

Day 1: Monday 26 October

Session 2 - 2.00 pm (AEDT)

ARDC Skills Landscape: Release, feedback and what's next

Hashtags

#ARDCSkills2020

Topic hashtags:

- #eResearchSkills
- #DataSkills
- #SoftwareSkills
- #Trainingresources

Group hashtag: #NeRDSkills





Australian Research Data Commons

Australian eResearch & Data Skills Landscape

An ARDC-National perspective

Kathryn Unsworth – Skills Consultant, ARDC



Muted during presentations



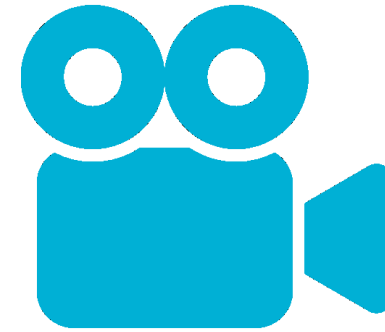
Unmute if you'd like to talk

For questions & comments



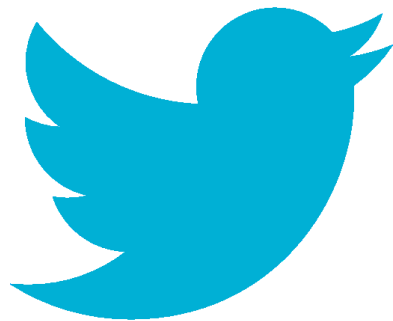
use the Chat box

Turn on your video

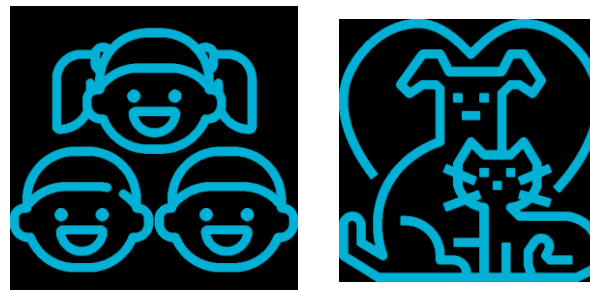


If you have the bandwidth

#ARDCskills2020



We welcome little people



& pets

Come prepared to contribute



and have fun!

We acknowledge and celebrate the First Australians on whose traditional lands we meet, and we pay our respect to the elders past and present. We extend that respect to Aboriginal and Torres Strait Islander peoples joining us today.



The Australian Research Data Commons purpose is to:

Provide Australian researchers with a competitive advantage through data



An Australian Research Data Commons

ARDC (the organisation) is working in concert with its partners to deliver the Australian Research Data Commons (the commons), which brings together people, data, **skills**, and resources to enable researchers to conduct world class data-intensive research.

The Commons is creating **transformational change** in the research data ecosystem, increasing the coherence and interoperability of existing investments and thereby increasing the effectiveness and efficiency of the system for researchers and producing more impactful outcomes for the nation.

ARDC Strategic Plan (2019-2023)



Transformational change...

*“And that is how change happens.
One gesture. **One person.** One
moment at a time.”*

— Libba Bray, [The Sweet Far Thing](#)



One researcher + Python = change

A trainee's story...

James Broadbent is an analytical chemist – part of a Molecular Analysis team at CSIRO.

His team specialises in proteome analysis, the large scale investigation of proteins.



Peptide specificity in the UniProt

Getting the information I need

Filter database search output file for peptides of interest

Peptide to search for: (K/R)**PEPTIDE**

Access API to get text output to search **Get organism name by regex**

```
>tr|Q4KY22|Q4KY22_PENCE Arginine kinase OS=Penaeus chinensis OX=139456 PE=2  
SV=1MADAAVIEKLEAGFKLEAATDCKSLKLYLTKAVFDQLKDKKTSLGATLLDVIQSGVENLD SGVGIYAPDAEAYTLFAPLFDPIIEDYHVG  
KQTDKHPNKDFGDVTSFVNVDPEGKYVISTRVRCGRSMEGYPFNPCLTEDQYKEMESKVSSTLSSLEGELKGTYYPLTGMGKEVQQKLIDDF  
LFKEGDRFLQAAANACRYWPSGRGIYHNKPEPTIDELRIISMQMGGDLGQVFRRLTSVAVNEIEKRIPFSHHDLRGLFTFCPTNLGTTVRASVHIK  
PKLAANRDKLEEVAGKYNLQVRGTRGEHTEAEGGIYDISNKRMRGLTEFQAVKEMQDGILELIKMEKEM
```

