



Neil Clarke
Director, ITS Research Support Services
11 February 2009

Monash Large Research Data Store (LaRDS)

Supporting the Data Lifecycle 11-Feb-2009



Australian Government
Australian Research Council

What is LaRDS ?

- LaRDS is Monash's central '*petascale*' research data store
- A petabyte is 1,000,000,000,000,000 bytes
 $\approx 250,000$ DVDs

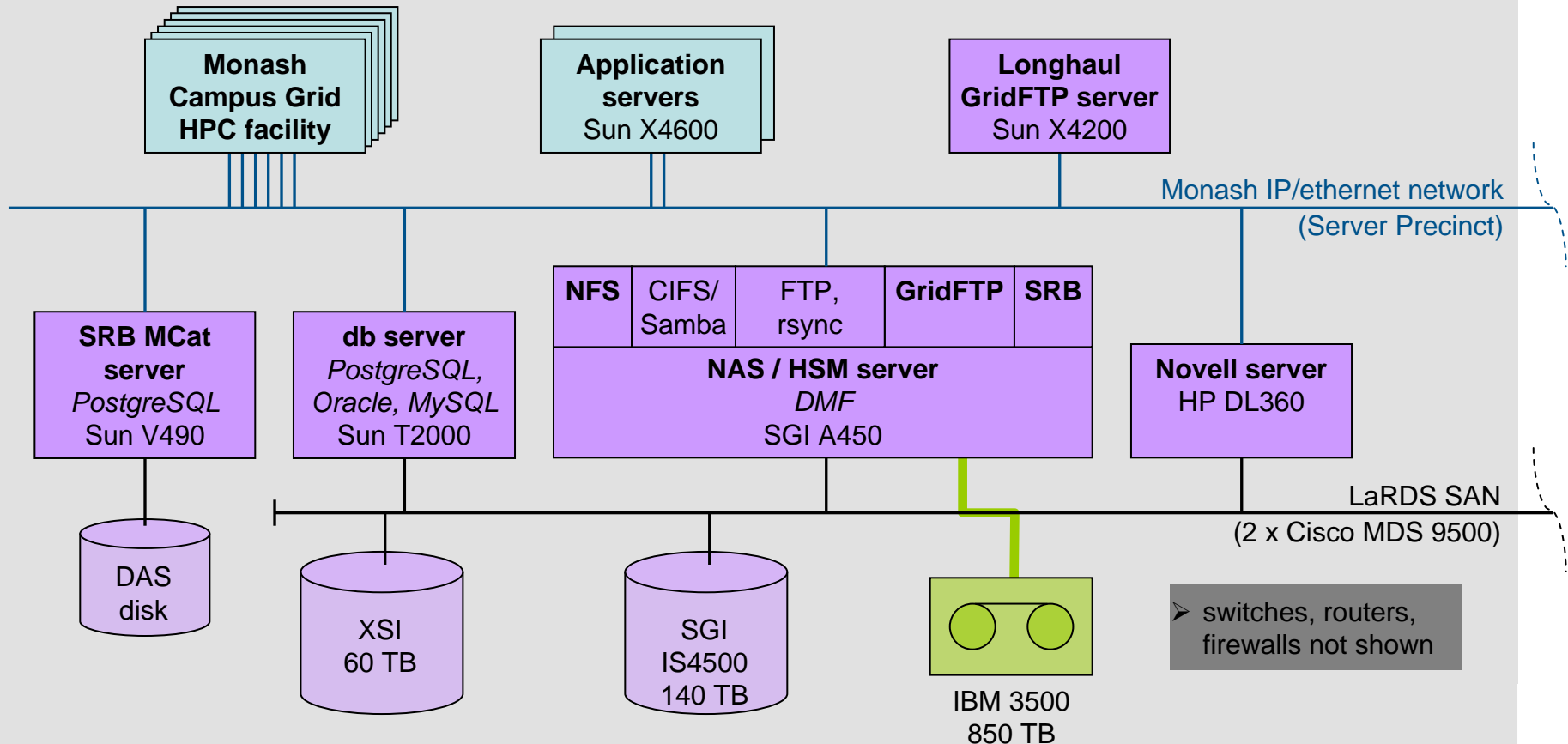
LaRDS drivers

- Improve the accessibility, preservation and reusability of research data
- Improve the management and organization of research data and researcher practice around data
- LaRDS approach: A spectrum of solutions, in small steps

What is LaRDS ?

- Not just a single... service, 'file system', GUI or software application
- LaRDS is a common infrastructure supporting a multiplicity of applications and access modes
- LaRDS can be used many different ways, and 'looks' very different in each
- All files in LaRDS are backed-up to tape across both Monash Data Centres

