Meeting skills demand for data stewardship

Liz Stokes, Skilled Workforce Team, ARDC
eResearch Skilled Workforce Summit 29-30 July 2019
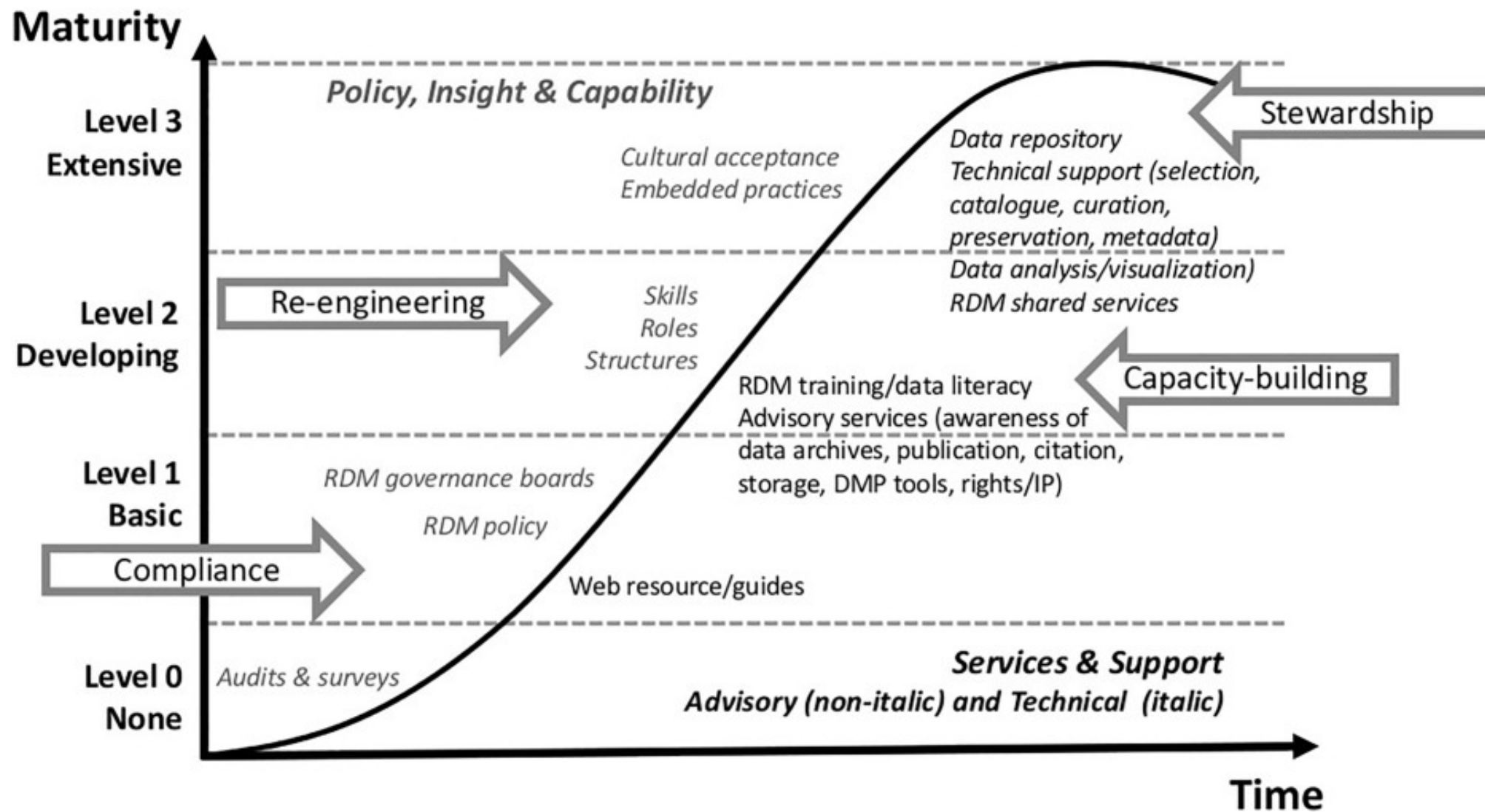New Law Building, The University of Sydney, Australia

ARDC
Australian Research Data Commons

NCRIS
National Research
Infrastructure for Australia
An Australian Government Initiative

What is data stewardship?

