Data & Digital Research Capabilities (or Skills)

WHAT THEY ARE, HOW TO GET THEM, WHERE YOU CAN GO

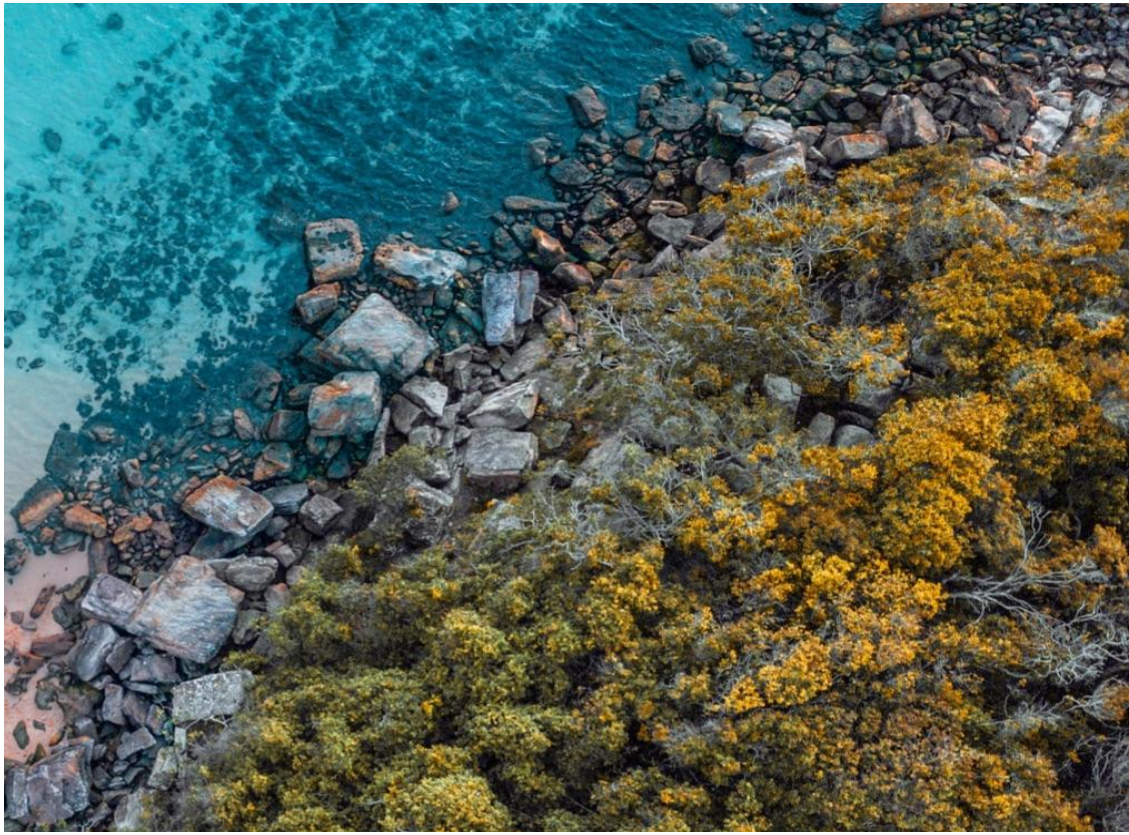
09 February 22

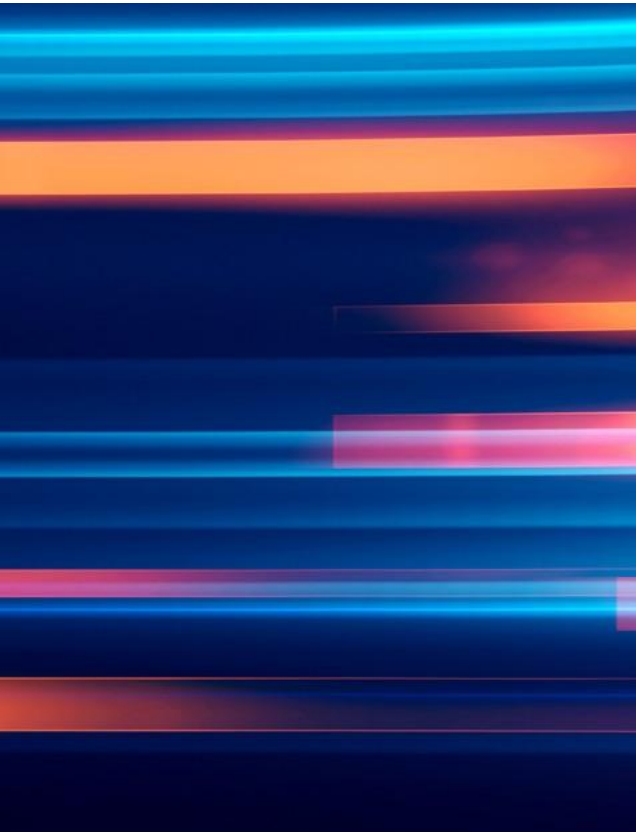
PRESENTED BY

Kathryn Unsworth - Manager, Skilled Workforce Development, ARDC

ACKNOWLEDGEMENT OF COUNTRY

We acknowledge and celebrate the First Australians on whose traditional lands we meet, and we pay our respect to their elders past, present and emerging.