LaRDS physical server architecture



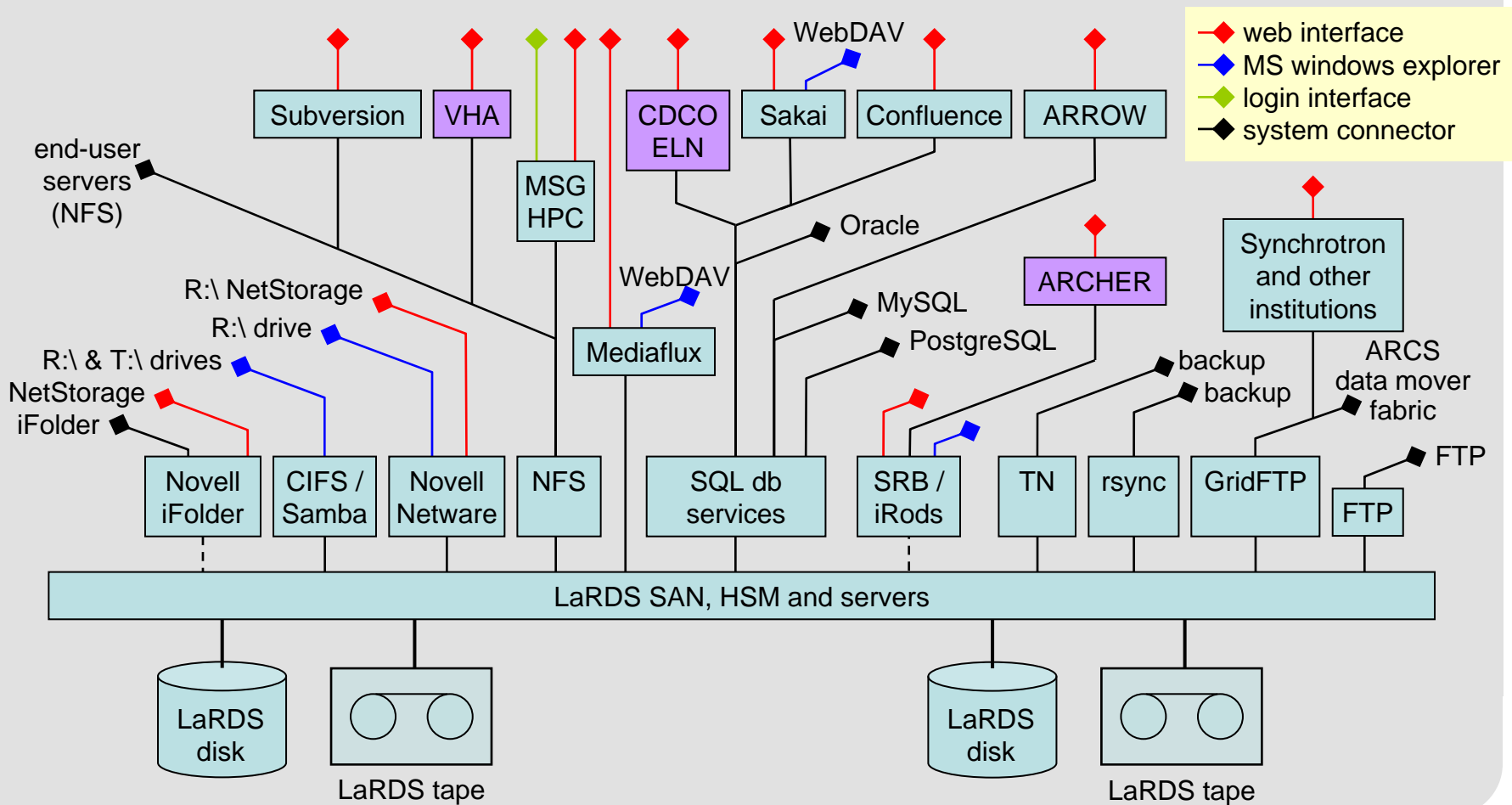
- All storage elements are replicated identically across both Data Centres

What can LaRDS 'look like' ?

LaRDS has a number of separate 'areas':

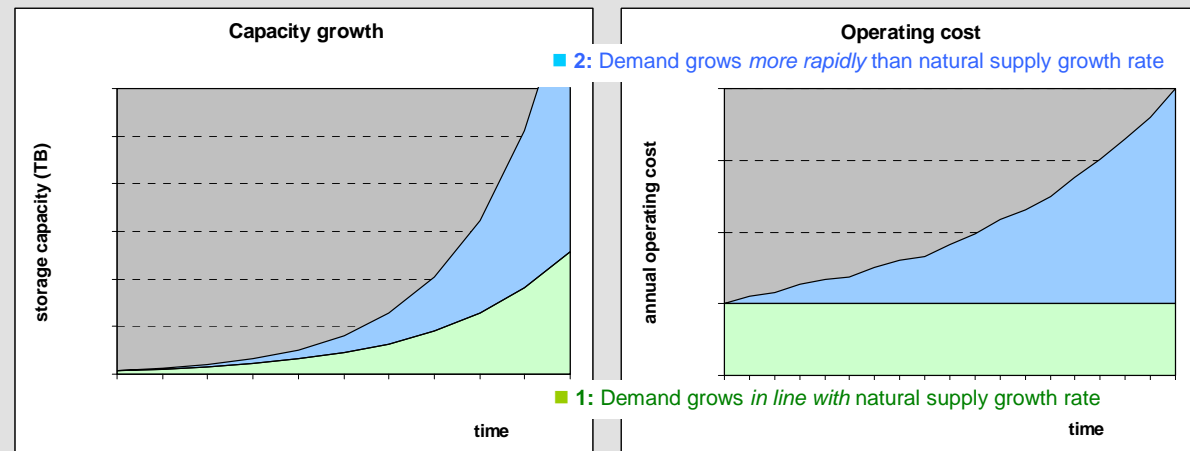
- Network drive:
 - NFS (unix), Samba (MS), Novell R:\ drive
- Database:
 - Oracle, MySQL, PostgreSQL, MS-SQL, SRB
- Digital Asset Management System (DAMS):
 - Mediaflux
- Collaboration (file sharing) environment:
 - Sakai VRE, Confluence wiki
- Custom Laboratory Information Management Systems (LIMS):
 - [CDCO:] idbsELN
 - [MISCL:] M-eNotes
- Custom applications:
 - Shoah Foundation: Visual History Archive
 - ARCHER tools
- Published repositories:
 - ARROW
- ARCS data mover fabric:
 - gridFTP, iRods [coming soon]
- Backup of local server systems:
 - rsync, TimeNavigator