- Repositories
- Data infrastructure –platforms, services
- Training and advocacy – generalist and specialist
- Ethics and research integrity
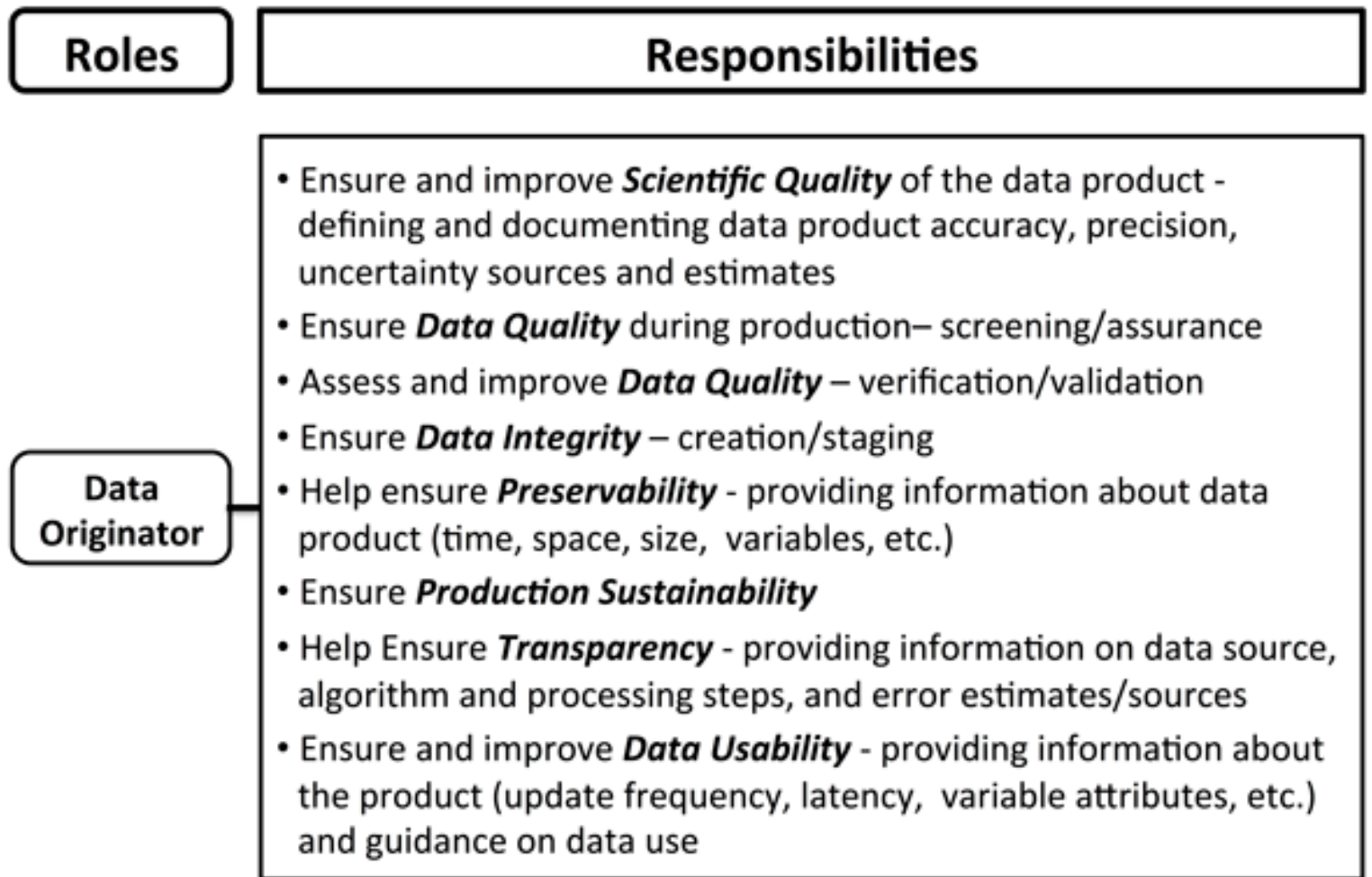- Data governance - DMPs