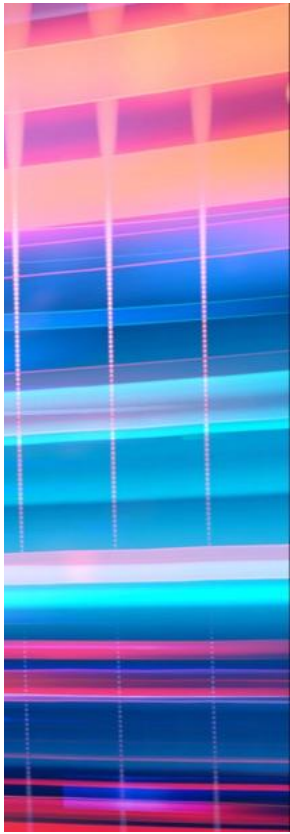
Australian Research Data

Purpose

To provide Australian researchers with competitive advantage through data.

Mission

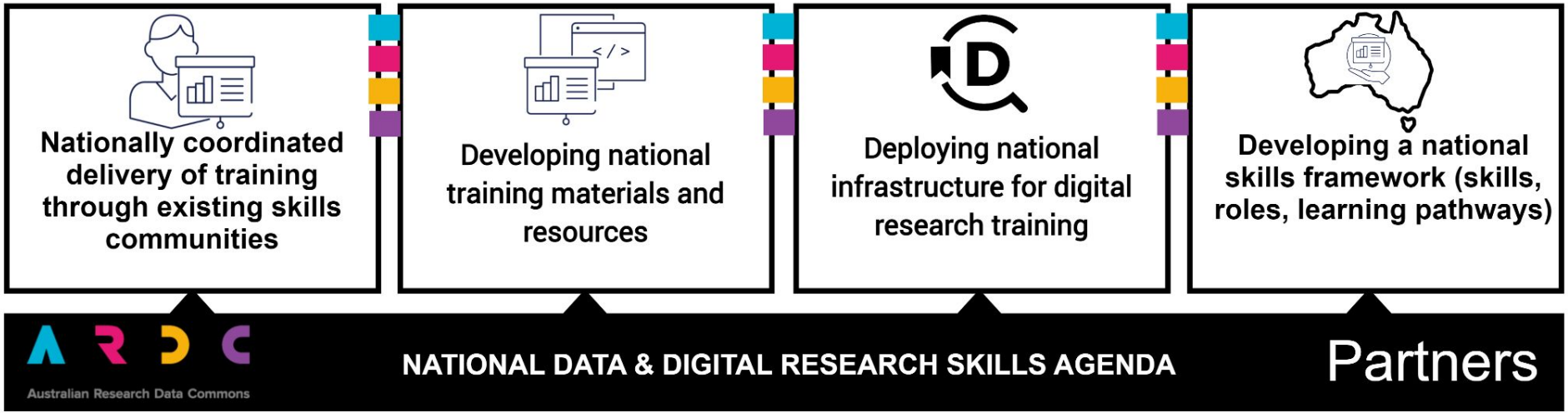
To accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.



NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

ARDC is enabled by NCRIS

A COMMUNITY DRIVEN, NATIONALLY COORDINATED, SECTOR WIDE APPROACH TO TRAINING

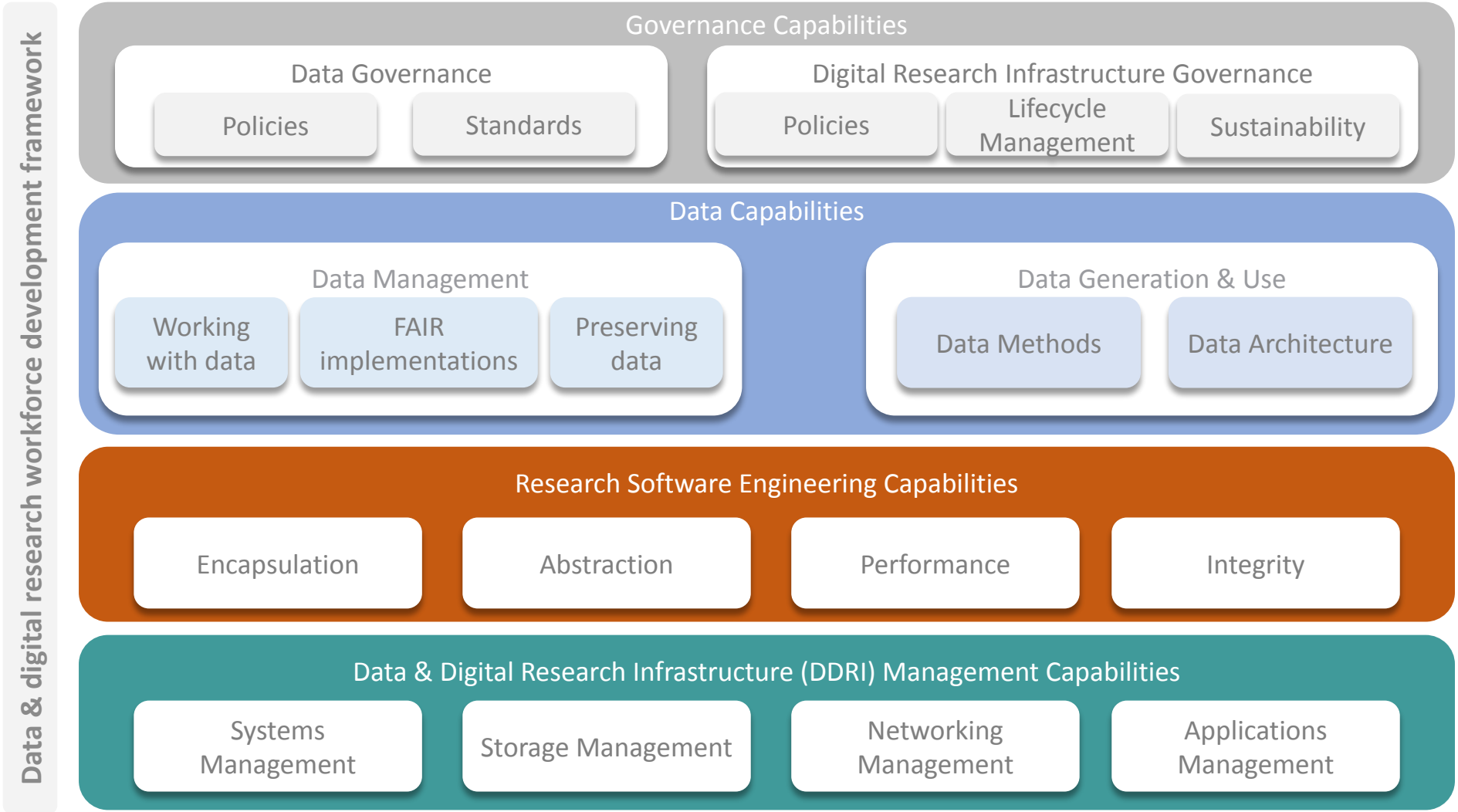


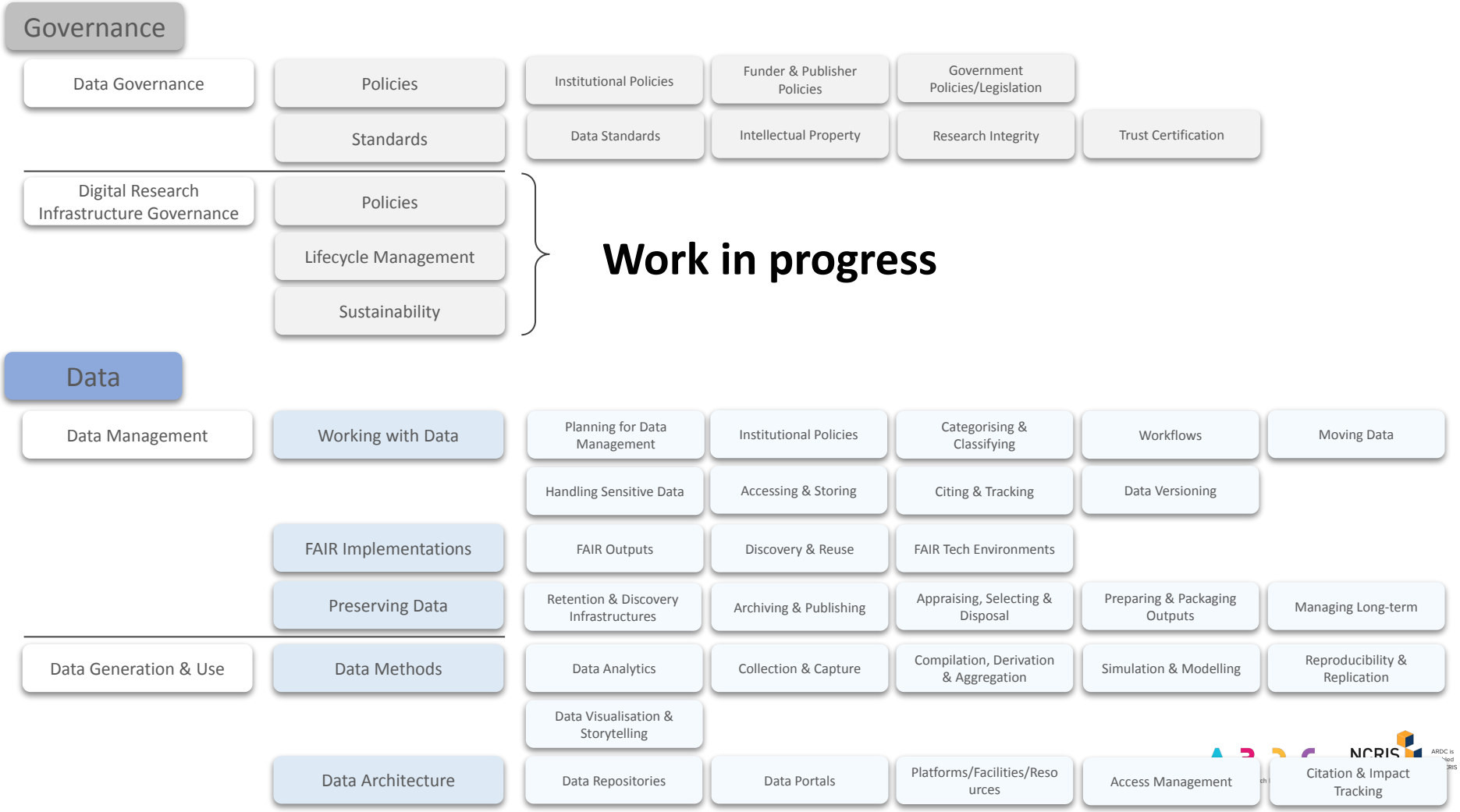
What they are

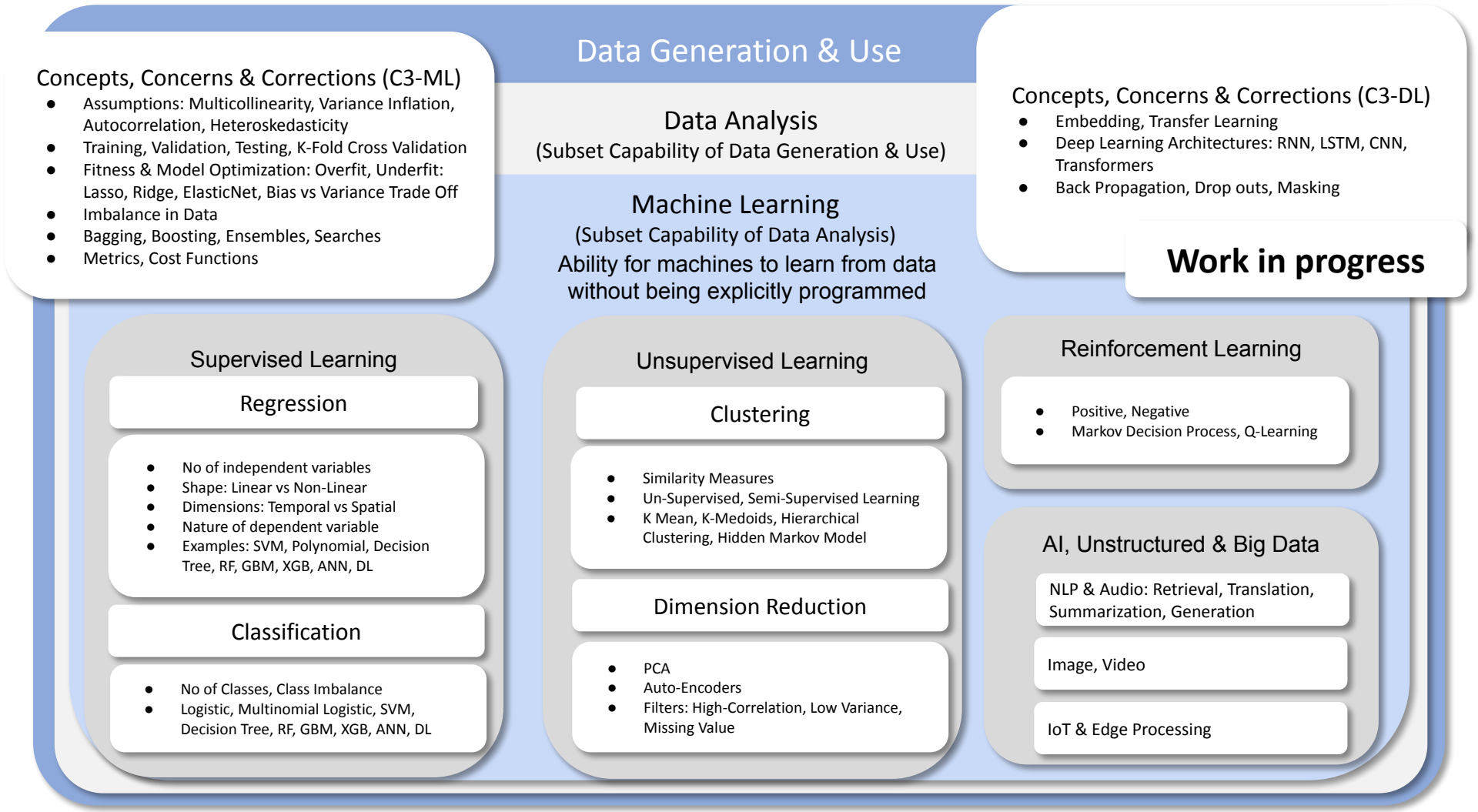
EXPLORING THE ARDC CAPABILITIES/SKILLS LANDSCAPE

WHY CREATE A CAPABILITIES - SKILLS LANDSCAPE?

Image by Peggy und Marco Lachmann-Anke from Pixabay


Data & digital research workforce development framework
Governance Capabilities
Data Governance
Policies
Standards
Digital Research Infrastructure Governance
Policies
Lifecycle Management
Sustainability
Data Capabilities
Data Management
Working with data
FAIR implementations
Preserving data
Data Generation & Use
Data Methods
Data Architecture
Research Software Engineering Capabilities
Encapsulation
Abstraction
Performance
Integrity
Data & Digital Research Infrastructure (DDRI) Management Capabilities
Systems Management
Storage Management
Networking Management
Applications Management


Governance
Data Governance
Policies
Institutional Policies
Funder & Publisher Policies
Government Policies/Legislation
Standards
Data Standards
Intellectual Property
Research Integrity
