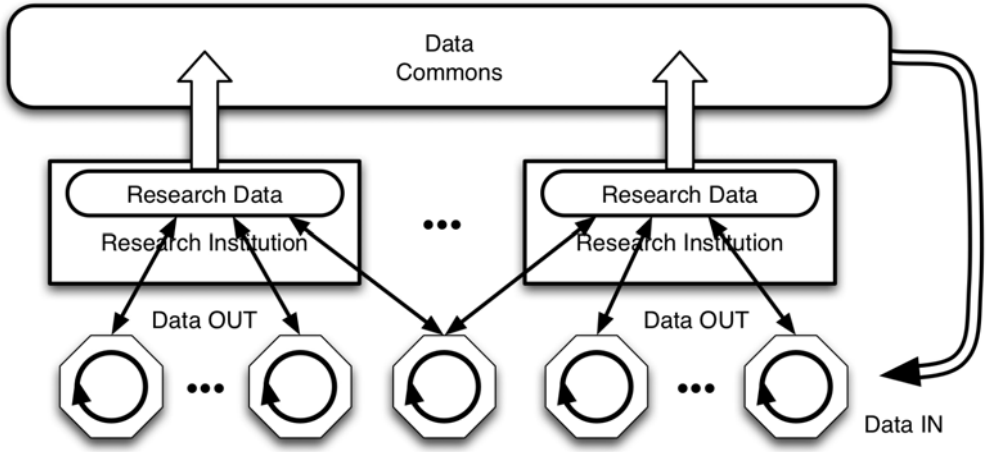
- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers publish and share data through the data commons

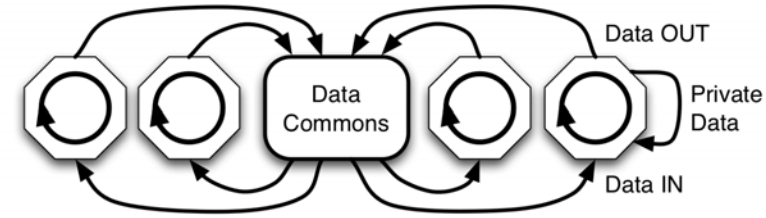


Data Federations: Australian National Data Service

- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers publish and share data through the data commons