Australian Research Data Commons

Cox, A. M., Kennan, M. A., Lyon, L. and Pinfield, S. (2017), Developments in research data management in academic libraries: Towards an understanding of research data service maturity. Journal of the Association for Information Science and Technology, 68: 2182-2200. doi:10.1002/asi.23781

"Technologies can be consolidated, and processes automated, but collection, creation, curation and research use/reuse of data is ultimately dependent on the domain and disciplinary expertise of the humans who know them best."

Hervé L'Hours, UK Data Archive
https://www.linkedin.com/pulse/formal-certification-data-repositories-key-fair-eosc-herv%C3%A9-l-hours/

Roles	Responsibilities
Data Originator	• Ensure and improve **Scientific Quality** of the data product - defining and documenting data product accuracy, precision, uncertainty sources and estimates • Ensure **Data Quality** during production– screening/assurance • Assess and improve **Data Quality** – verification/validation • Ensure **Data Integrity** – creation/staging • Help ensure **Preservability** - providing information about data product (time, space, size, variables, etc.) • Ensure **Production Sustainability** • Help Ensure **Transparency** - providing information on data source, algorithm and processing steps, and error estimates/sources • Ensure and improve **Data Usability** - providing information about the product (update frequency, latency, variable attributes, etc.) and guidance on data use

Peng, G. et al, 2016, Scientific Stewardship in the Open Data and Big Data Era — Roles and Responsibilities of Stewards and Other Major Product Stakeholders, D-Lib Magazine Vol 22 No 5/6, 10.1045/may2016-peng
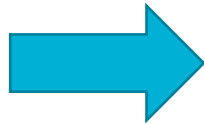
Diversified data stewardship

Roles	Responsibilities
Data Steward	• Ensure *Data Integrity* – ingest and archive • Ensure and improve *Data Provenance* and *Traceability* • Improve *Data Quality* metadata • Ensure and improve archiving requirements
Scientific Steward	• Assess/improve *Data Quality* – Evaluation/verification • Promote and improve *Data Usability* – Characterization • Help ensure and improve *Data Quality* metadata • Ensure and improve data quality and usability requirements
Technology Steward	• Ensure *Data Integrity* – ingest, archive retrieval, data access, and file system and technology upgrade • Ensure and Improve *Data Accessibility* and *Discoverability* • Promote and improve *Data Interoperability* • Ensure and improve software and system requirements

Australian Research Data Commons

Diversified data stewardship

Repository / operational
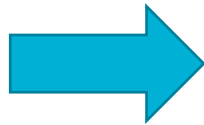
Embedded data steward, data champion

Research software engineer

Roles	Responsibilities
Data Steward	• Ensure **Data Integrity** – ingest and archive • Ensure and improve **Data Provenance** and **Traceability** • Improve **Data Quality** metadata • Ensure and improve archiving requirements
Scientific Steward	• Assess/improve **Data Quality** – Evaluation/verification • Promote and improve **Data Usability** – Characterization • Help ensure and improve **Data Quality** metadata • Ensure and improve data quality and usability requirements
Technology Steward	• Ensure **Data Integrity** – ingest, archive retrieval, data access, and file system and technology upgrade • Ensure and Improve **Data Accessibility** and **Discoverability** • Promote and improve **Data Interoperability** • Ensure and improve software and system requirements

ARDC

Australian Research Data Commons

Peng, G. et al, 2016, **Scientific Stewardship in the Open Data and Big Data Era — Roles and Responsibilities of Stewards and Other Major Product Stakeholders, D-Lib Magazine Vol 22 No 5/6, 10**.1045/may2016-peng

But wait, don't forget!