Trust Certification
Digital Research Infrastructure Governance
Policies
Lifecycle Management
Sustainability
Work in progress
Data
Data Management
Working with Data
Planning for Data Management
Institutional Policies
Categorising & Classifying
Workflows
Moving Data
Handling Sensitive Data
Accessing & Storing
Citing & Tracking
Data Versioning
FAIR Implementations
FAIR Outputs
Discovery & Reuse
FAIR Tech Environments
Preserving Data
Retention & Discovery Infrastructures
Archiving & Publishing
Appraising, Selecting & Disposal
Preparing & Packaging Outputs
Managing Long-term
Data Generation & Use
Data Methods
Data Analytics
Collection & Capture
Compilation, Derivation & Aggregation
Simulation & Modelling
Reproducibility & Replication
Data Visualisation & Storytelling
Data Architecture
Data Repositories
Data Portals
Platforms/Facilities/Resources
Access Management
Citation & Impact Tracking


Data Generation & Use
Concepts, Concerns & Corrections (C3-ML)
Assumptions: Multicollinearity, Variance Inflation, Autocorrelation, Heteroskedasticity
Training, Validation, Testing, K-Fold Cross Validation
Fitness & Model Optimization: Overfit, Underfit: Lasso, Ridge, ElasticNet, Bias vs Variance Trade Off
Imbalance in Data