LaRDS access points & connectors



LaRDS sustainable funding model

- Initial capital: Monash and government funds to establish LaRDS
- Ongoing: Exponentially growing capacity for a fixed annual operating cost



- Two tier funding model:
 - Green: Centrally funded (no volume charging)
 - Blue: User pays (marginal cost per TB)

LaRDS allocations Jan 2009

+ more will be added during the year

Faculty	Disk	Tape	Total
Art & Design	6	24	30
Arts	36	144	180
Business & Economics	31	124	155
Education	14	56	70
Engineering	46	184	230
Information Technology	14	56	70
Law	5	20	25
Medicine, Nursing & Health Sciences	178	712	890
Pharmacy	11	44	55
Science	36	144	180
Gippsland	10	40	50
DVC-R (MUARC, MCSS, MCEM, etc.)	13	52	65
Total	400	1600	2000

LaRDS capacity

- **Current capacity:**
 - + 400 TB on-line disk
 - + 1690 TB near-line tape ... and growing ...
- **Aim: Keep capacity ahead of demand**
 - So far: ‘high achievers’ have not been charged the user pays per-TB rate

LaRDS AUP & SLA

- Use of LaRDS is welcomed and encouraged by all Monash researchers, HDR students and research collaborators at other institutions
- Simple modification of existing Monash internal AUP:
 - the Monash participant (team leader) is responsible for actions of external collaborators
 - SLA for external collaborators is “all care no responsibility”
 - reflects that this is not a paid “commercial” service

Important points

- LaRDS is different things to different people
- LaRDS is secure:
 - all data is backed up to two data centres
 - data is only available to the people you want it be available to :
 - individual
 - workgroup
 - public

Questions ...

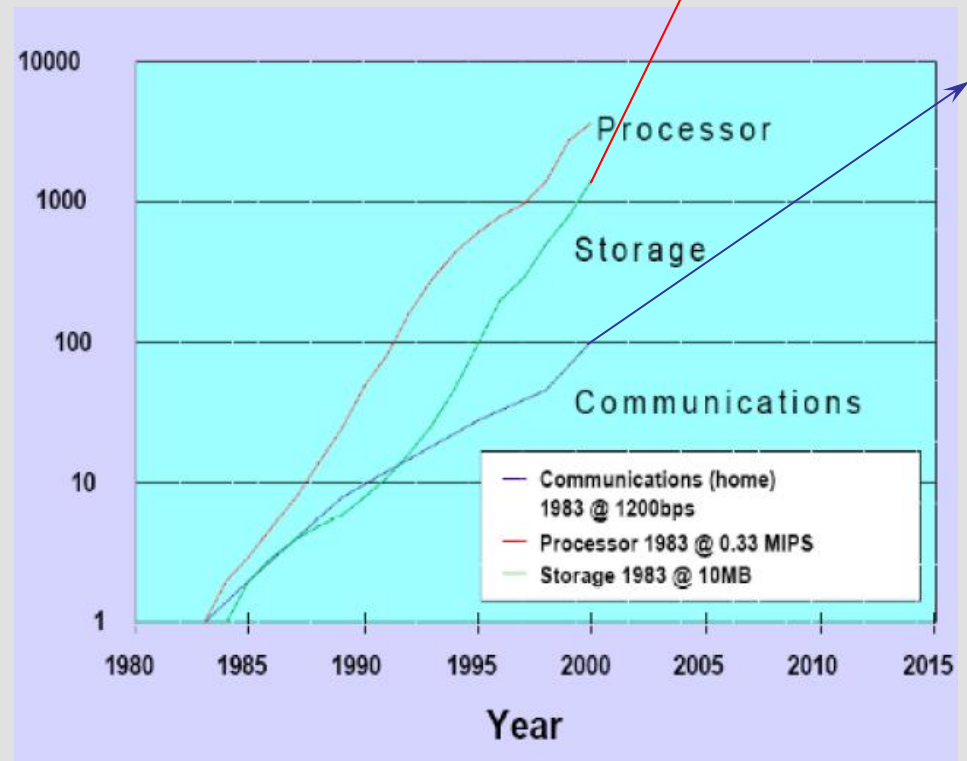
www.monash.edu/eresearch/services/lards/

Neil.Clarke@its.monash.edu

Data is growing fast - moving it is hard

- **And will get harder:**
- Doubling times for:
 - data: <12 months
 - network: ~18 months
- Data is accelerating away from network

(Moore's Law)