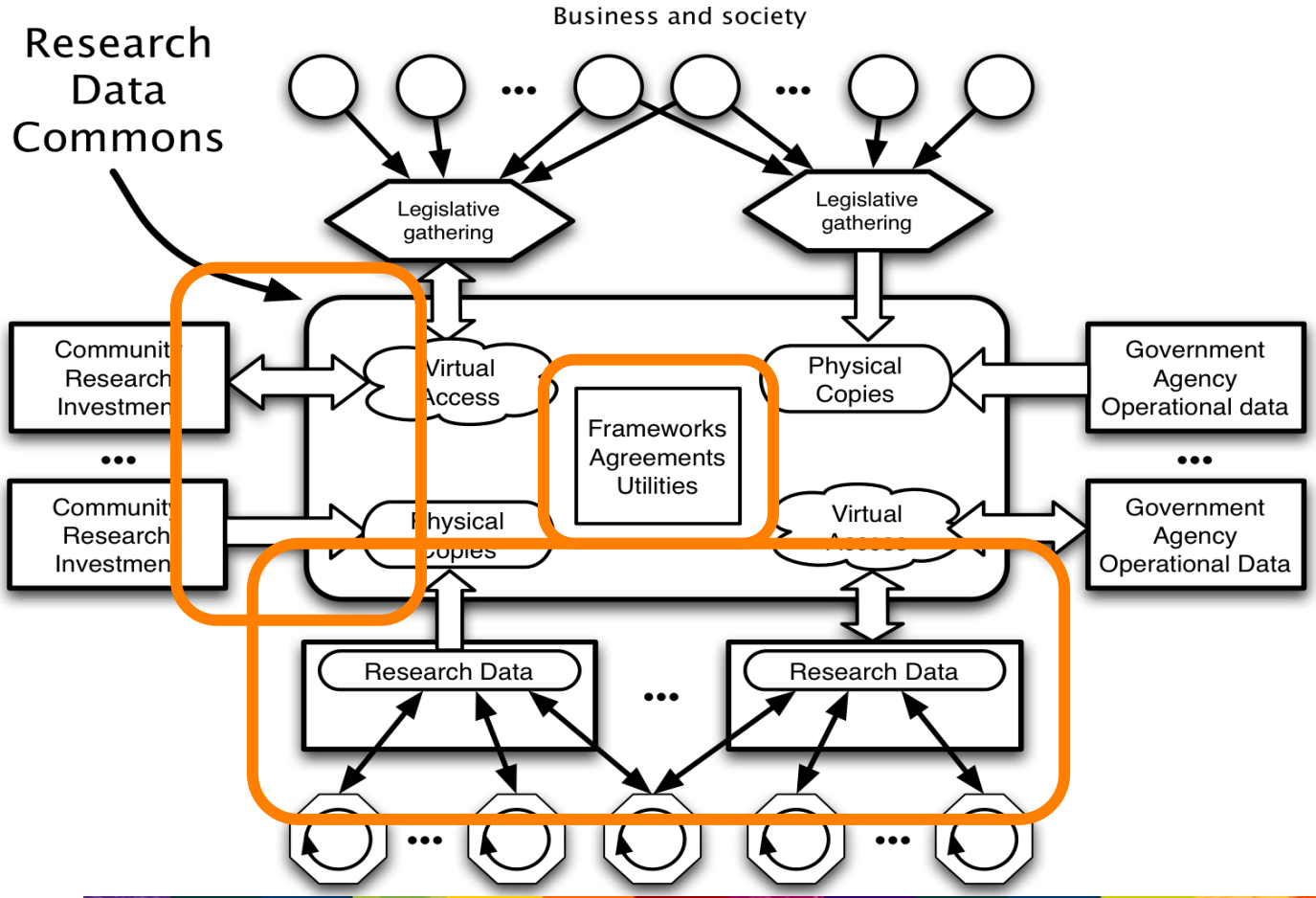






Data Federations: Australian National Data Service

- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers publish and share data through the data commons



Why Does ICT matter?

Our society was shaped by a truth: only people process information

For most of history it was IMPOSSIBLE to speak to anybody at a distance

- From 1870 to 1990: ICT delivered that ability to 8% of the world's population
- From 1990 to 2010: ICT will grow that to 50% of the world's population

ICT means that voice and images can be sent from anyone in the world to anyone else in the world, anywhere, at a moments notice.

In the past, no-one could publish information and have it accessible to everyone

- This possibility is now available to more than 15% of the world's population
- The web shows the highest growth rate of any technology adoption ever

Until now, it was only possible to socialise, work or play with neighbours

- ICT is drawing this "tyranny of distance" to a close
- New collaboration, immersion and remote presence technologies stand poised to redefine human interaction

Why Does ICT matter?

Our society was shaped by a truth: only people process information

- communication was between people
- everything made was hand made
- all decisions were a product of mind