Create a table showing species versus peptide sequences

Output:

	DWKFLVWNEID	Tap Mix 1	Tap Mix 2	Tap Mix 3	Tap Mix 4	Tap Mix 5	Tap Mix 6	Tap Mix 7	Tap Mix 8
Penaeus chinensis	1	0	0	0	0	0	0	0	0
Organism 2	0	0	0	0	0	0	0	0	0
Organism 3	0	0	0	0	0	0	0	0	0
Organism 4	0	0	0	0	0	0	0	0	0
Organism 5	0	0	0	0	0	0	0	0	0
Organism 6	0	0	0	0	0	0	0	0	0
Organism 7	0	0	0	0	0	0	0	0	0
Organism 8	0	0	0	0	0	0	0	0	0
Organism 9	0	0	0	0	0	0	0	0	0
Organism 10	0	0	0	0	0	0	0	0	0
Organism 11	0	0	0	0	0	0	0	0	0

Digital Toolbox:

- Python
 - Jupyter Notebooks environment
 - Python libraries
 - For Requests
 - APIs that interact with UniProt
 - Code built to access large volumes of data via APIs (~500,000,000 sequences)
 - For data munging
 - For data manipulation
 - For data visualisation
 - Regex – cleaned the noise from the data served up by the UniProt API



Peptide specificity in the UniProt

James reflects:

“Now when I’m looking at any specific task the first thing that I do is ask myself: “Am I better off scripting this?”; my manager has also started to ask the same thing and we’re really changing the way our team works when it comes to data analysis and saving time!”



James →
Skills Landscape...

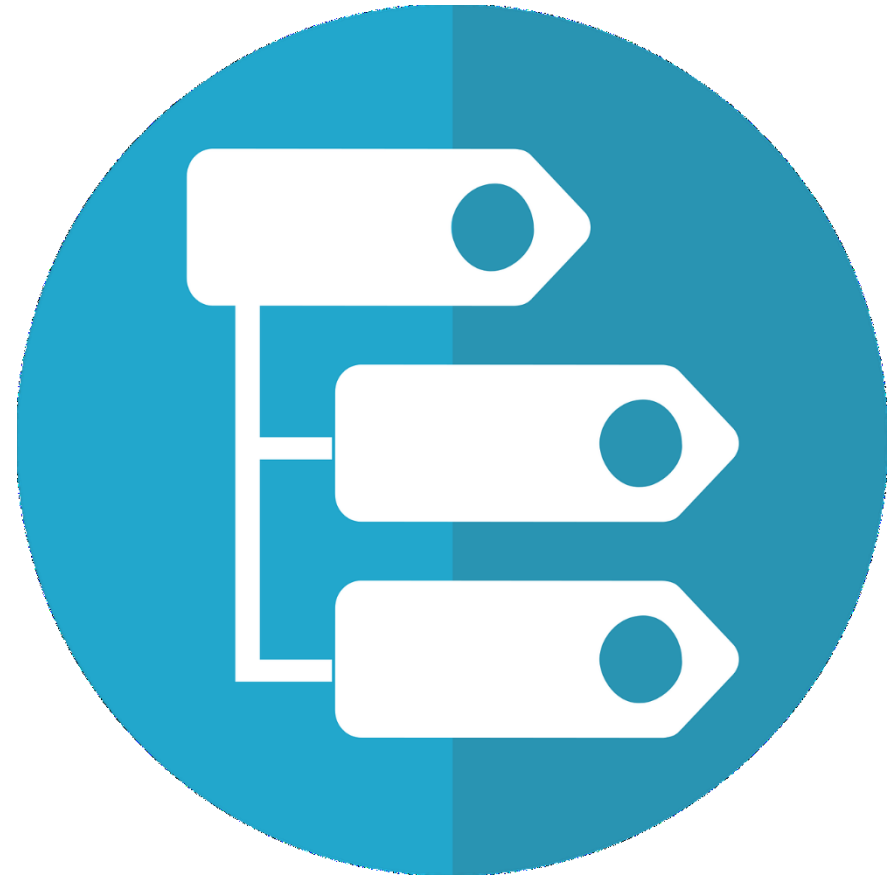
Why create a Skills
Landscape?