Bagging, Boosting, Ensembles, Searches
Metrics, Cost Functions
Data Analysis
(Subset Capability of Data Generation & Use)
Machine Learning
(Subset Capability of Data Analysis)
Ability for machines to learn from data without being explicitly programmed
Concepts, Concerns & Corrections (C3-DL)
Embedding, Transfer Learning
Deep Learning Architectures: RNN, LSTM, CNN, Transformers
Back Propagation, Drop outs, Masking
Work in progress
Supervised Learning
Regression
No of independent variables
Shape: Linear vs Non-Linear
Dimensions: Temporal vs Spatial
Nature of dependent variable
Examples: SVM, Polynomial, Decision Tree, RF, GBM, XGB, ANN, DL
Classification
No of Classes, Class Imbalance
Logistic, Multinomial Logistic, SVM, Decision Tree, RF, GBM, XGB, ANN, DL
Unsupervised Learning
Clustering
Similarity Measures
Un-Supervised, Semi-Supervised Learning
K Mean, K-Medoids, Hierarchical Clustering, Hidden Markov Model
Dimension Reduction
PCA
Auto-Encoders
Filters: High-Correlation, Low Variance, Missing Value
Reinforcement Learning
Positive, Negative
Markov Decision Process, Q-Learning
AI, Unstructured & Big Data
NLP & Audio: Retrieval, Translation, Summarization, Generation
Image, Video
IoT & Edge Processing