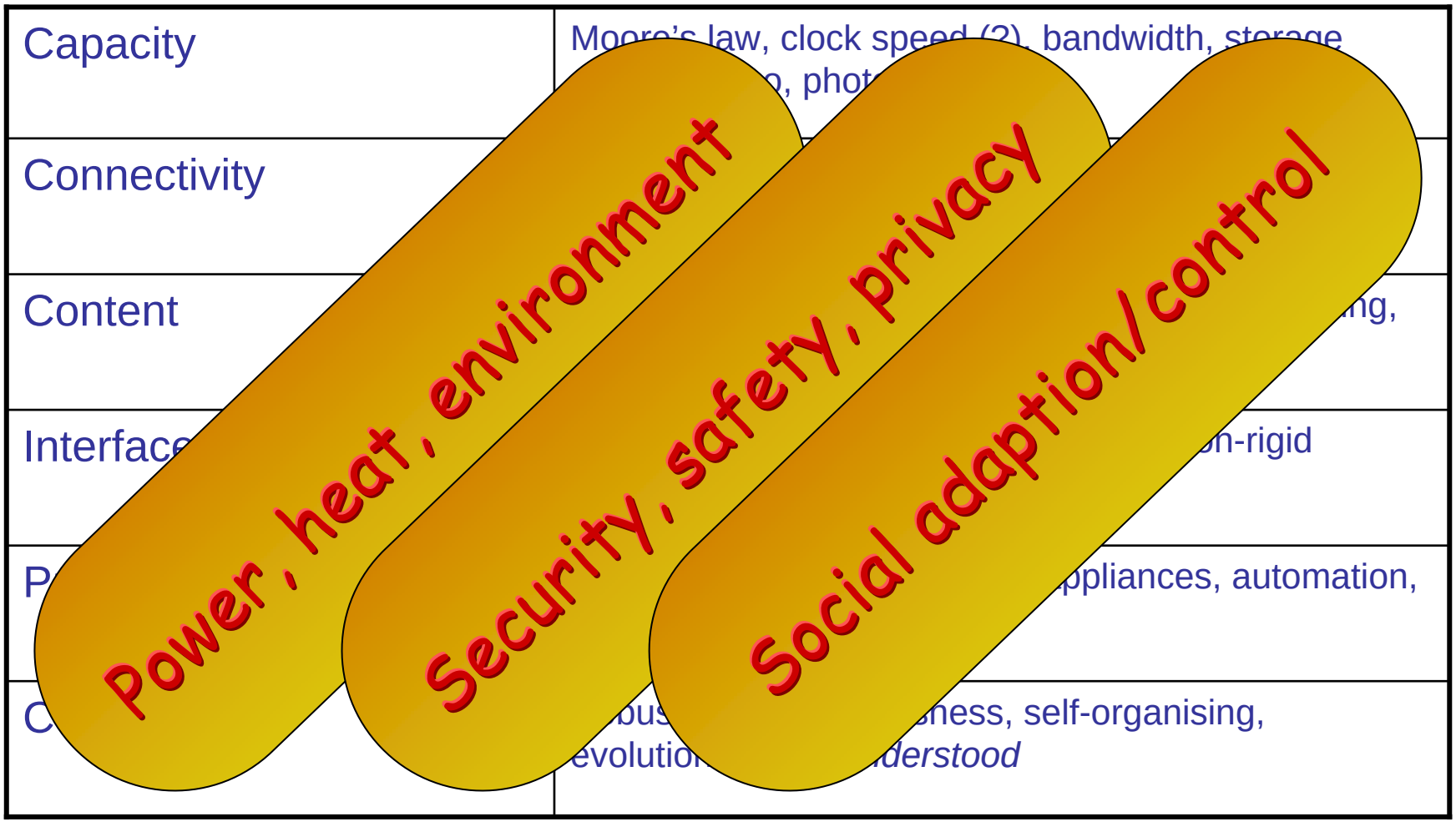
Now, ICT gathers and processes information faster, more reliably and in inhuman quantities, and we have:

- economic globalisation
- remote monitoring and control
- integrated supply chains
- entirely new knowledge

The promise is expansive:

- Improved efficiency, reduced waste, shared experiences, enhanced communities, increased variety
- Information, entertainment and knowledge gathered, kept, synthesised, processed and delivered to suit the needs of any moment and every individual at marginal cost
- Nothing ever lost, nothing ever accidentally forgotten or missed, nobody ever without help or assistance

