eResearch: librarians pushing technology to perform

Martie van Deventer

Portfolio Manager

CSIR Information Services

IUG Directors' meeting

Port Elizabeth 12 November 2008

CSIR
our future through science

Changes in Traditional Research

Visualisation

Cyber infrastructure

Global Collaboration

Personalisation

Big science

e-Research

Previously impossible research

VRE

Opportunities

New technologies

South Africa

Threats

HPC

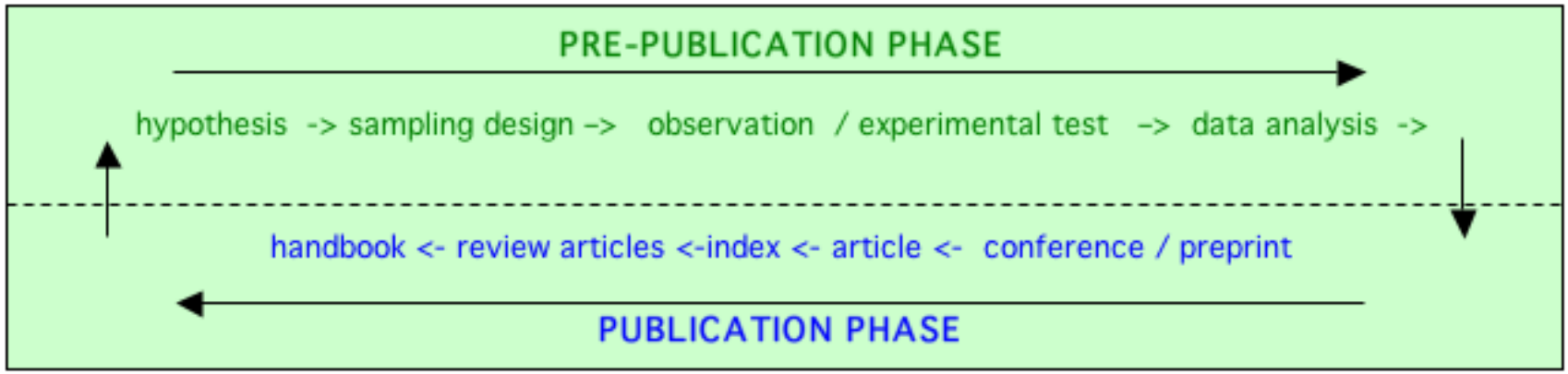
International

Multidisciplinary

Definitions

CSIR
our future through science

Research is changing



(Gold, 2007, Part 1)

Today much of the work of scientists is **internet-based**. From their desks scientists access special hardware, software, data and applications as parts of distributed systems. Typically access to them has to be made explicit by keying in specific data. **Cooperation** with other scientists (e.g. the **exchange of interim or final results** of research for annotation and further use) typically takes place via the internet. (Osswald, 2007: 517)

Even if publications are still the currency of tenure, data is the currency of science, (Gold, 2007) … !

Ready for eResearch?

- My research involves **working with researchers** from other organisations in South Africa/ Africa and/or overseas …
- Data and/or **equipment** that would be useful to me are located **elsewhere …**
- I have data and/or equipment that other researchers would like to access …
- I need to send/receive large quantities of data (e.g. audio, video, datasets) …
- I am aware of international research initiatives in my area and would like to participate in them …

Based upon: http://www.mcs.vuw.ac.nz/EResearch/KAREN

CSIR
our future through science

Ready for eResearch?

- I do not **collaborate with other researchers** in other organisations but I would like to in future
- I organise / attend meetings regularly with other researchers
- I have data and/or equipment that other researchers would like to access
- I use software programs to analyse data - it is very complex / takes a long time but produces a **graphic** that I can interpret
- I often email/phone colleagues elsewhere in South Africa/ Africa and overseas to discuss research projects and common research interests

Based upon: http://www.mcs.vuw.ac.nz/EResearch/KAREN

CSIR
our future through science

eResearch: challenges for the librarian

- Most librarians are much **less familiar with the data-generating research phases** of the scientific research cycle than with post-research
- **Data science** and data management are an **awkward fit** with the **text**-oriented constructs and systems that still dominate library relationships with science communication and publishing
- **Becoming literate** in cyberinfrastructure means understanding cyberinfrastructure, eResearch, collaboratories, collaboration science, computational and grid science, data curation, the Semantic Web, open data, data archiving, digital preservation, and data management, and how they relate to each other
- **Phases** of reporting, communication and publication are **less distinct**
- When used in discussions of eResearch, a **vocabulary** familiar to librarians (archival, curation, stewardship, provenance) takes on new or specialized meaning … major cause of **confusion**

CSIR
our future through science

- **What is currently the same**
  - Having to adapt to change
  - Researchers currently still regard libraries as trusted service providers … but for how long?

- **What's Different**
  - Collaboration … science has to become big and multi-disciplinary to solve the really big challenges
  - Research often relies on visualisation
  - Researchers need partners in research rather than just suppliers and support staff
  - Expanded skill sets – need to learn to do different tasks

CSIR
our future through science

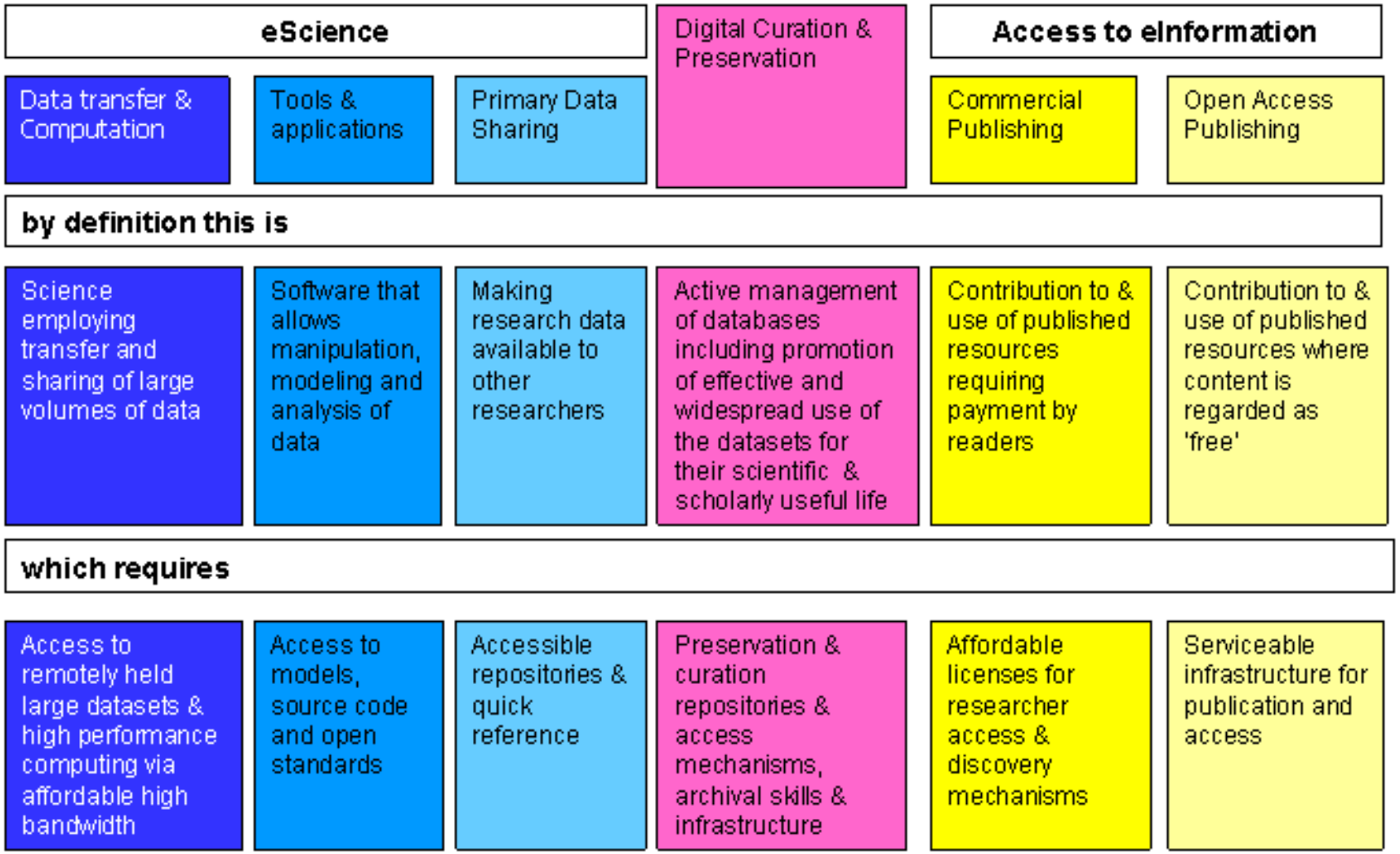
What is eResearch?