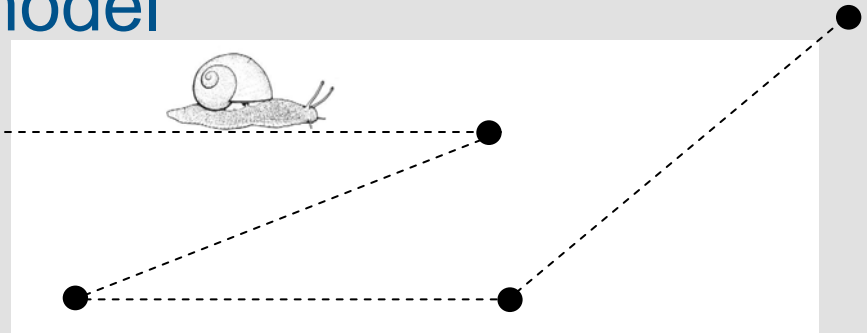


Moving data is hard

- traditional workflow model

- multiple moves

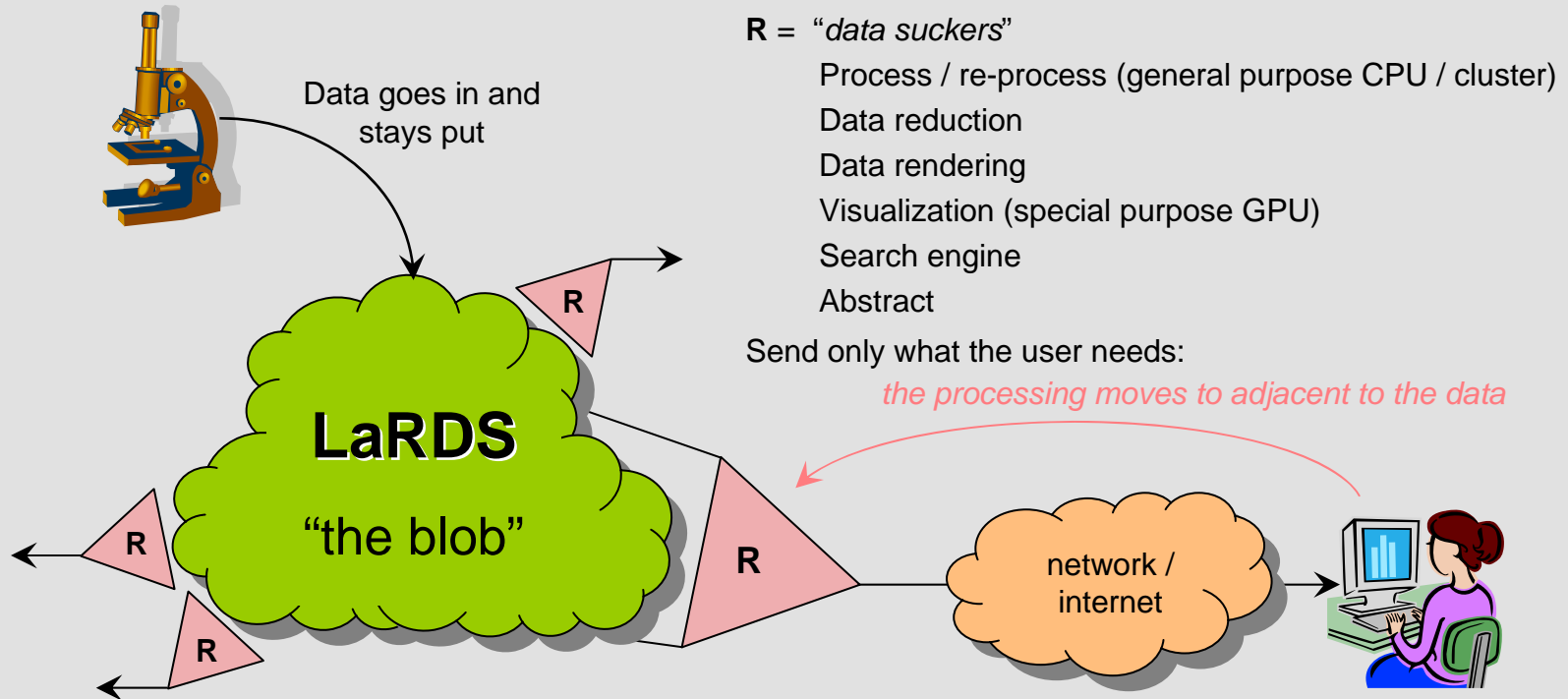
VS



- “park and use” model

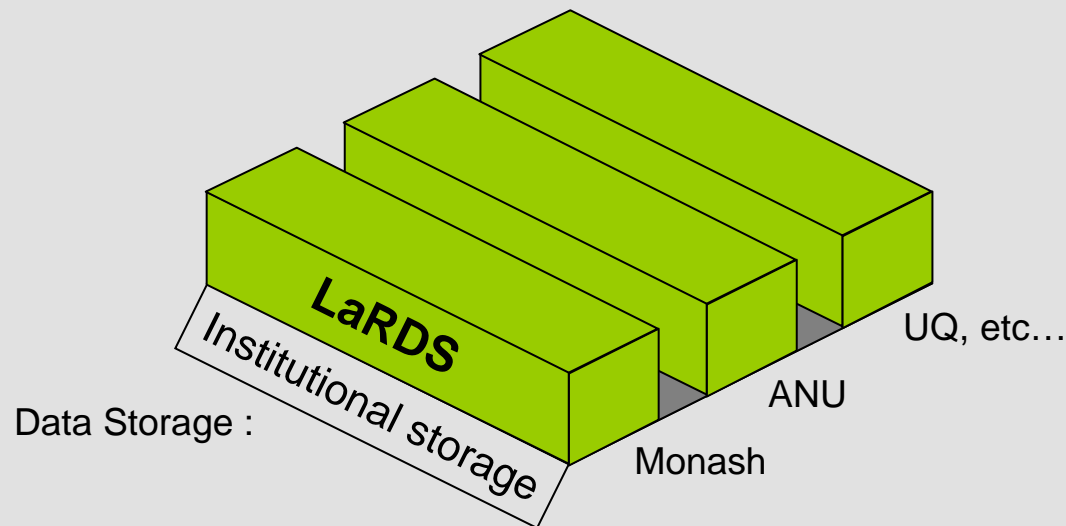
- near its source /
- near where it will mostly be used