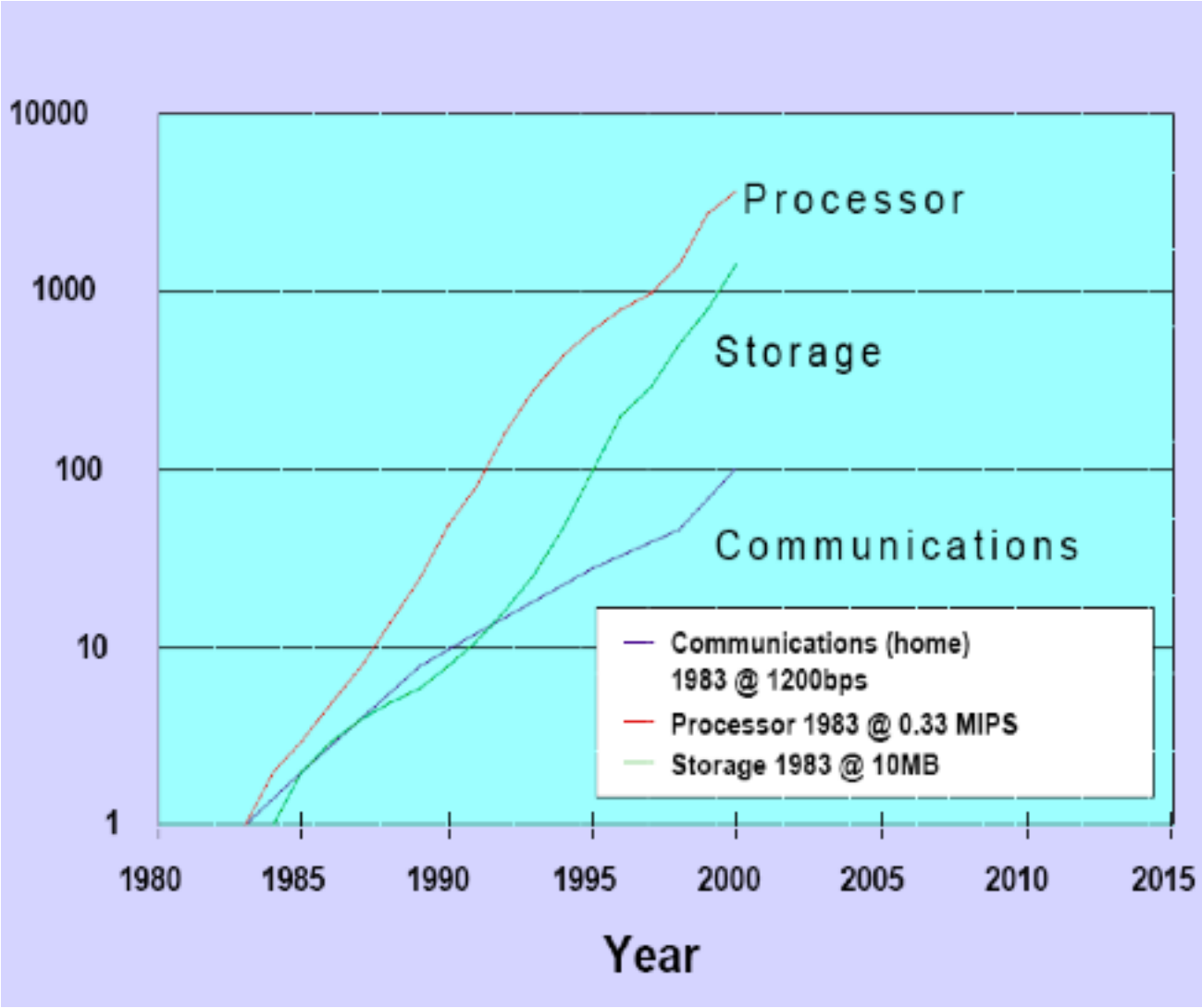
From a technology point of view



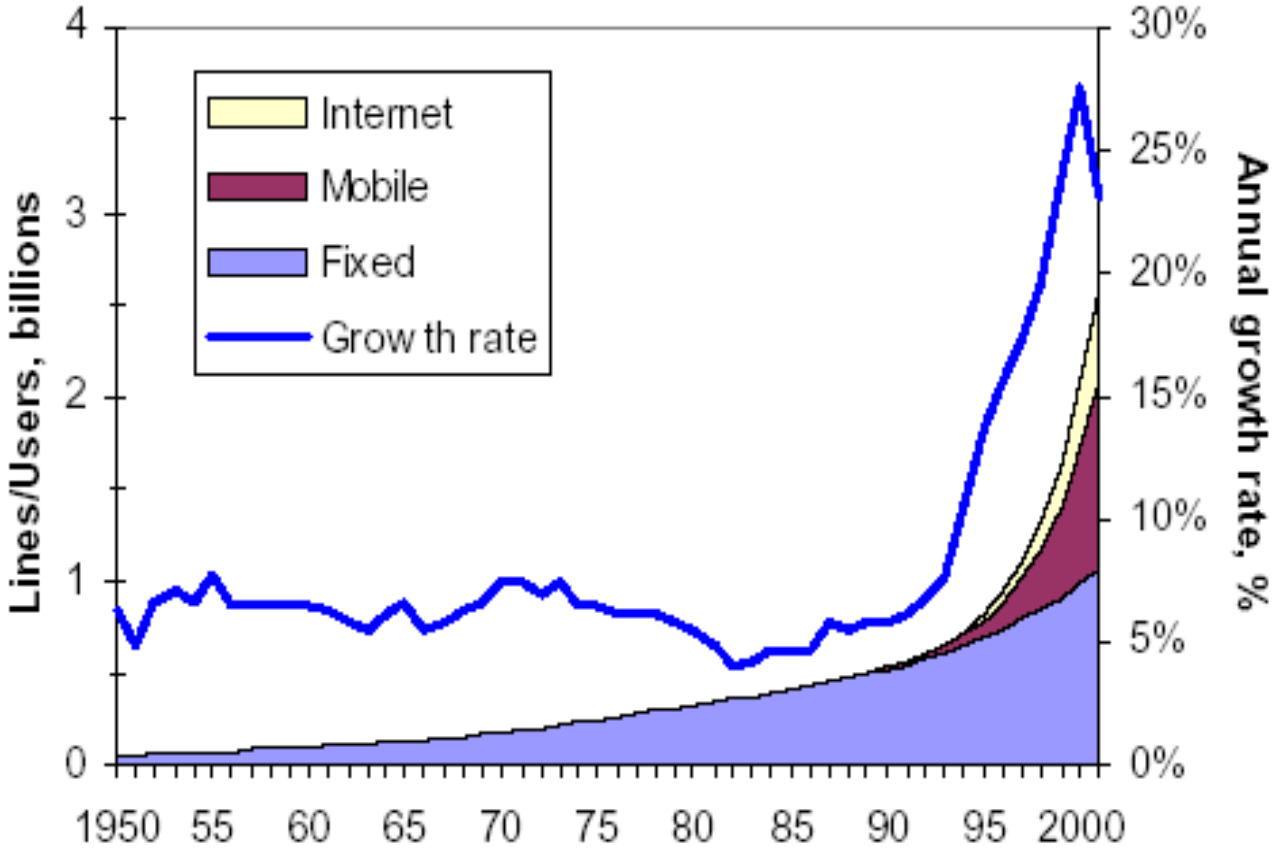
From a technology point of view

Capacity	Explosive growth started 50-60 years ago
Connectivity	Explosive growth started 15-20 years ago
Content	Explosive growth started 3-5 years ago
Interface	The soft visual, smart speech, sensed interface maybe around in 10 years time
Products	Smart autonomous devices maybe 25-100 years
Complexity	Major problem, complexity still exploding, maybe a forever challenge