- eResearch is research done **faster**, to
- a **better quality**, or
- by **different methods**, using
- **advanced digital tools** and
- services that enable **diverse research expertise** to be
- assembled in **global teams** focused on
- **very specific research problems**

Hine, 2008, http://www.mcs.vuw.ac.nz/twiki/pub/EResearch/Resources/symposium_hine.pdf

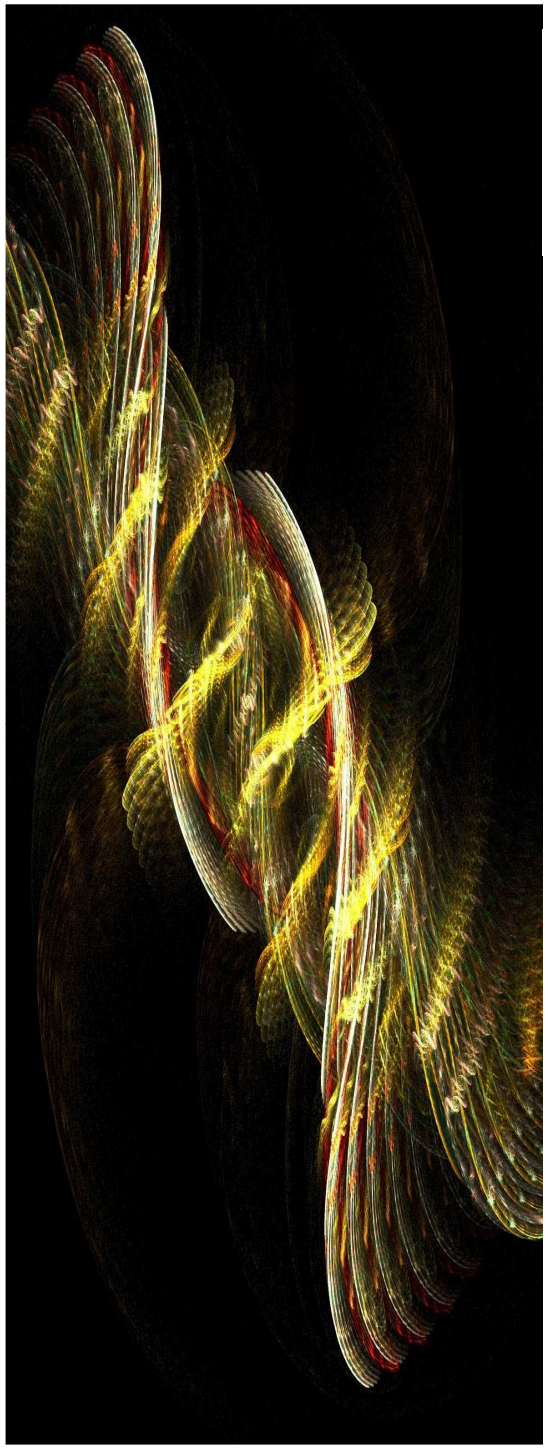
CSIR
our future through science

Components of eResearch



Researcher Requires: Perpetual access, Curation, Training, Marketing

Supplier must ensure: Security - Access, Authorization, Authentication



Examples of eResearch: Serious Disease Genes Revealed

- Wellcome Trust Case Control Consortium
- 50 research groups
- 200 scientists
- DNA from 17,000 patients
- 15,000 polymorphic markers
- Learned more in 12 months than last 15 years

http://www.mcs.vuw.ac.nz/eresearch

CSIR
our future through science

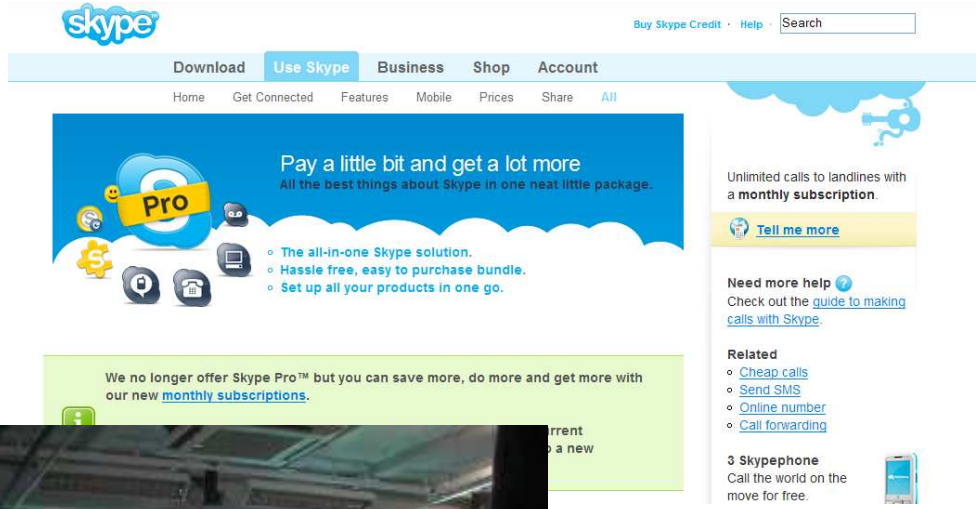
Functional MRI (fMRI) Data Centre

- Online repository of neuroimaging data
- Typical study comprises:
  - 3 groups
  - 20 subjects/group
  - 5 runs per subject
  - 300 volumes per run
  - 90,000 volumes,
  - 60GB raw data
  - 1.2 million files processed
- 100s of such studies

CSIR
our future through science

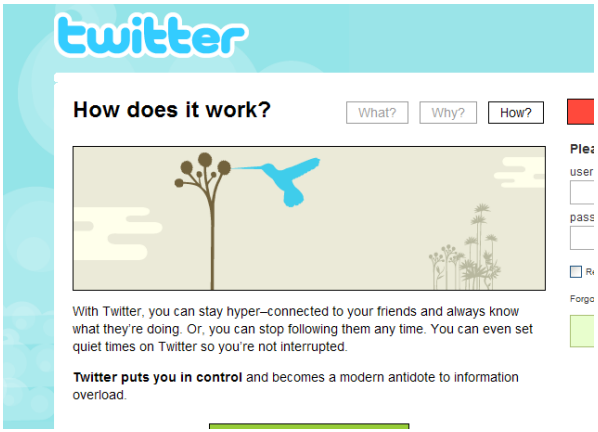
Collaboration, co-ordination & communication

Desktop video-conferencing



Skype

Video-conferencing



Twitter: 21st century telegraph

CSIR
our future through science

Visualisation

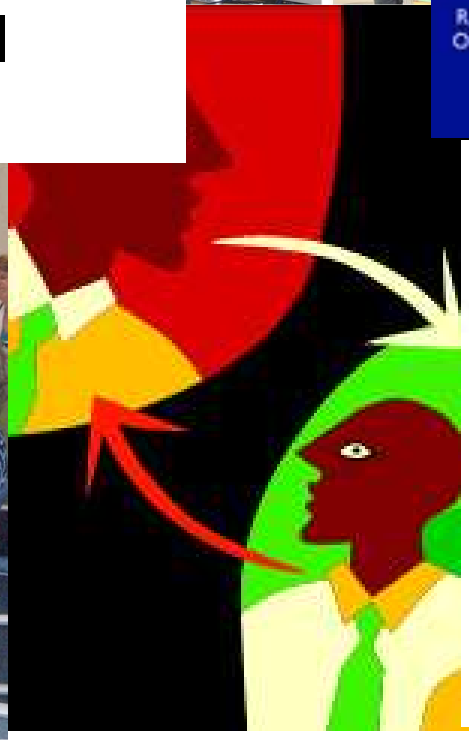
Tools: proprietary eg Mathlab but open source is rapidly gaining market share