Data “Park and Use” model

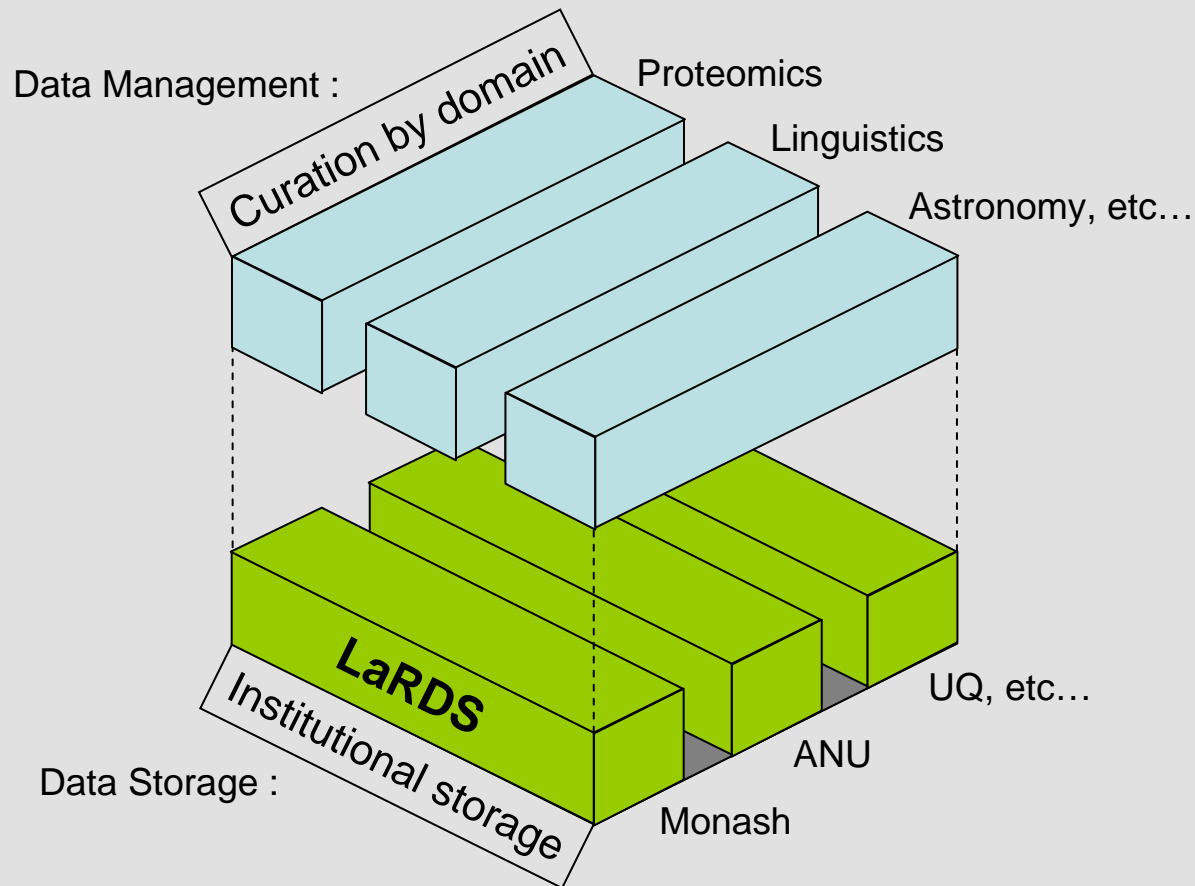


A dilemma

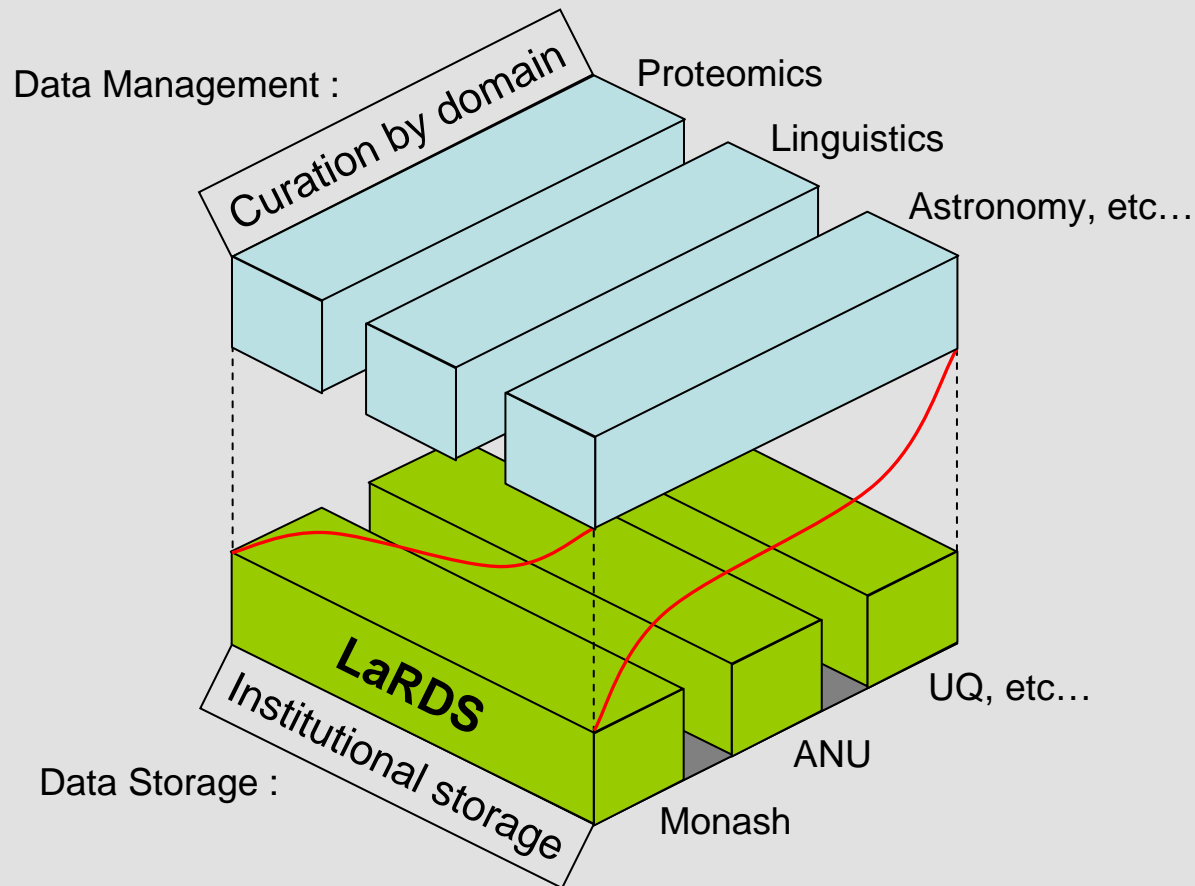
- Data storage is by institution, ...but...
- Data management and curation is by research domain