Research Software Engineering

Encapsulation	Domain Knowledge	Modular Design for Reusability	Interfaces – Programming, Graphical & Web Services	Containerisation	
Abstraction	Domain Knowledge	Algorithm Design & Development	Data Structures	Computational Workflows??	Domain-specific Languages
Performance Optimisation	Domain Knowledge	Compile and/or Functional Languages	Parallelisation	HPC Development, Code Optimisation, Compilers	GPU Programming
Integrity	Domain Knowledge	Unit Testing & Other Testing Paradigms	Numerical Stability	Version Control	Documentation – Design, Specifications, Usage

Digital Research Infrastructure Management

Systems Management	Data Centre Inf Management	Disaster Recovery and Backups	Capacity Planning	System Administration	Helpdesk & Service Desk Management	Asset Lifecycle Management
Networking Management	Network Administration	Network Maintenance	Network Operation	Network Provisioning	Network Security	
Storage Management	Capacity Planning	Storage Protocols	Asset Management	Backup Recovery		
Application Management	Application Lifecycle Management	Application Performance Management	Application Service Management	Application Configuration		

Roles (generalised - bearing in mind people and organisations often cover more than one of these roles)

Data Owner:

Data owner refers to those who not only have possession of but also have responsibility for data. The control of data includes not just the ability to access, create, modify, package, derive benefit from, sell or remove data, but also the right to assign these access privileges to others.

Data Governor:

Defines the availability, usability, integrity and security of data, based on data standards and policies that also control data usage. Effective data governance ensures that data and data use are consistent and trustworthy and generate benefit for the data owner(s).

Data Steward:

A person responsible for keeping the quality, integrity, and access arrangements of data and metadata in a manner that is consistent with applicable law, institutional policy, and individual permissions. Data stewardship implies professional and careful treatment of data throughout all stages of a research process. A data steward aims at guaranteeing that data is appropriately treated at all stages of the research cycle (i.e. design, collection, processing, analysis, preservation, data sharing and reuse). https://www.lcrdm.nl/en/glossary

Data User / Generator:

Generates, accesses and/or analyses data to derive a conclusion within a data governance framework that benefits from the tools, resources, skills and workflows provided by data stewards. These processes may result in data ownership.

Research Software Engineer:

Research software engineer (RSE) - A growing number of people in academia combine expertise in programming with an intricate understanding of research. These RSEs may start off as researchers who spend time developing software to progress their research or they may come from a more conventional software-development background and are drawn to research by the challenge of using software to further research. https://www.lcrdm.nl/en/glossary

Data Infrastructure Manager:

A Data Research Infrastructure Support Professional is an ICT expert who manages and operates research infrastructures and the necessary services for the storage, preservation and processing of research data. Digital skills for FAIR and Open Science - https://op.europa.eu/en/publication-detail/-/publication/af7f7807-6ce1-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-1906942

Governance Capabilities

Data Capabilities

Research Software Engineering Capabilities

Data & Digital Research Infrastructure Mgt Capabilities

Data Analytics

A data lifecycle stage that involves the techniques that produce synthesized knowledge from organized data/information.

Data Methods

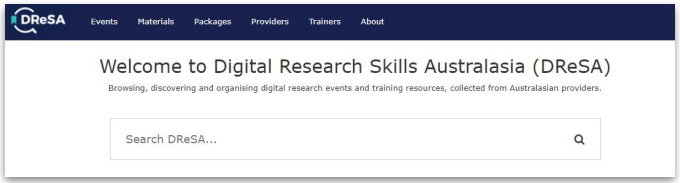
Superset capabilities

Data Analytics

Subset capabilities

- Handling Geospatial Data
- Statistical Inference
- Time-series Analysis
- Advanced Linear Modelling
- Uncertainty Propagation
- Spatial Analysis
- Scientific Workflow
- Nonlinear Modelling
- Qualitative Analysis
- Bayesian Techniques
- Machine Learning