http://www.mcs.vuw.ac.nz/EResearch/Symposium

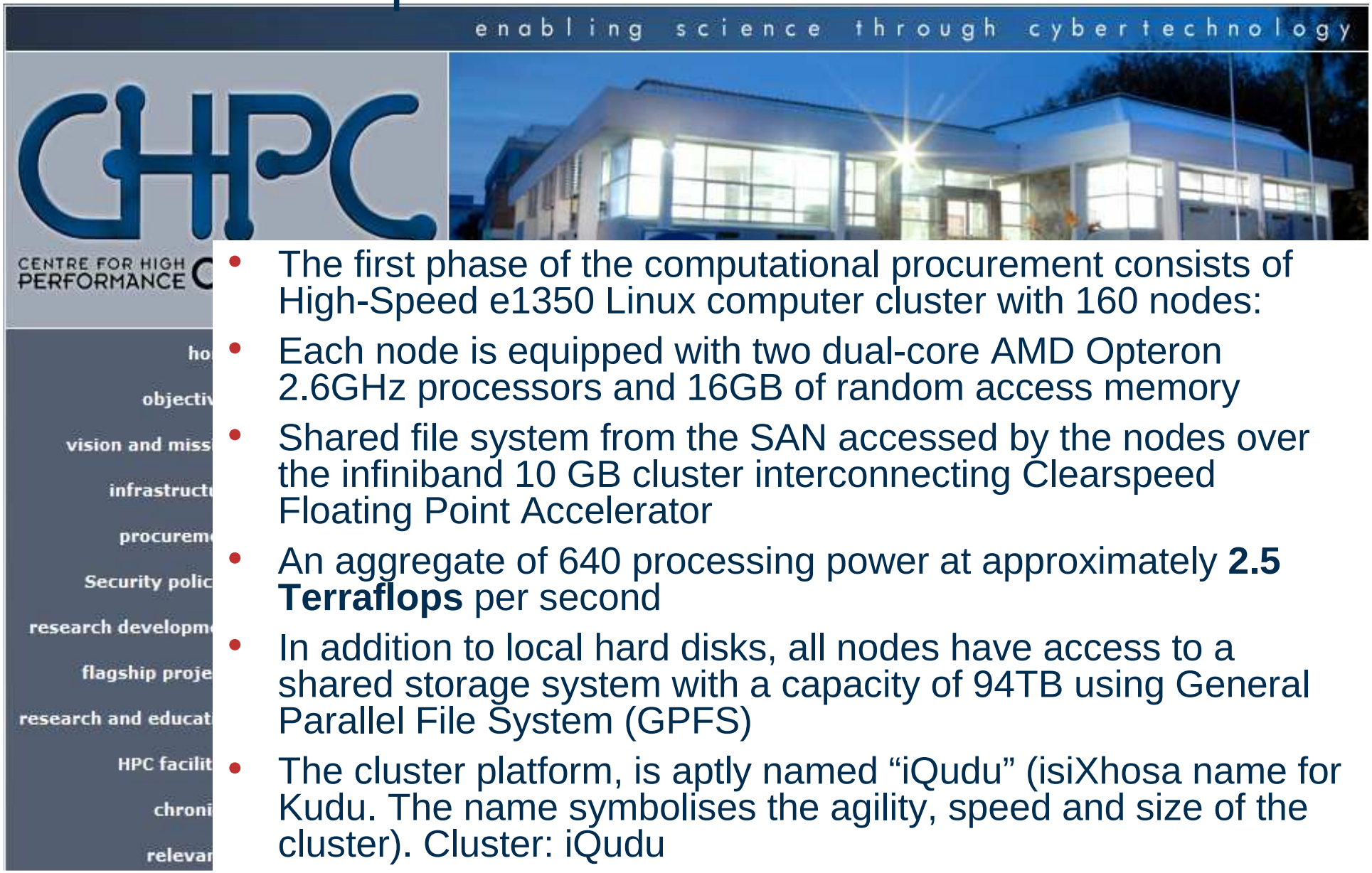
CSIR
our future through science

Working in multidisciplinary teams

Expect: jargon confusion, language barriers, cultural barriers, etc

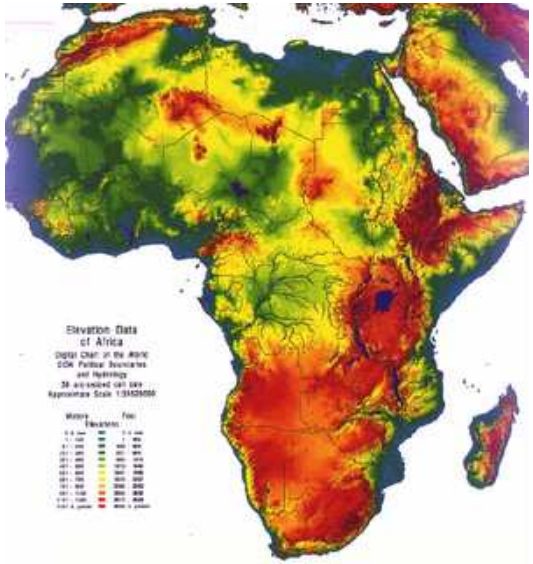


Centre for High Performance Computing – Rosebank Cape Town



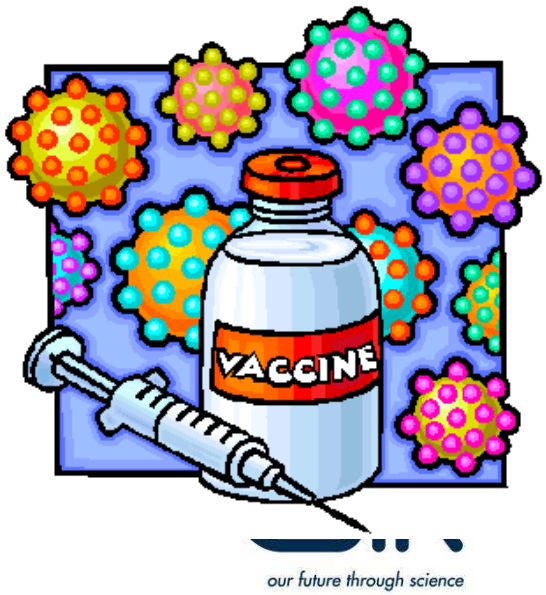
- The first phase of the computational procurement consists of High-Speed e1350 Linux computer cluster with 160 nodes:
- Each node is equipped with two dual-core AMD Opteron 2.6GHz processors and 16GB of random access memory
- Shared file system from the SAN accessed by the nodes over the infiniband 10 GB cluster interconnecting Clearspeed Floating Point Accelerator
- An aggregate of 640 processing power at approximately **2.5 Terraflops** per second
- In addition to local hard disks, all nodes have access to a shared storage system with a capacity of 94TB using General Parallel File System (GPFS)
- The cluster platform, is aptly named “iQudu” (isiXhosa name for Kudu. The name symbolises the agility, speed and size of the cluster). Cluster: iQudu

Examples of what will CHPC be doing



- Develop new regionally specific **climate change scenarios for 2025**; resource planning in the Western Cape based on models developed from quantifying rainfall vulnerability coupled with impact of increased population growth in the region.

Speed up the development of vaccines by studying fragments of 'foreign' proteins that have been shown to trigger a response by the immune system; search for common patterns in the sequences or structures of those proteins and use that information to create a theoretical model of the features that cause the fragments to trigger the immune response. These methods are specifically applicable to the AIDS and malaria vaccine projects.



http://www.chpc.ac.za/challenges.html

What are eResearchers worried about?

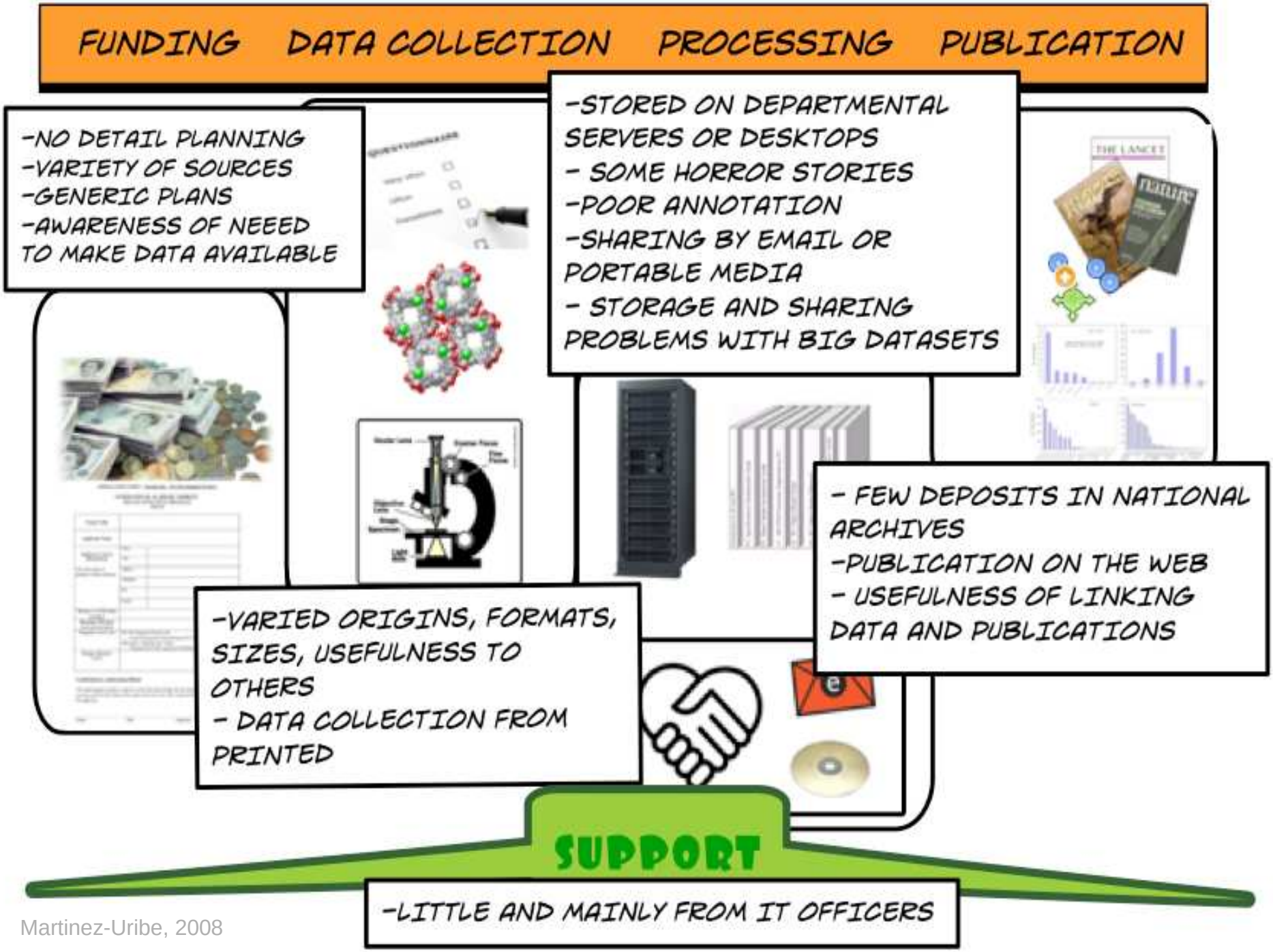
- How do we describe & catalogue the artefacts, resources, experiments & knowledge we have so that others can find, use, and understand these things?
  - What are effective carriers of meaning (between researchers) in our field?
  - Can these carriers be represented / emulated in systems? How?
  - How much do we need to rely on top-down imposition of meaning? And how much on bottom-up?
  - Can we reason across geological settings?
- How do we fit knowledge computing into current work practices?
  - How do we engineer **knowledge capture** so that it is as **unobtrusive** as possible?
  - How do we maximize utility of captured knowledge in future work?
  - Governance: **who gets to make the rules**?