A dilemma

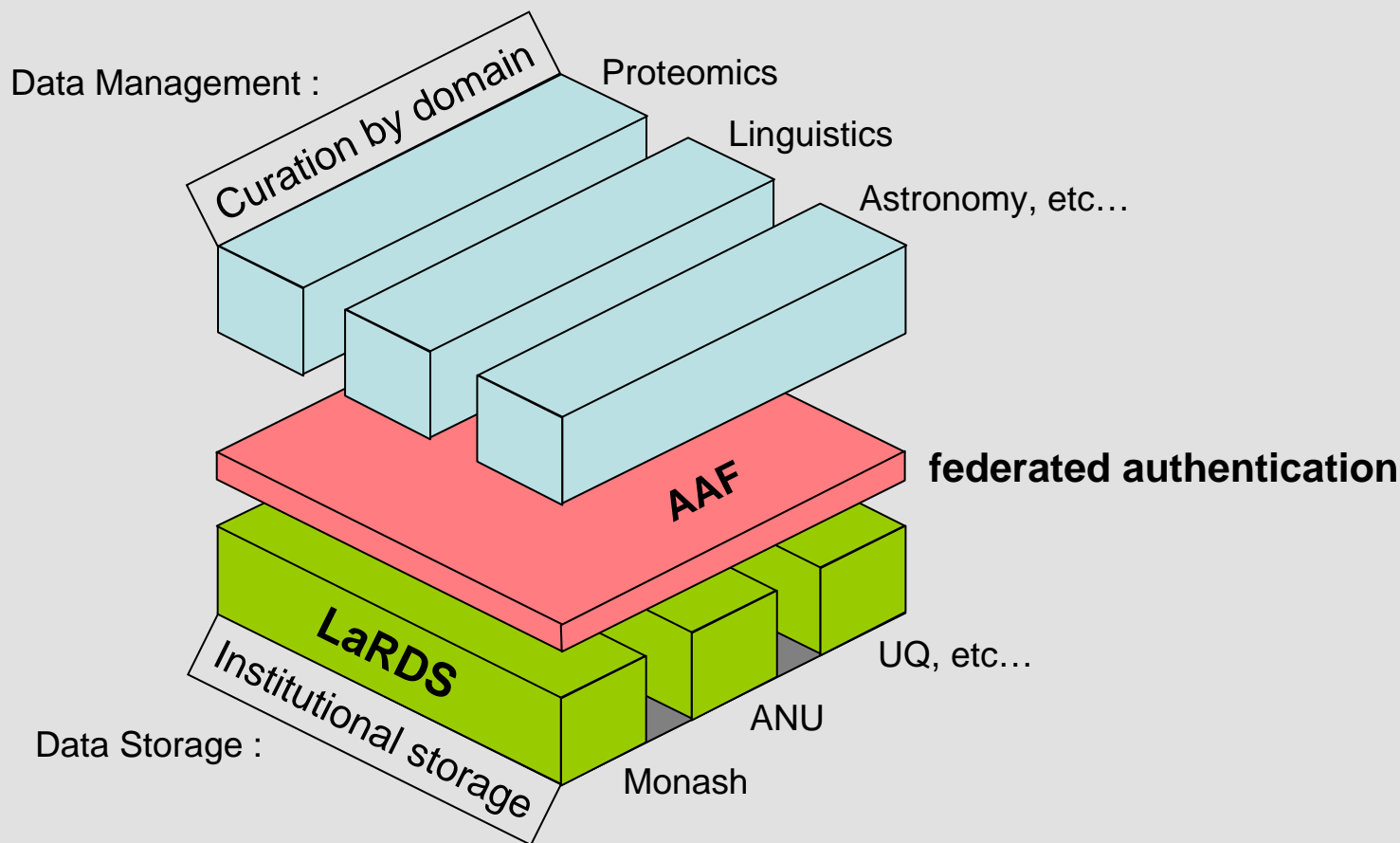


A dilemma – a twist

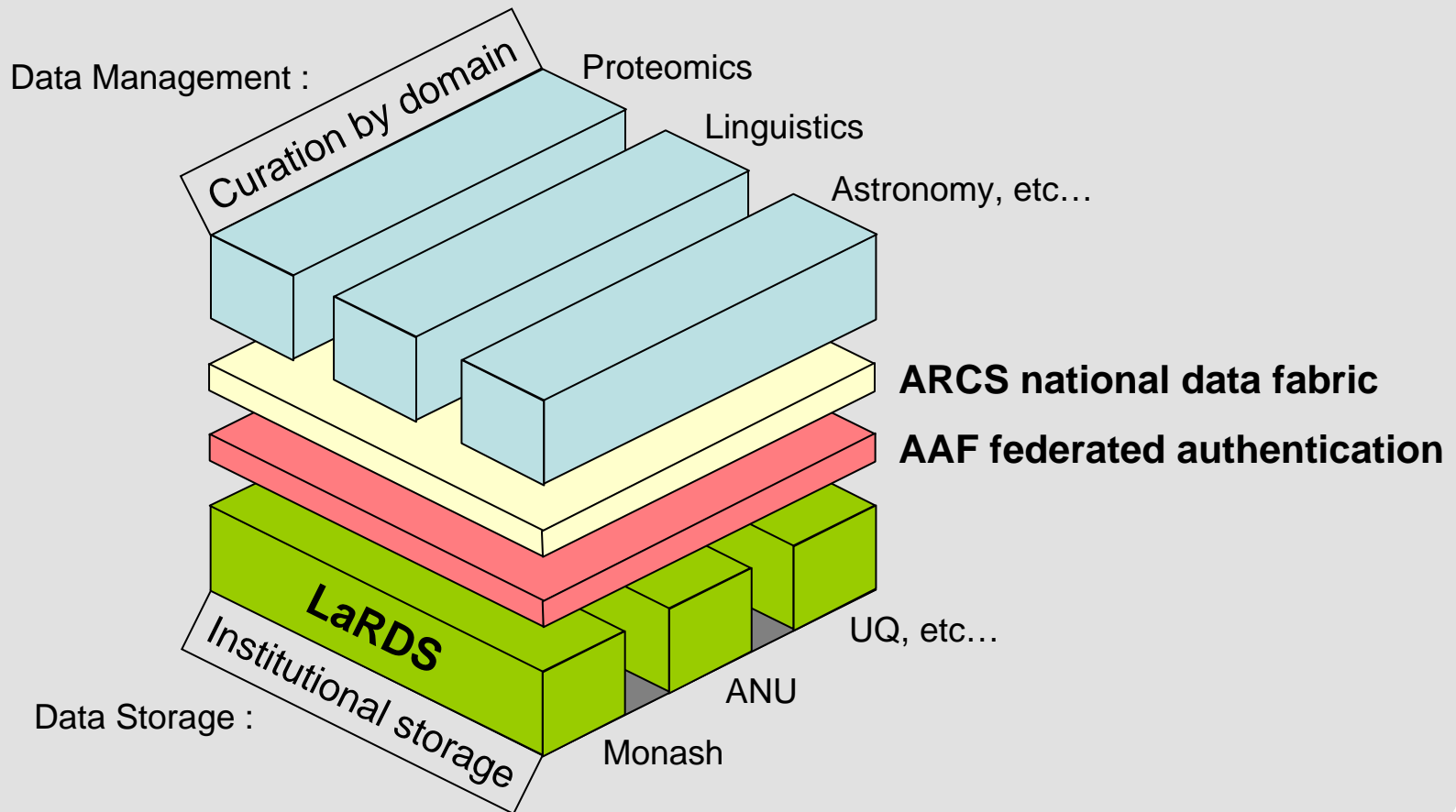