See more



Governance Capabilities

Data Capabilities

Research Software Engineering Capabilities

Data & Digital Research Infrastructure Mgt Capabilities

Machine Learning

Ability for machines to learn from data without being explicitly programmed

Data Analytics

Superset capabilities

Machine Learning

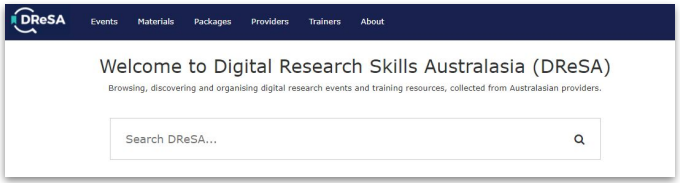
Subset capabilities

Supervised Learning

Unsupervised Learning

Reinforcement Learning

AI, Unstructured & Big Data

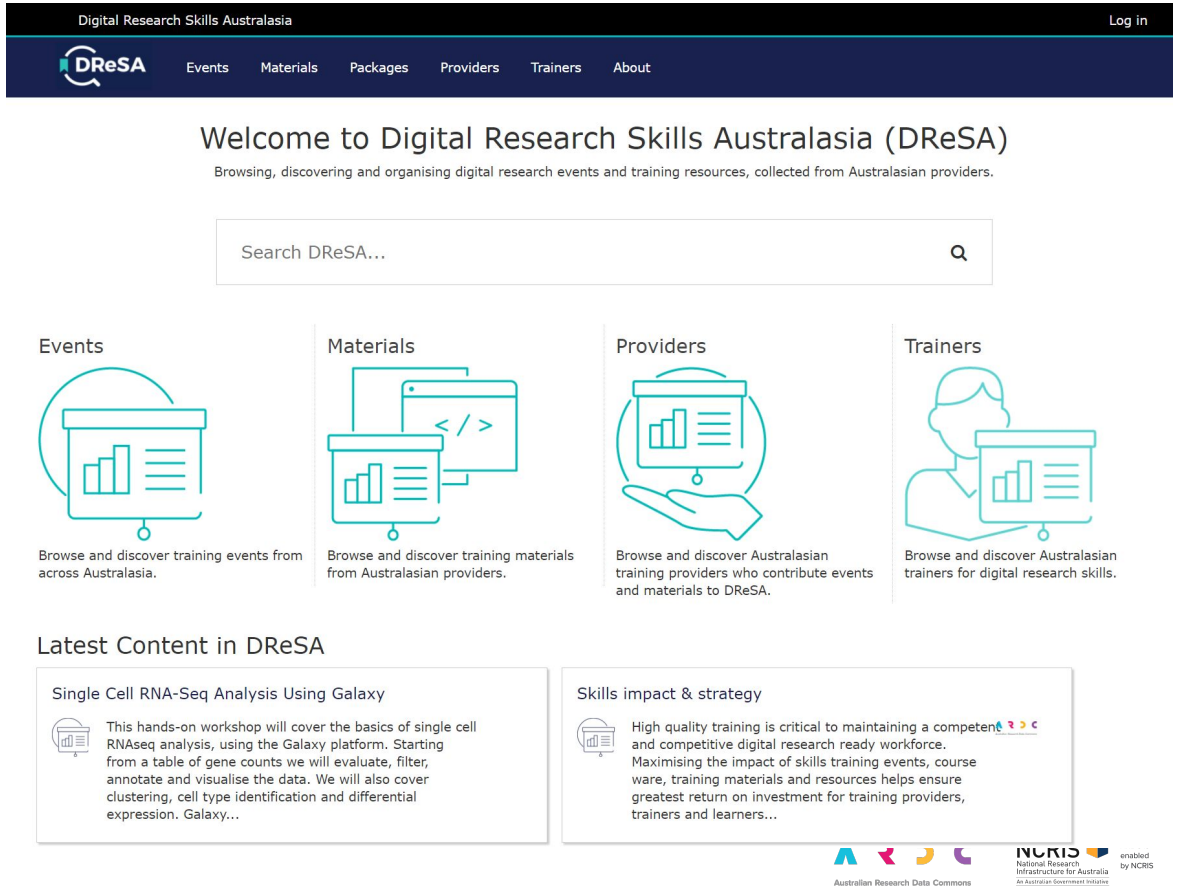


How to get them

DISCOVERING TRAINING THROUGH THE DIGITAL RESEARCH SKILLS AUSTRALASIA (DReSA) TRAINING PORTAL

What is DReSA?

Find digital research skills-focused educational events and resources with DReSA



DReSA makes it easier for learners, trainers, and training providers to find digital research skills-focused educational events and resources.

Who is DReSA for?

Individuals seeking training

- Find training and materials by topic, location, organiser, audience, etc.
- Subscribe to new items on current searches

Training providers/Trainers

- Promote and share training events and materials
- Contribute to a growing catalogue of materials
- Establish collaborative training partnerships