CSIR
our future through science

Further challenges facing eResearchers

- Preserving digital content and even understanding what "preservation" means
- The absence of practices for refreshing and migrating both data and relevant retrieval software
- Capture (archiving) and selection, as well as providing sufficient metadata or other descriptive or administrative information to ensure adequate (as well as legal) access and retrieval over the long term to electronic content created
- Insufficient knowledge of the software applied and an inability to find relevant information
- Missing or ignoring standards as well as obstacles related to licensing and accountability
- **Bottom line: without 'librarian'-skills (adapted to the eResearch environment) researchers will not be able to benefit fully from their own past – they do not wish to be librarians**

CSIR
our future through science



Martinez-Uribe, 2008

When asked, Oxford scientists requested:

Luis Martinez Uribe
Digital Repositories Research Co-ordinator

- ***Advice on practical issues*** *related to managing data across their life cycle. This help would range from assistance in producing a data management/sharing plan; advice on best formats for data creation and options for storing and sharing data securely; to guidance on publishing and preserving these research data.*
- *A* ***secure and user-friendly solution*** *that allows storage of large volume of data and sharing of these in a controlled fashion way allowing fine grain access control mechanisms.*
- *A* ***sustainable infrastructure*** *that allows publication and long-term preservation of research data for those disciplines not currently served by domain specific services such as the UK Data Archive, NERC Data Centres, European Bioinformatics Institute and others.*
- ***Funding*** *that could help address some of the departmental challenges to manage the research data that are being produced.*

SAMI VRE components

Red: none
Orange: some
Yellow: all

Repositories: research results; experiments; literature & documents

Web/wiki/blog: search engines, databases; researchers & topics; funders, portals, communication, projects

Identification of research area

Literature review & indexing

Internal shared database of indexed articles

Dissemination & artifacts

Identification of collaborators

Document management system

Skype, smart board, video conferences

Real time communication

Proposal writing

Generic software e.g. MS / Open Office

Training / mentoring etc

Identification of funding sources

E-learning system for researchers

Scientific workflow

Project management

(Collaborative) Electronic Lab book

Project management system

Servers with data files

Sophisticated instruments that generate digital information and data

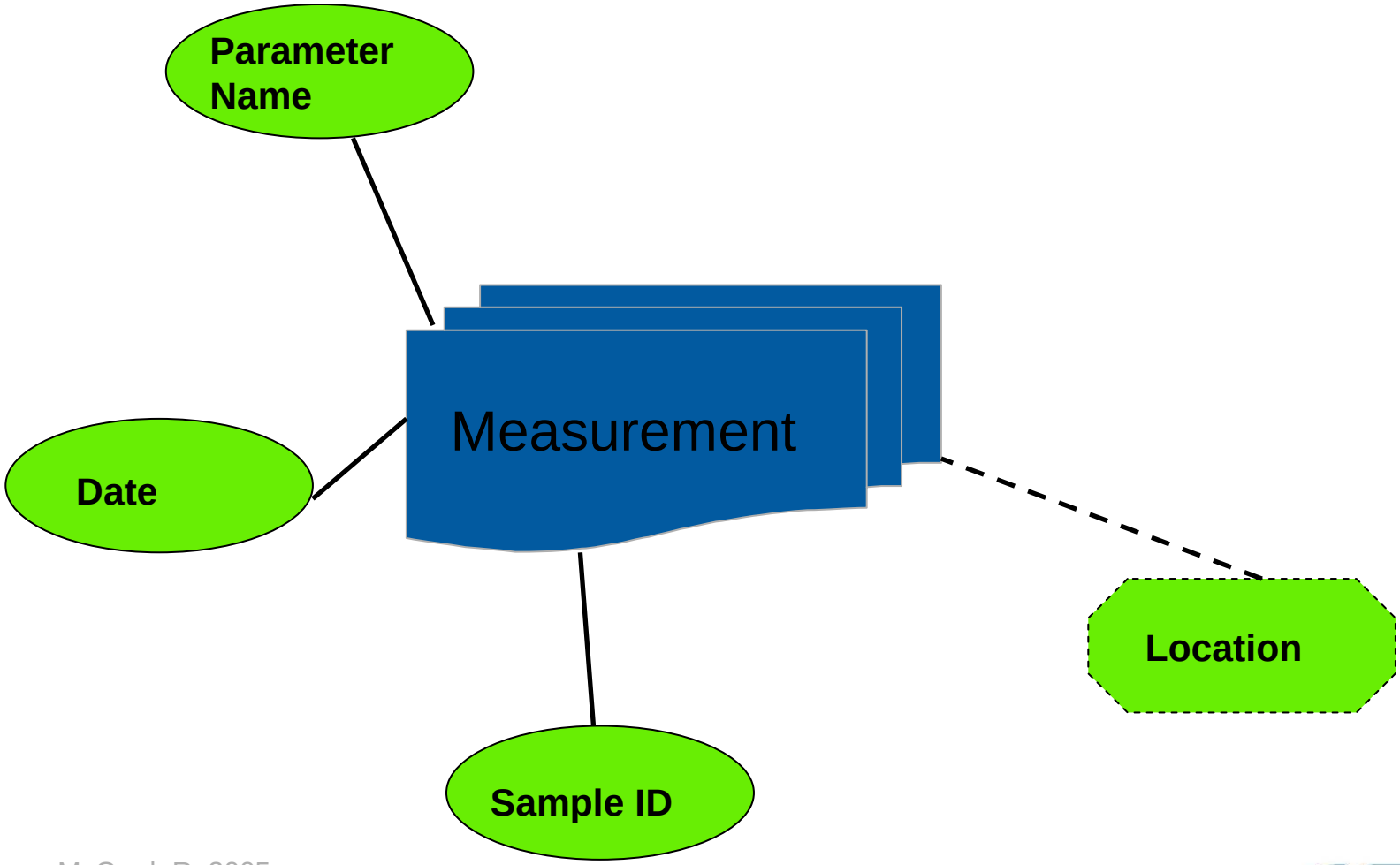
Access to research networks & super computers; access to labs with in silico screening +

Integrated data management system

Mathematical modelling tools; numerical algorithm tools; simulation software; in silico experiments