Federated authentication

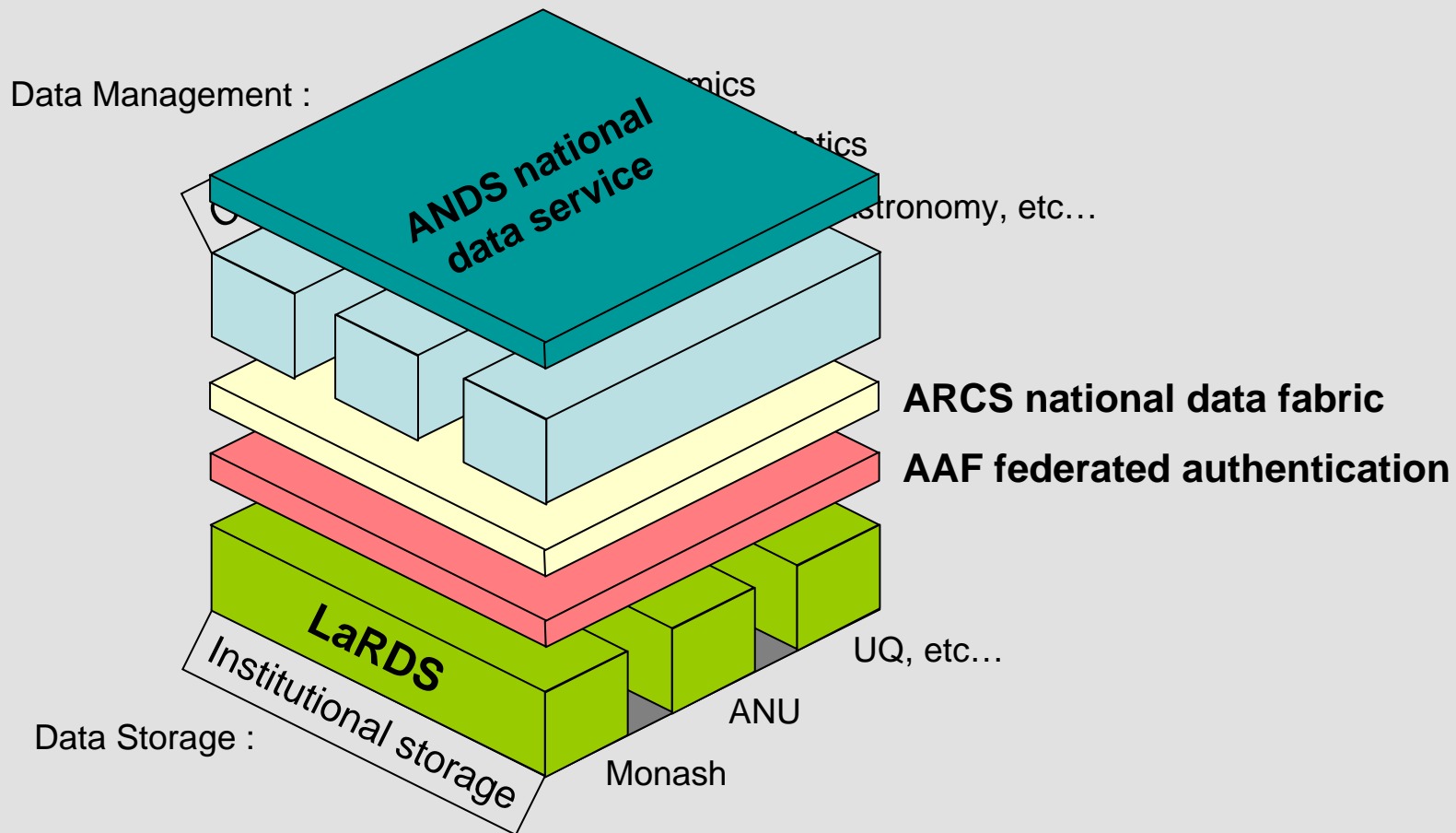
– the Australian Access Federation (AAF)