dresa.org.au

ARDC Australian Research Data Commons

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

ARDC is enabled by NCRIS

THE DReSA PARTNERS

Product owner group

Australian Research Data Commons

Australian BioCommons

INTERSECT RESEARCH FASTER

pawsey

Core partners

Australian Space Data Analysis Facility

CSIRO

DEAKIN UNIVERSITY

Federation UNIVERSITY • AUSTRALIA

MONASH University

NCI AUSTRALIA

QCIF

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

THE UNIVERSITY OF QUEENSLAND AUSTRALIA CREATE CHANGE

THE UNIVERS SYDN

Invaluable assistance from

AUSTRALIAN ACCESS FEDERATION

Australian Research Data Commons

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

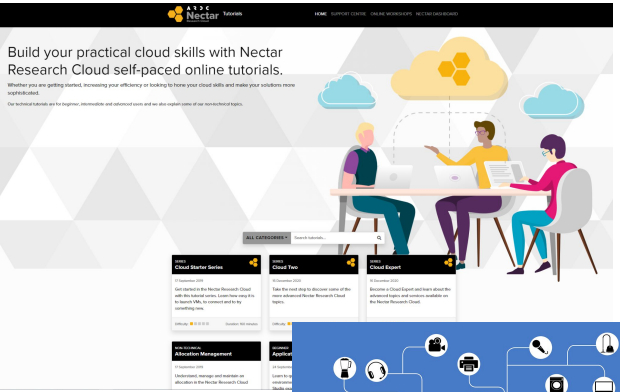
ARDC is enabled by NCRIS

Where you can go

WHAT YOU CAN ACHIEVE

SKILLS TRAINING

Where it can take you, the heights *or depths* you can reach



MACHINE LEARNING

IMAGE - GDJ - 2729781 - PIXABAY

IMAGE - JEFERRB - 782707 - PIXABAY

SMART URBAN IRRIGATION:

SAVING WATER & PROTECTING THE GREAT BARRIER REEF

IMAGE — MAREK OKON - TWWCQIMIUMG / UNSPLASH

Researchers from Central Queensland University worked alongside Cairns Regional Council and irrigation equipment manufacturer Rain Bird Australia to optimise urban irrigation using artificial intelligence, the ARDC Nectar Research Cloud and the Internet of Things, resulting in significant environmental benefits.

ARDC Australian Research Data Commons

NCRIS National Research Infrastructure for Australia — An Australian Government Initiative

ARDC is enabled by NCRIS

Australian Research Data Commons. (2021). Leading Australia to data-driven research impact. Zenodo. https://doi.org/10.5281/zenodo.5644236

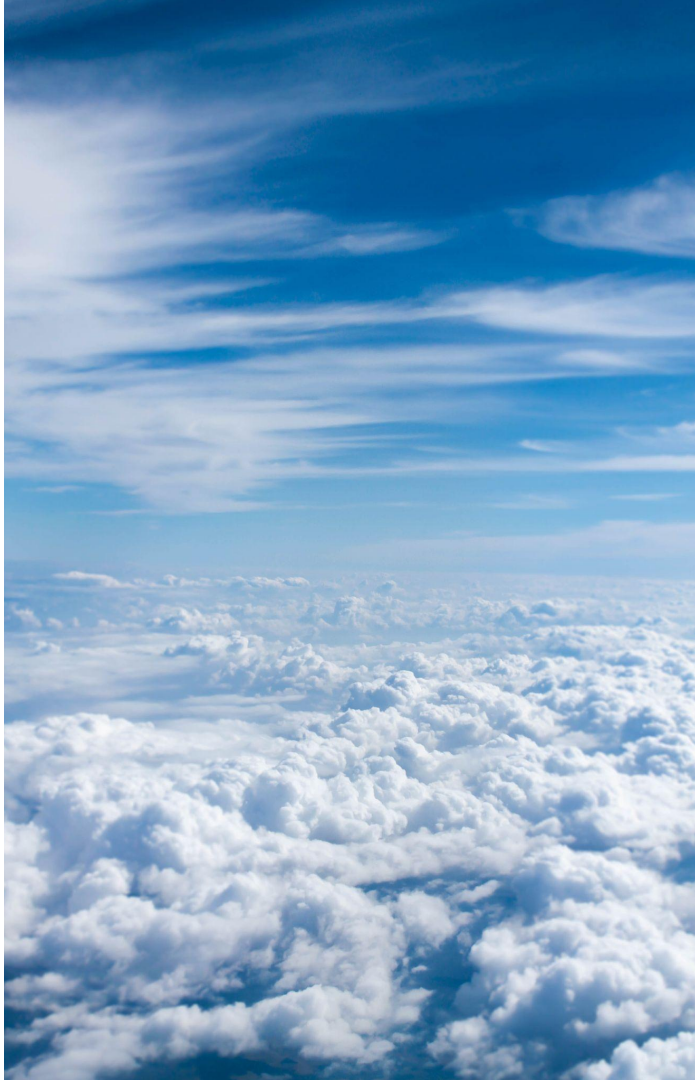


IMAGE - Taylor Van Riper on Unsplash

SKY'S THE LIMIT

Bringing it all together

- What are the skills needed to:
  - Solve research problems?
  - Create efficiencies in research practices?

Capability frameworks

- Where to find digital research skills training?

Training portals like DReSA

- What can be achieved through digital research skills training?

Subscribe to the **ARDC CONNECT** newsletter

THANK YOU

ardc.edu.au

contact@ardc.edu.au

+61 3 9902 0585

@ARDC_AU

Australian-Research-Data-Commons