(Free) Data analysis software

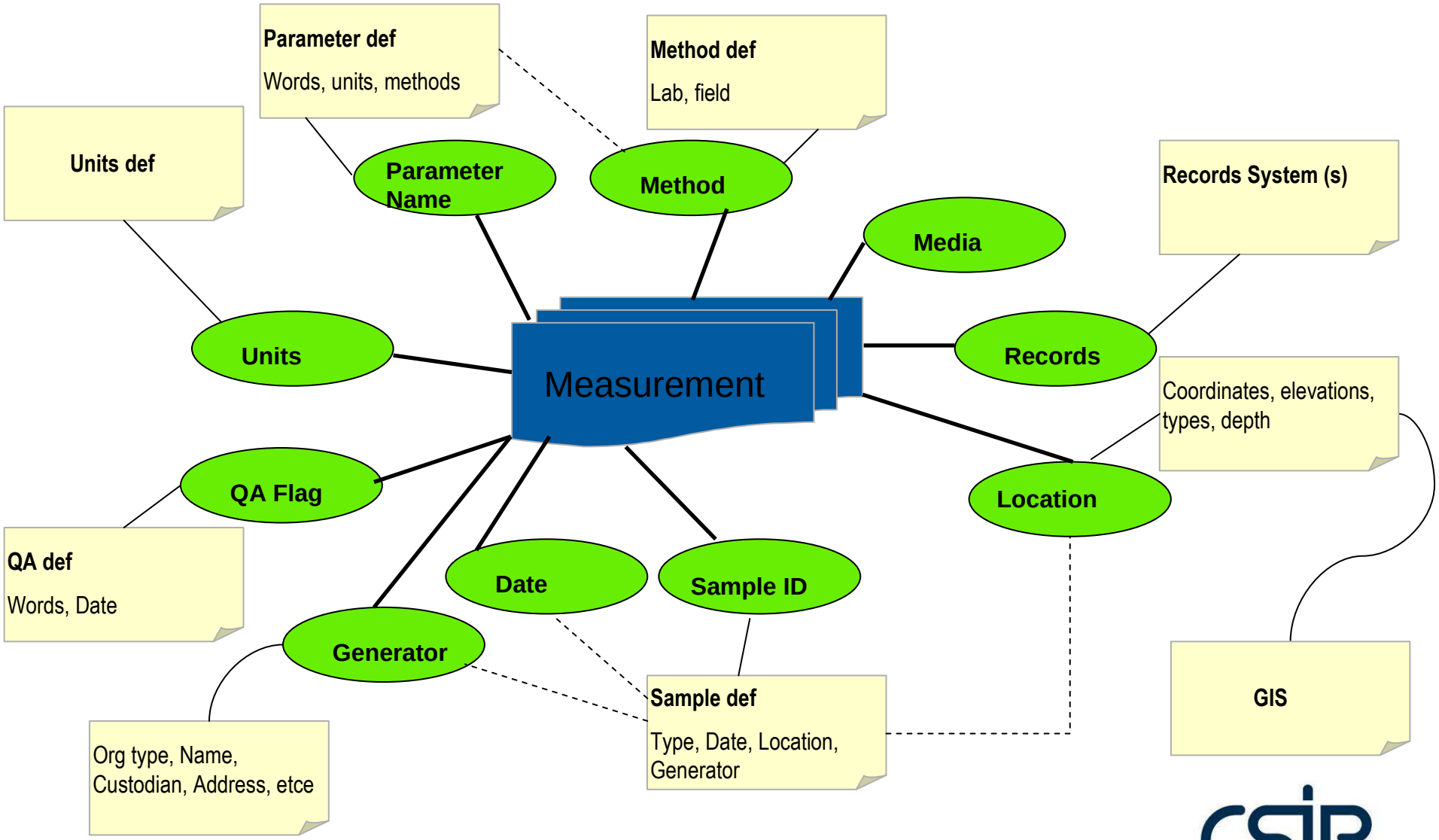
Single experiment view of data



Source: McCord, R. 2005

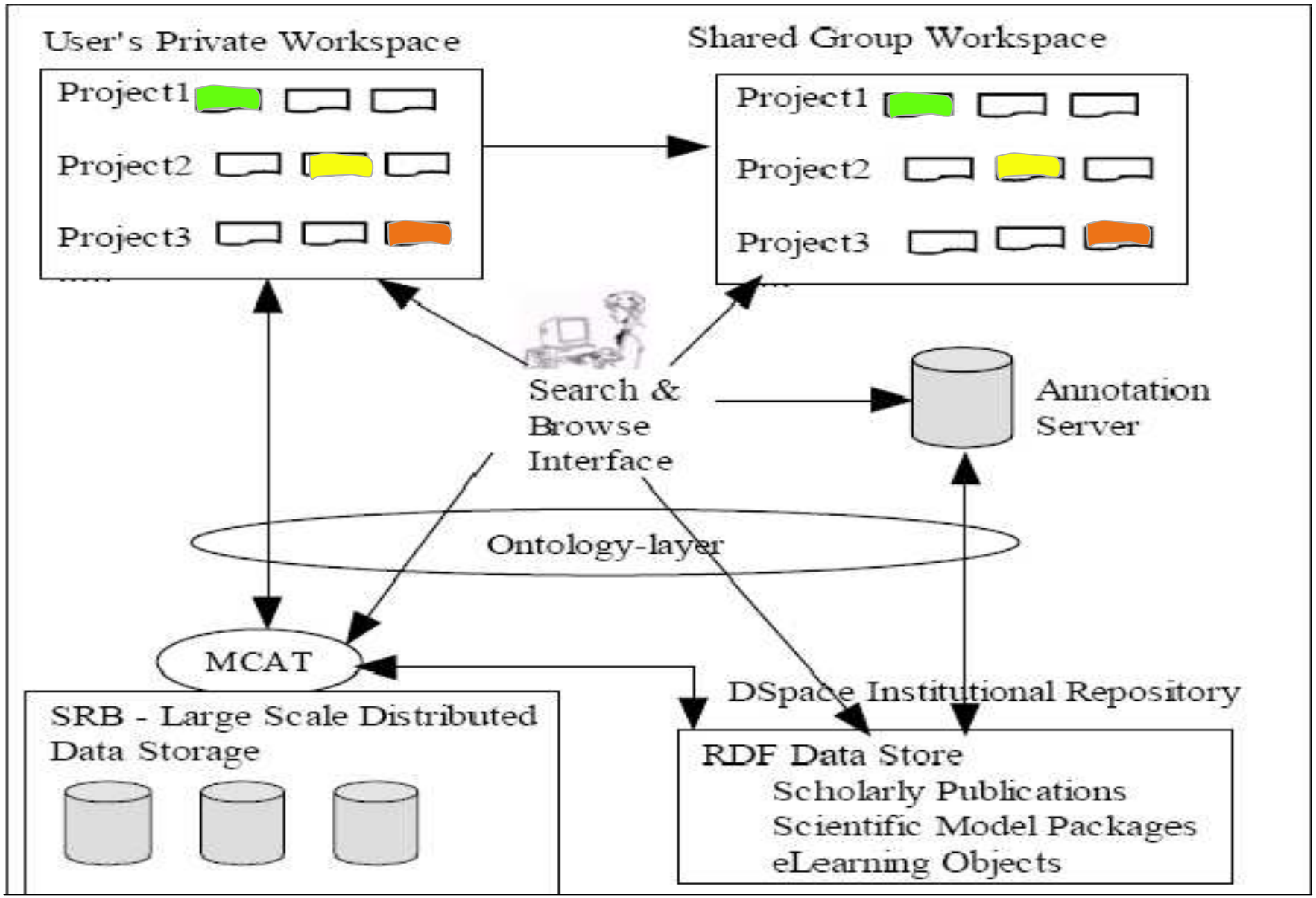
CSIR
our future through science

Integrated System & Curation View



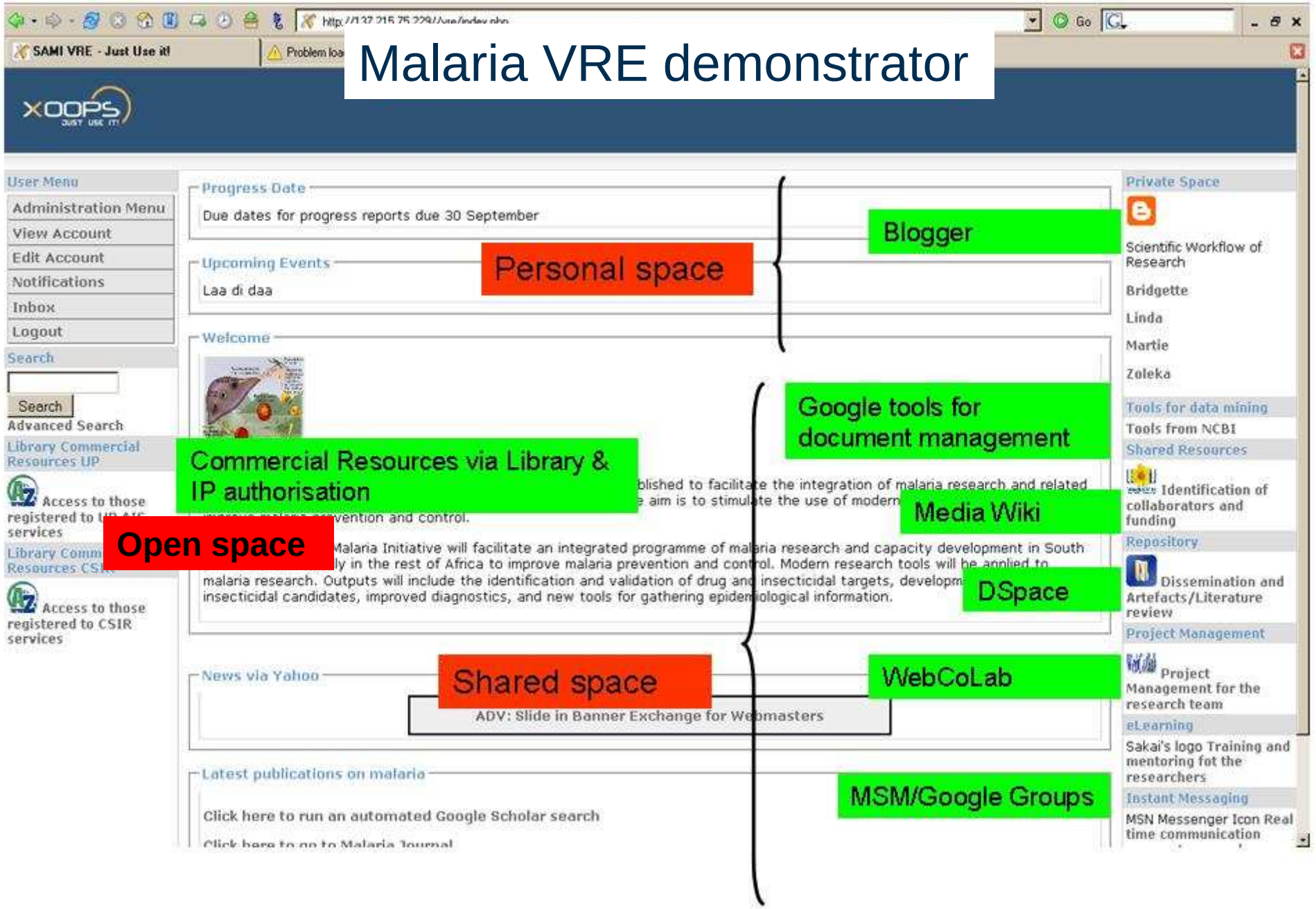
Source: McCord, R. 2005

CSIR
our future through science



Different storage areas envisaged in an scientist's ideal environment, and the relationships between them (Hunter)

Malaria VRE demonstrator



What can librarians do? … Traditional …

- Select, acquire, and license data and data sets;
- Create metadata (or metadata standards) for discovery and description of data sets … but at eResearch scale
- Creating and/or organize documentation related to data
- Offering preservation services for digital data
- Advise on the appraisal and selection of what data to keep for the long term