Repositories, publishers ➡

Data governance ➡

Research project manager ➡

Researchers, could be anywhere ➡

Roles	Responsibilities
Data Distributor	• Ensure and improve *Representation* of data quality information • Ensure and improve *Traceability* of data quality information • Ensure user feedback • Help improve data quality and usability requirements
Sponsor	• Define *Data Quality* and *Usability* requirements • Require data quality oversight and monitoring • Encourage *Transparency* in data quality procedures and practices
Manager	• Help increase awareness of *Data Quality* and *Usability* • Help improve data quality and usability requirements • Help ensure *Data Interoperability*
End-User	• Request *Transparency* in data quality procedures and practices • Request *Provenance* of the data product • Request evaluation results of product, stewardship, and service maturity of the data product • Provide feedback on *Quality* and *Usability* of the data product

ARDC
Australian Research Data Commons

- Research information manager
- Data stewardship coordinator
- Research data management coordinator
- Specialist in scientific information for natural sciences
- Senior research data manager

- Data steward
- Data manager
- Research Data Manager
- Consultant Data Management & Designer
- Data Scientist
- Data Scientists: statistical analysts, machine learning specialists & scientific programmers
- Bioinformatician

Policy, strategy and coordination

Embedded and operational

Generic and advice

- Data steward (4x)
- Research Data management Specialist (3x)
- Research Data manager junior or medior
- Research Data Management Consultants
- Data Science Research Assistant

ARDC
Australian Research Data Commons

Data stewardship is everyone's business.

Data stewardship has something for everyone.

Australian Research Data Commons

Data stewardship use cases

- As a **senior administrative data steward**, I want to ensure research data and infrastructure are FAIR so that I can mitigate risk and monitor for integrity.

- As a **graduate research training coordinator,** I want discipline agnostic examples that teach the FAIR data principles which I can use in my training programs to facilitate better RDM practice at my institution.

- As a **faculty dean**, I want graduate attributes that prove our graduates have a data advantage to our industry partnerships.

- As a **repository manager**, I want a community or practice to engage in next generation repository design which supports my case for sustained funding and a future roadmap pitch to

the senior administration of my institution.

- As an **eResearch analyst,** I need to provide specialised data management advice to researchers in their grant and ethics applications.

- As a **faculty liaison librarian**, I want introductory training in research data management, so that I can facilitate access to tools and best practice in research data management to researchers and HDR students within my faculty.

- As a **member of an Ethics approvals committee,** I want to ensure researchers can manage and share their data ethically and with integrity.