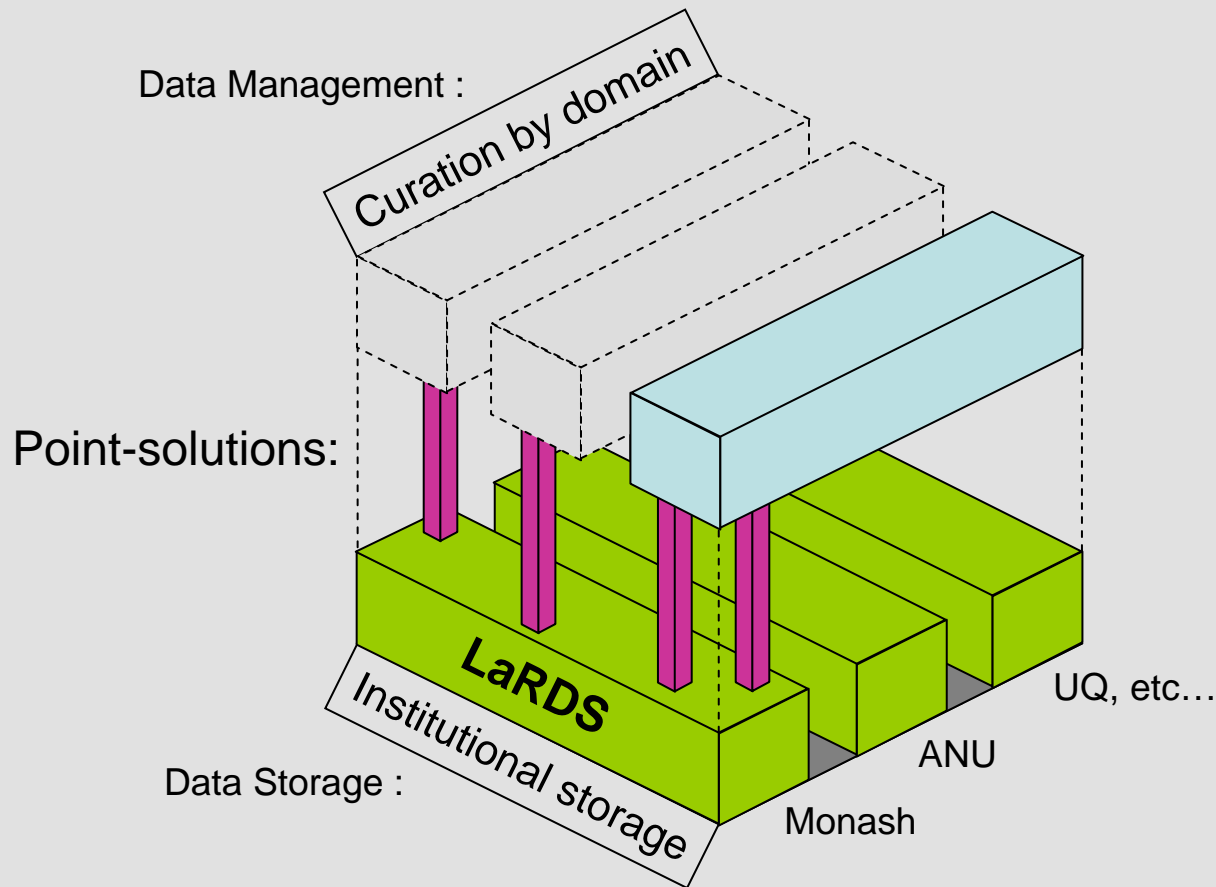
Australian Research Collaboration Service (ARCS) national data fabric



Australian National Data Service (ANDS)



But: Where are we now ?



Context: What's going on in e-Research ?

- Custom-built point-solutions to meet the specific needs of individual research teams ('growing downwards')
- Research domain-specific data management standards ('growing from the middle')
- Custom and generic collaboration environments ('growing from the middle')
- Generic e-infrastructure ('growing upwards')