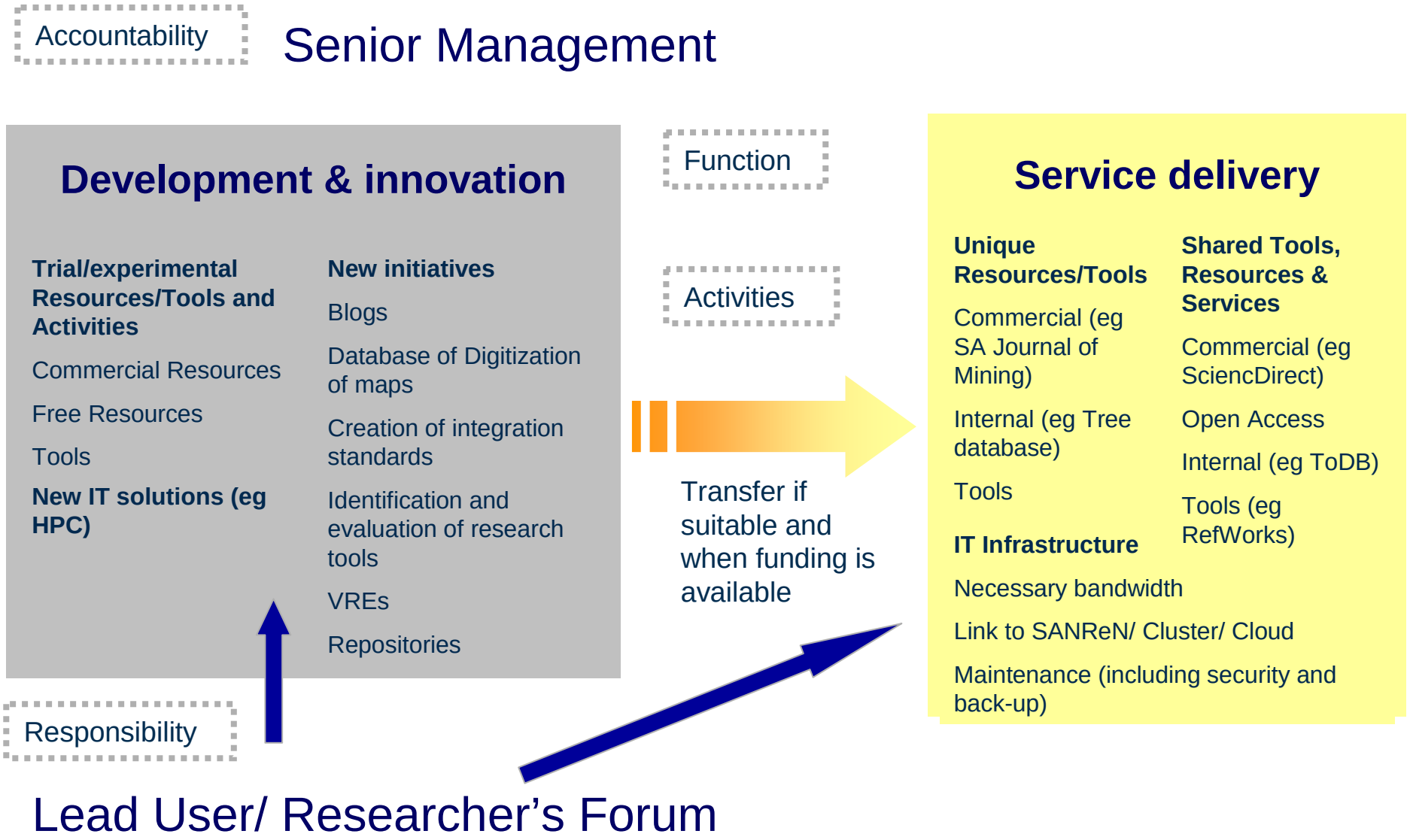
What can librarians do? … Traditional 2 …

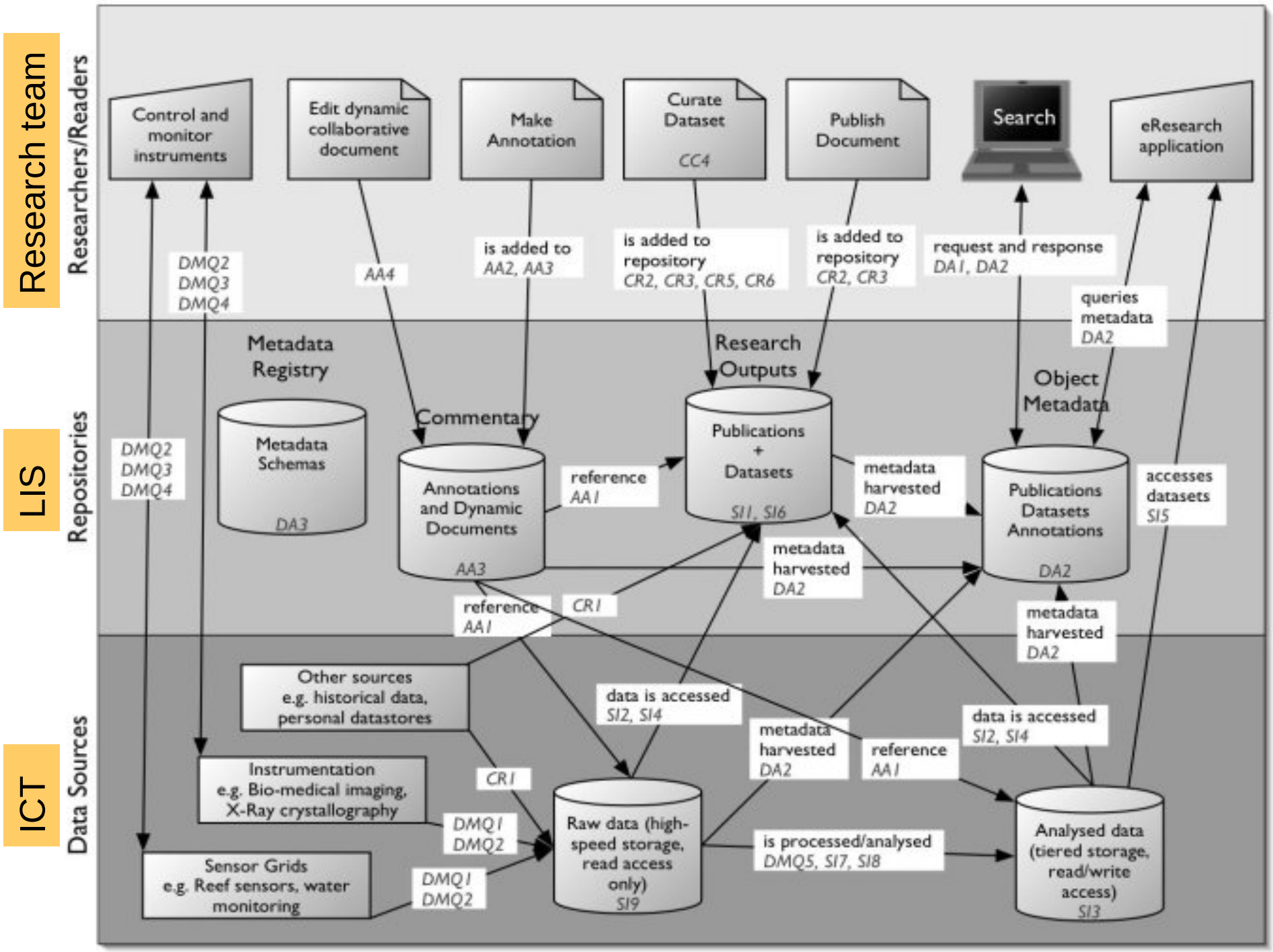
- Assist users with finding data relevant to their research, using third-party high level directories and data discovery sources
- Develop data publication standards and systems
- Publishing workflows, global identifier schemes, linking schemes, standards for data clean-up and normalization, and also standards for providing credit and recognition to data authors
- Develop practices for rights management (– how do you site data sets?)
- Offer long-term repositories of scholarly output

What can librarians do? … New

- Position themselves as partners in research – embedding themselves
- Create more dynamic repositories that support pre-publication workflows
- Learn to practice eResearch themselves!