Research Enterprise Management

Team Data Management

Personal Information Management

Figure 1: Domains Represented in the Matrix

Australian Research Data Commons
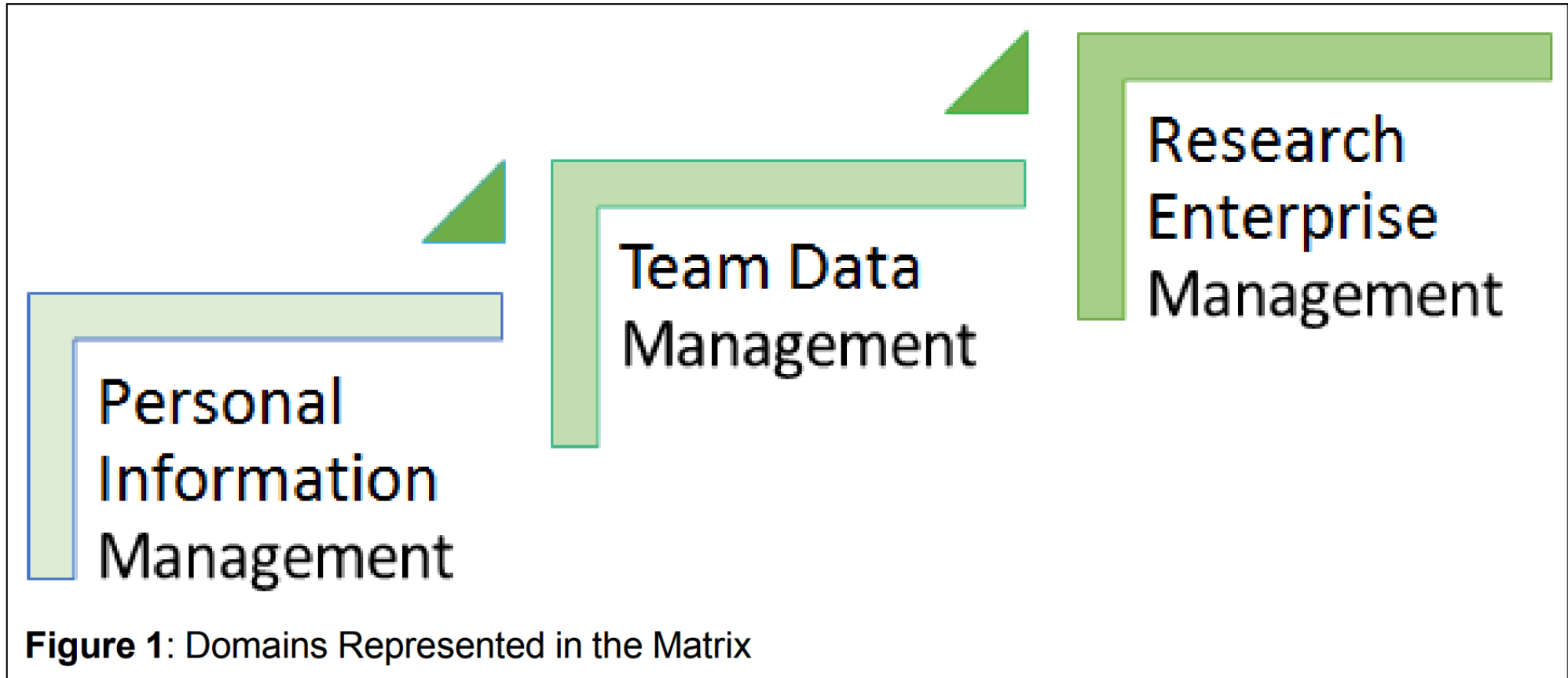
Skills training to foster communities of practice

- Need to collaborate and share
- Situate learning to address local needs
- Internal processes could benefit from data skills
- Need to be not alone – people over technology

In meeting the skills demand

- Can we agree on core competencies for different levels?
- Can we join the dots from students to researchers?
- Can we take our own medicine?
- Can we build and sustain communities of practice?
- How do we relate to industry and government sectors
- How can we reward, motivate and incentivise data stewardship?

ARDC
Australian Research Data Commons
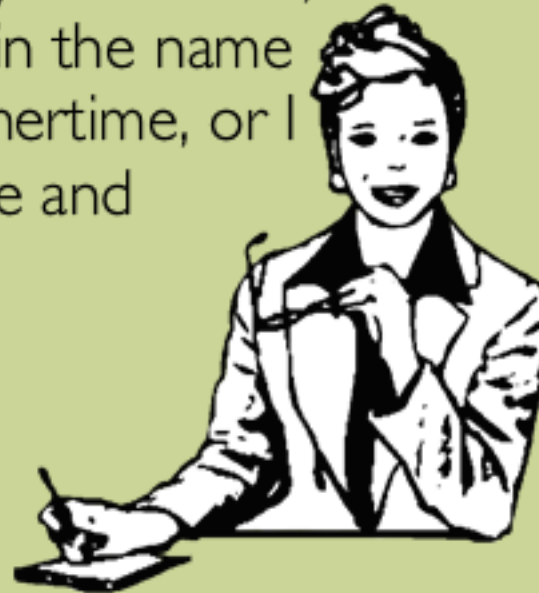
Liz.stokes@ardc.edu.au

Tw: @ragamouf