Raw technology

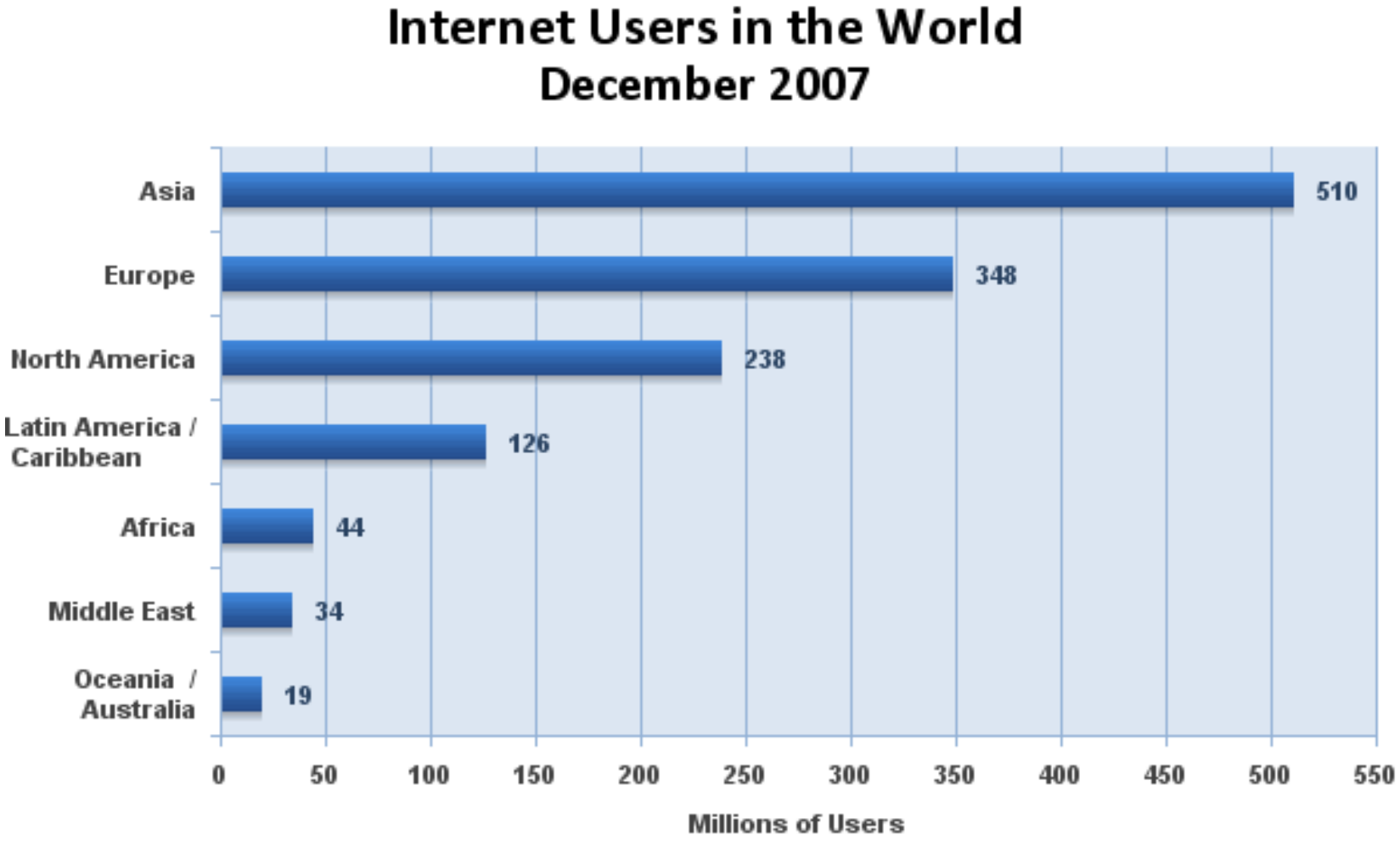


The world is connecting - it's a hand held world



Worldwide	Cell Phones	Internet	PCs
2005	32%	17%	14%
2010	48%	26%	20%

No longer an “English speaking” internet



Note: Total World Internet Users estimate is 1,319,872,109 for year-end 2007
Copyright © 2008, Miniwatts Marketing Group - www.internetworldstats.com

Information Explosion? (Yes, but person to person)

- New stored information grew about 30% a year from 1999 to 2002
- We stored about 5 exabytes of new information in 2002
  - On print, film, optical storage and magnetic (92% - mostly hard disks)
  - 5 exabytes
    - Sufficient to digitize the library of congress 37,000 times
    - Sufficient to store all sentences ever spoken in human history
- Our telephone calls worldwide (fixed and mobile) contained 17.3 exabytes of new information if stored in digital form
  - voice was 98% of all new information transmitted (person to person)
  - radio amounted to about 3.5 petabytes and TV about 70 petabytes
- The internet contained about 0.5 exabyte
  - 83% in email, 17% in the ‘deep web’, a negligible amount in the ‘web’