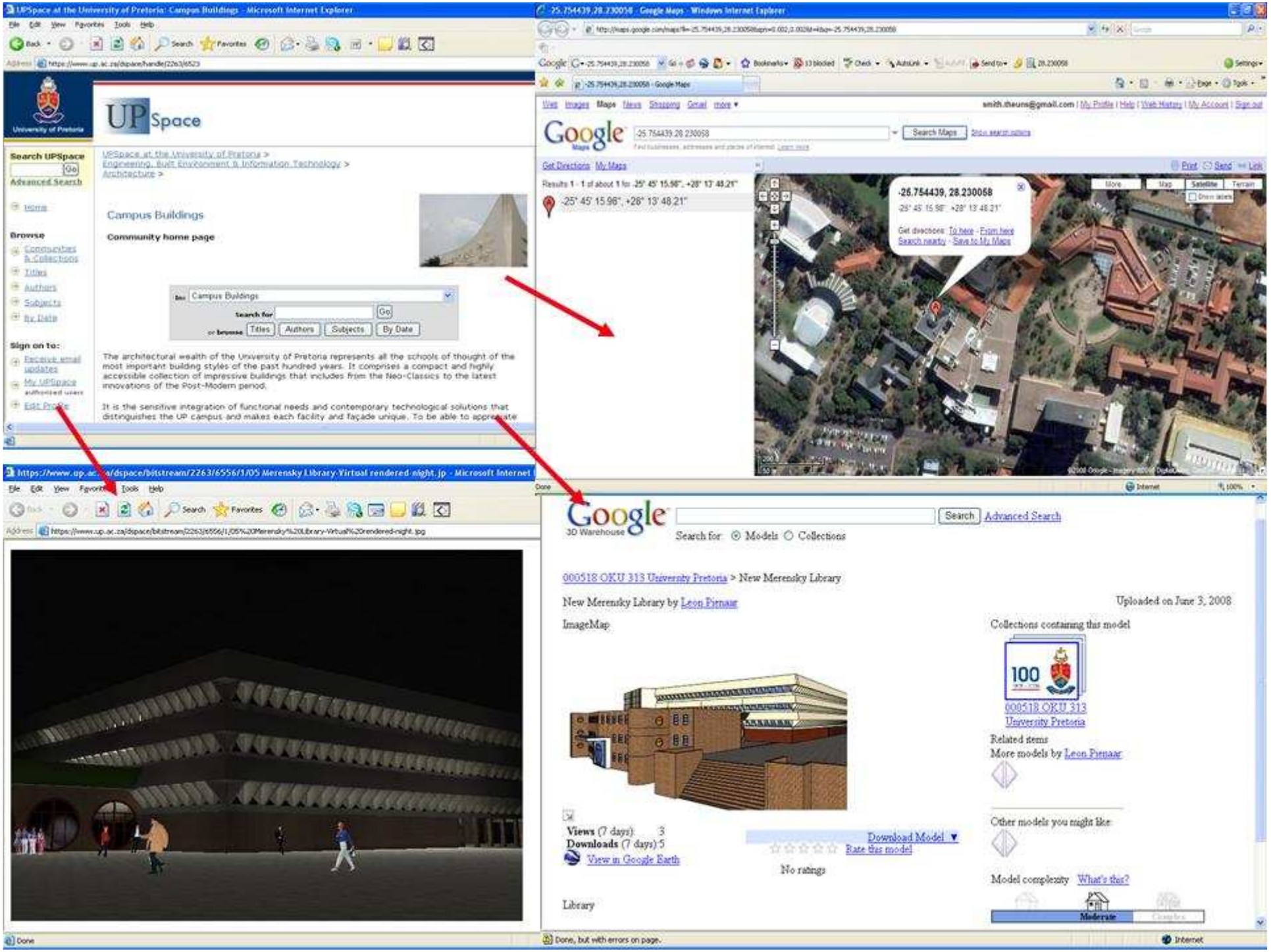
eResearch Support Framework



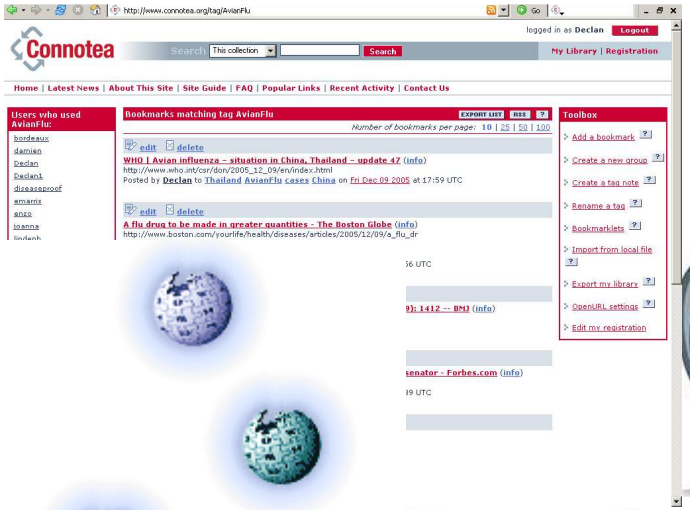

Research team
LIS
ICT
Researchers/Readers
Repositories
Data Sources
Control and monitor instruments
Edit dynamic collaborative document
Make Annotation
Curate Dataset
CC4
Publish Document
Search
eResearch application
DMQ2
DMQ3
DMQ4
AA4
is added to
AA2, AA3
is added to repository
CR2, CR3, CR5, CR6
is added to repository
CR2, CR3
request and response
DA1, DA2
queries metadata
DA2
Metadata Registry
Metadata Schemas
DA3
Commentary
Annotations and Dynamic Documents
AA3
Research Outputs
Publications + Datasets
SI1, SI6
Object Metadata
Publications Datasets Annotations
DA2
accesses datasets
SI5
reference
AA1
metadata harvested
DA2
CR1
Other sources
e.g. historical data, personal datastores
Instrumentation
e.g. Bio-medical imaging, X-Ray crystallography
Sensor Grids
e.g. Reef sensors, water monitoring
DMQ1
DMQ2
Raw data (high-speed storage, read access only)
SI9
data is accessed
SI2, SI4
is processed/analysed
DMQ5, SI7, SI8
Analysed data (tiered storage, read/write access)
SI3

Learn to visualise (eg find out more about our users)

CSIR
our future through science

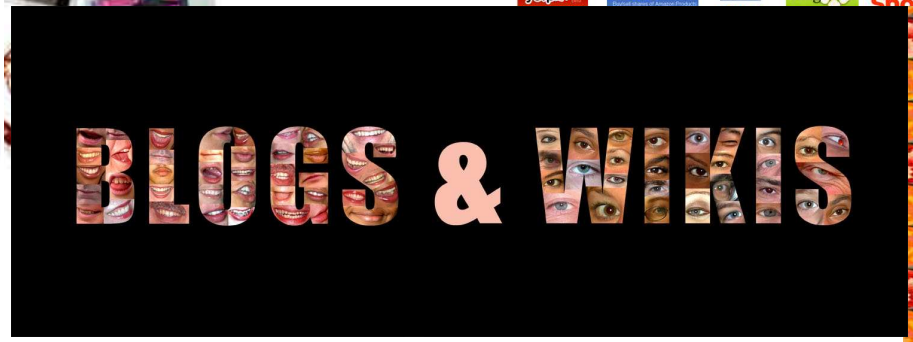

UPSpace at the University of Pretoria: Campus Buildings - Microsoft Internet Explorer
UPSpace
Campus Buildings
Community home page
The architectural wealth of the University of Pretoria represents all the schools of thought of the most important building styles of the past hundred years. It comprises a compact and highly accessible collection of impressive buildings that includes from the Neo-Classics to the latest innovations of the Post-Modern period.
-25.754439, 28.230058 - Google Maps - Windows Internet Explorer
-25° 45' 15.98", +28° 13' 48.21"
Google 3D Warehouse
New Merensky Library by Leon Pirnaar
ImageMap
Views (7 days): 3
Downloads (7 days) 5
View in Google Earth
Download Model
Rate this model
No ratings
Collections containing this model
000518 OKU 313 University Pretoria
Related items
More models by Leon Pirnaar
Other models you might like
Model complexity What's this?
Moderate
Library

Learn to collaborate and co-create


Connotea
Bookmarks matching tag AvianFlu
Toolbox

flickr

09:00 am May 8 07
movie
weather
message
music
map
web
game
headlines
clock
phone
contacts
photo
search


BLOGS & WIKIS

13, 23, 43 Things: Create. Share. Network

- Blogging
- Photos & Images – Flickr & mashups
- RSS & Newsreaders
- Social Tagging – del.icio.us, Connotea
- Social Networking – Facebook, Linkedin
- Tagging, Folksonomies & Technorati
- Wikis
- Online Applications & Tools – online office/ design
- Podcasts, Video & Downloadable Audio - youTube
- Next Generation OPACs

CSIR
our future through science

Next generation OPACs

- As times have changed, our OPACs have not kept up. End-users want more from their library catalogs. They want the ability to create mash-ups with other services (e.g., LibraryThing), save catalog data in new and different ways (e.g., Zotero), and much, much easier findability (e.g., through faceted browsing.)

CSIR
our future through science

LibraryThing BETA

What's on your bookshelf?
Over thirty million books on members' bookshelves.

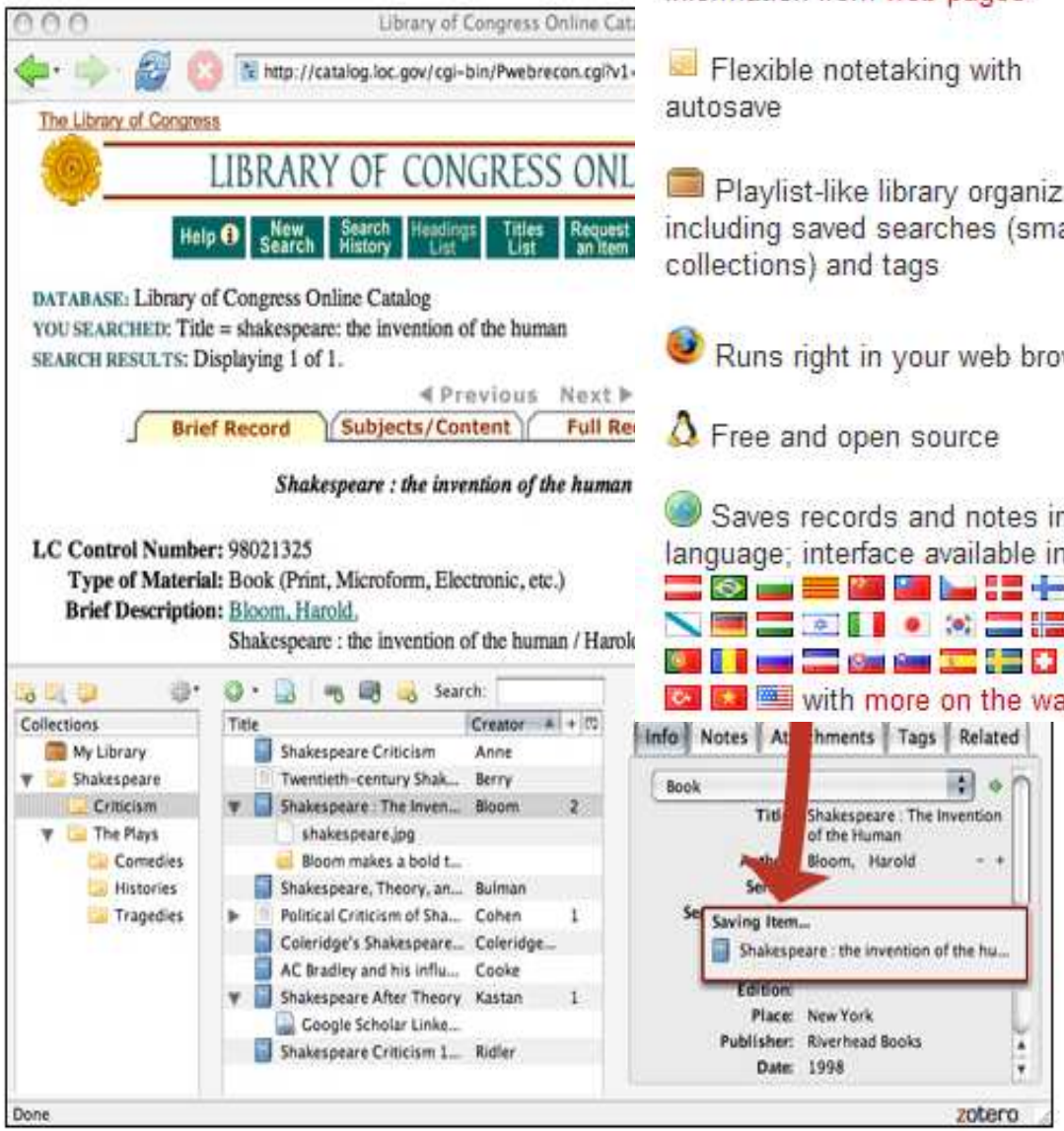
- "If you love books, and love people who love books, LibraryThing is for you. Start by using the service to catalog your book collection: Tag your books by topic, share your catalog with others, and then endlessly browse the titles that they have on their shelves. The utterly book obsessed can add the LibraryThing widget to a blog to show visitors what they have been reading lately." PCWorld (February 2006)

http://www.librarything.com/

CSIR our future through science

zotero

The next-generation research tool.