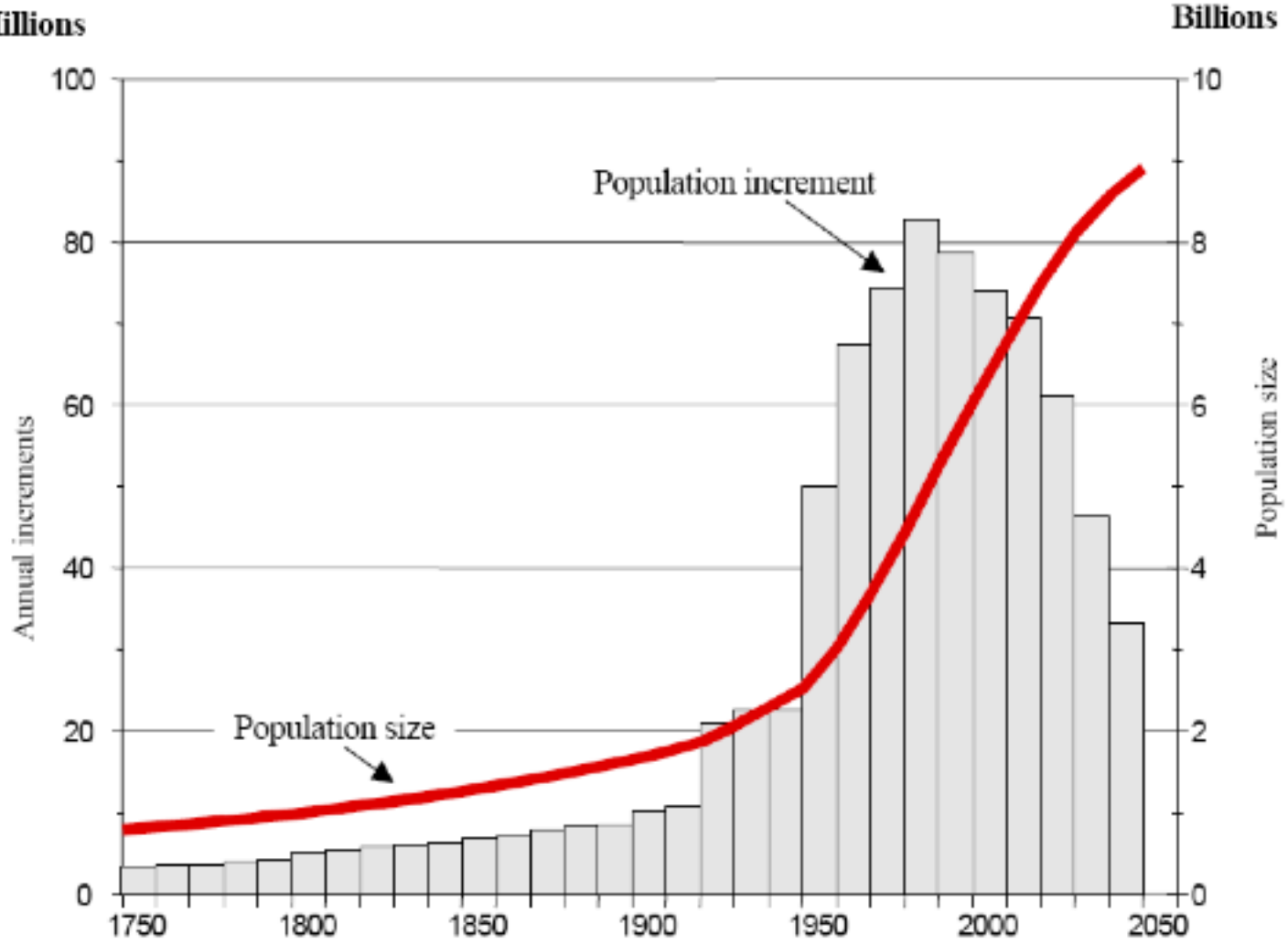
Population

1 AD
~200 million

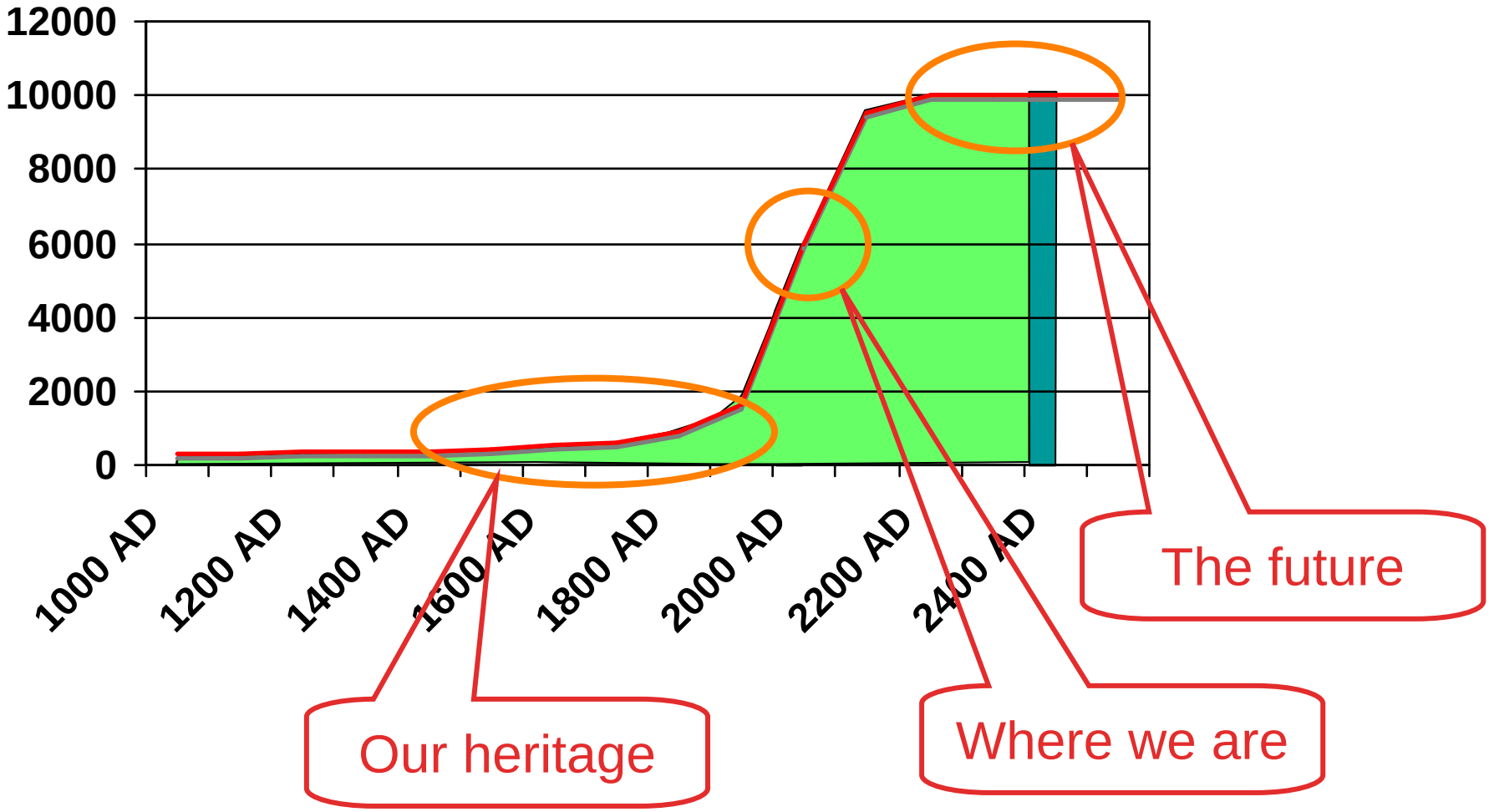
1000 AD
~300 million

1500 AD
~500 million

1750 AD
800 million



The future researchers



The spike or the ’S’ curve

- Obviously the population explosion is being contained so that the raw input to intellectual output (people) will level off
- Soon everybody will be connected so connectivity will level off
- Never ending growth in compute power seems fanciful
- The growth in sensors and laboratory automation is still in its infancy, so dramatic scale ups in data acquisition are foreseeable and remain to be coped with
  - The entire human genome project could be completed in a day in a few years time

- DATA IS THIS GENERATION’S PROBLEM
- But overall, the result trends towards an increasing store of ‘old’ knowledge and increasingly less (as a %) ‘new’ knowledge in any year!!

The human dilemma

- A step change in population and its urbanisation provides a step change in the amount of research in progress
- An avalanche of information results from an explosion in connectivity
  - to each other and to computing and data systems
- Vast amounts of information arise from sensor development
  - but no person can peruse it
- What if we suppose there will be no corresponding substantive change in a person's ability to learn and understand
  - the information space understood by any researcher must shrink as a fraction of the total aggregate information and as a fraction of the information visible to them
- So how do we conduct research by the increasingly ignorant?
  - Or, more politely, how do we conduct research in an information or knowledge rich environment

There are various paths forward

- Work with higher bandwidth human faculties
  - The interactive modeling and visualisation theme
- Make knowledge more accessible
  - The smarter tools solution
- Learn to work in teams
  - The collaborate or perish proposition
- Build servants more intelligent than us
  - The magic wand approach
- Extend the human ability to learn
  - Open pandora's box, augmentation
- … others ?

Why eResearch, and why now?

- We are in the midst of the information explosion
- We are in the midst of change
- eResearch is about establishing the foundations we need in order to apply ICT to knowledge in three ways:
  - Enhan
    - We
    - We
  - Enable
    - We
    - We
  - Encod
    - We
    - We

Lofty goals

that we are only now

building the means

to achieve

Thank you

For further information, contact:

Rhys Francis

Executive Director

Australian eResearch Infrastructure Council

Email: rhys@pfc.org.au

Website: www.pfc.org.au