Automatic capture of citation information from web pages

Flexible notetaking with autosave

Playlist-like library organization, including saved searches (smart collections) and tags

Runs right in your web browser

Free and open source

Saves records and notes in any language; interface available in

with more on the way

Storage of PDFs, files, images, links, and whole web pages

Fast, as-you-type search through your materials

Platform for new forms of digital research that can be extended with other web tools and services

Formatted citation export (over 1100 styles and growing)

Integration with Microsoft Word and OpenOffice

Integration with WordPress and other blogging software

CSIR

our future through science

Librarian 2.0

Michael Stephens

- Recognises that the Library is human
- Plans for the users: breaks down barriers and allow users access wherever they are: home, work, commuting,
  school, or at the library.
- Embraces Web 2.0 tools recognizing how services might be enhanced by the Read/Write web and how new services might be born in a climate of collaboration.
- Controls techno-lust - does not buy/use technology for the sake of technology.

See: http://www.oclc.org/nextspace/002/3.htm

CSIR
our future through science

Web 2.0 Librarian cont

- Makes good, yet fast decisions: comfortable with perpetual beta & redesigns for ease of use, user involvement and easily added/re-configured pieces
- Trendspotter :
  - seeks out information and news that may impact future services
  - reads outside the profession and
  - watches for the impact of technology on users and new thinking on business
- Knows how to get content - understands that the future of libraries will be guided by how users access, consume and create content
- Listens to staff and users when planning, tells the stories of successes and failures, learns from both, celebrates those successes, allows staff time to play and learn, and never stops dreaming about the best library services

CSIR
our future through science

In the Job Jar of Oxford's Digital Repositories Research Co-ordinator

- Capture and document researcher's requirements for digital repositories services to handle research data
- Participate actively in the development of an interoperability framework for the federated digital repository at Oxford
- Make recommendations to improve and coordinate the provision of digital repository services for research data

CSIR
our future through science

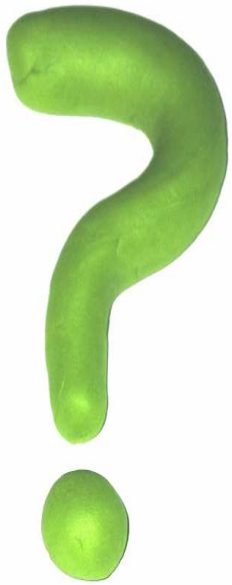
In the Job Jar of Oxford's Digital Repositories Research Co-ordinator 2

- Initiate and develop collaborations with the different repository activities already occurring to ensure that communication takes place in between them
- Raise awareness at Oxford of the importance and advantages of the active management of research data
- Communicate significant national and international developments in repositories to relevant Oxford stakeholders, in order to stimulate the adoption of best practices

CSIR
our future through science

Key question: do you manage a …

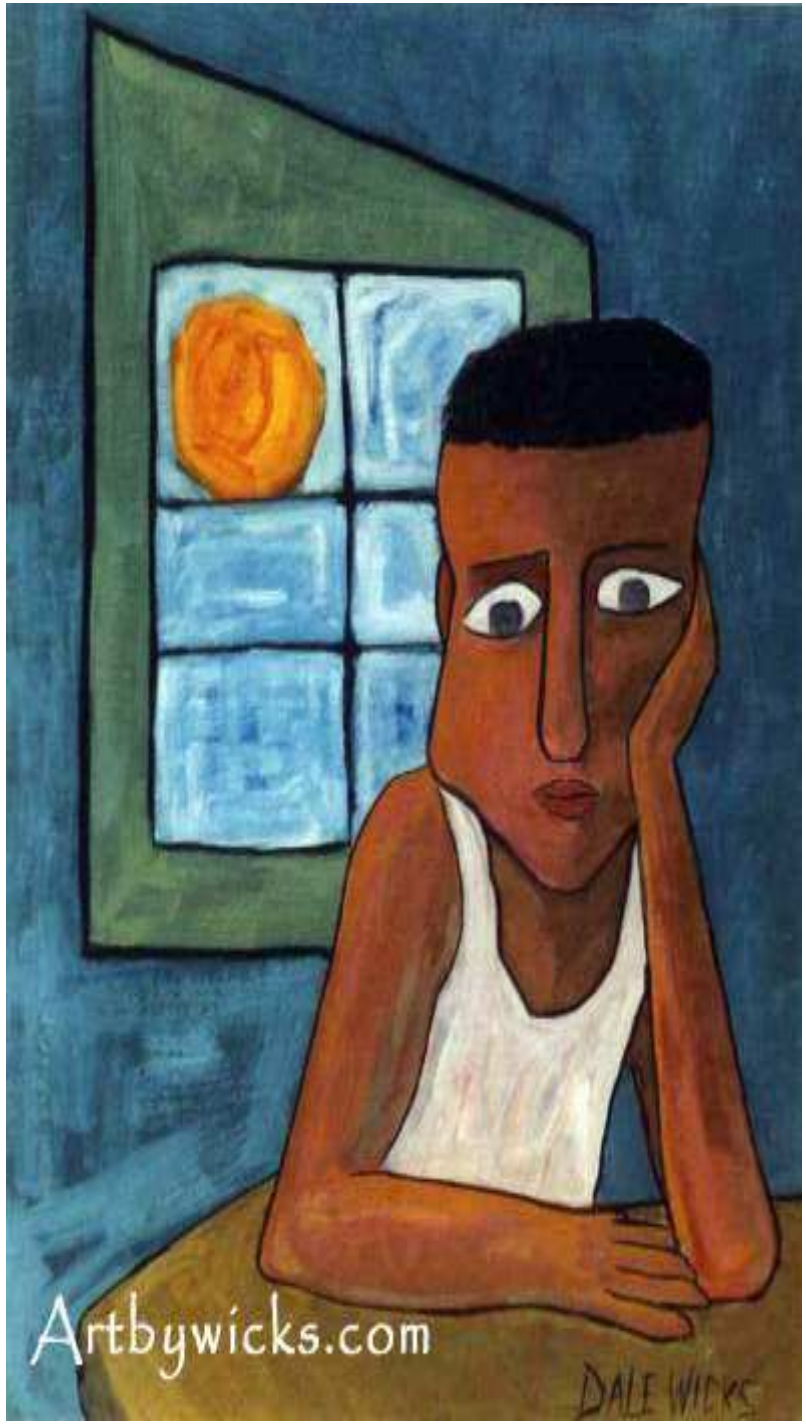
- Paper based library which also gives access to electronic resources?



- Electronic resources library which also gives access to paper collections?

- Challenge: perhaps the library should become more like a lab and less like a warehouse! (… with apology to Jim Jacobs – retired data-librarian University of California, San Diego)

CSIR
our future through science



"The world we have created today has problems which cannot be solved by thinking the way we thought when we created them."

-Albert Einstein

… also applicable to librarians

CSIR
our future through science

Scientific progress depends on speedy and open access to the full spectra of scientific data and derived products

(Hunter, 2006)

… remember …

CSIR
our future through science

CSIR
our future through science

Additional references

- **Gold, A 2007, Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians. D-Lib Magazine, V13, 9/10**
- **Gold, A 2007, Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries. D-Lib Magazine, V13, 9/10**
- Hunter, J. 2006, **Scientific Models – A User-oriented Approach to the Integration of Scientific Data and Digital Libraries. Available: http://www.valaconf.org.au/vala2006/papers2006/55_Hunter_Final.pdf Accessed 01/11/2008**
- **Lyon, L. 2007. Dealing with data: roles, rights, responsibilities and relationships. Available: http://www.jisc.ac.uk/media/documents/events/2007/06/liz_lyon.pdf Accessed: 01/11/2008**
- Osswald, A. 2007. E-science and information services: a missing link in the context of digital libraries. Online Information Review, Vol. 32 No. 4, 2008, pp. 516-523
- Martinez-Uribe, L. 2008. Scoping digital repository services for research data management. Available: http://www.google.co.za/url?sa=t&source=web&ct=res&cd=1&url=http%3A%2F%2Fora.ouls.ox.ac.uk%2Fobjects%2Fuuid%3A4e2b7e64-d941-4237-a17f-659fe8a12eb5%2Fdatastreams%2FATTACHMENT02&ei=FXcMSYDKGJ-k0QSth-XFBA&usg=AFQjCNEtVa6yBhSoontsorlkCuSN75235w&sig2=bdmTzbr-W0VmashGmMnSwg Accessed on 1/11/2008
- Schroeder, R. Fry, J, and De Beer, J. 2008. E-research infrastructures and scientific Communication. Available: http://www.iatul.org%2Fdoclibrary%2Fpublic%2FConf_Proceedings%2F2007%2FSchroeder_R_full.pdf&ei=7YQVSc7IJ4ie1wat6JDmCg&usg=AFQjCNGJ2sksjFkWUOyYhPqXJhJFdW5aAA&sig2=HTs8zIFed-l2YoCM8CCF4w
- Stoker, J. 2008. eResearch: Access and Support to University Researchers. 29th annual IATUL Conference. Available: http://www.iatul.org/doclibrary/public/Conf_Proceedings/2008/JudyStokker.pdf

CSIR
our future through science