What is the eResearch & Data Skills Landscape?

The Landscape approach is one way of classifying research data skills – concepts, principles and relationships. Some might call it a...

Data Skills Taxonomy



What does the Skills Landscape help us do?

Helps us answer some of the following questions:

- What data skills are needed for data-intensive research?
- What skills and capability levels are needed for the various data-related roles?
- Who is developing and delivering these skills development programs?
 - Is there overlap and duplication?
 - Are there gaps?



Australia's eResearch & Data Skills Landscape

Why develop a Skills Landscape?

To identify:

- Data skills
- Skills ARDC focuses on
- Skills others in the sector focus on
- Overlaps and gaps

Points to note:

- Data Integrity happens because of good DG, FAIR & DM
- Data Sovereignty is captured under Gov policy & legislation
- Change management -> Cultural change – important, but a separate set of skills

Who is the Skills Landscape for?

- Skills training developers & trainers
- Data Stewards
- Researchers, but **not directly**

Other considerations:

- Skills Landscape takes a course/unit view
- Slides 2 to 6 focus on a generalised identification of skills, **not roles**
- Slide 8 describes four **key/generalised** roles
- Roles are not exclusive, i.e. an individual could undertake more than one data role at a time

Communities for skills and workforce development

Data Stewardship Skills

Data Governance

Policies & Standards

FAIR Data Principles

FAIR Outputs

Discovery & Reuse

FAIR Technical
Environment

Data Management

Working with Data

Preserving Data

Data Generation and Use Skills

Data Methods

Data Infrastructures



Question 1 for the audience:

What skills do you think ARDC should focus on to enable users to gain greatest benefit from the Commons?

Keep in mind:

- ARDC's remit for delivering the Commons is shared with our partners
- ARDC is not a training organisation (not funded for this purpose)
- ARDC's Skilled Workforce team is currently 3 FTE

Menti.com

Code: 77 22 28 1



Trainer / Educator

Train the Trainer

Materials /
Infrastructure

Data Stewardship Skills

Data Governance Skills

The process of creating and complying with data standards and policies which manage the availability, usability, integrity, use, and security of data. Effective data governance ensures that data is consistent and trustworthy, doesn't get misused, and generates value for the data owner(s).

Policies & Standards

Institutional Policies

Funders & Publishers Policies

Government Policies/Legislation

Intellectual Property

Research Integrity

Trust Certification

Trainer / Educator

Train the Trainer

Materials /
Infrastructure

Data Stewardship Skills

FAIR Data Principles Skills

Skills that are useful to create and use FAIR (Findable, Accessible, Interoperable, and Reuseable) data outputs and infrastructures that enhance the ability of machines and people to find, access, and use or reuse data.

These skills help facilitate trust through improved transparency and reproducibility.

FAIR Outputs

Metadata

Persistent Identifiers

Open Formats

Discovery & Reuse

Repositories & Discovery Portals

Standardised Communications Protocols

Provenance

Terms of Access

Licensing

FAIR Tech Environments

Semantic Resources

Question for the audience:

At your organisation –

Are you including FAIR Data Principles in your training?

Menti.com

Code: 77 22 28 1



Trainer / Educator

Train the Trainer

Materials /
Infrastructure

Data Stewardship Skills

Data Management Skills

The operational management and oversight of data assets to help provide users with high-quality data that is easily accessible in a manner consistent with the data governance framework, i.e. these are *tactical* skills.

Working with Data

Planning for Data
Management

Cleaning & Validating

Categorising

Structuring Data

Workflows

Handling Sensitive Data

Applying Local Protocols

Citing & Tracking

Moving Data

Accessing & Storing

Retention & Discovery
Infrastructures

Appraising, Selecting &
Disposing of Outputs

Preparing & Packaging
Outputs

Managing over Long-
term

Trainer / Educator

Train the Trainer

Materials /
Infrastructure

Data Generation & Use Skills

Skills which are useful for researchers and other data generators to ensure their data is, at the outset, structured and managed in such a way as to facilitate (re)use, high quality, and reflection of impact.

Data Methods

Data Analytics

Collection & Capture

Compilation, Derivation &
Aggregation

Simulation & Modelling

Reproducibility &
Replication

Data Visualisation &
Storytelling

Data Infrastructures

Data Repositories

Data Portals

Platforms/Facilities/
Resources

Access Management

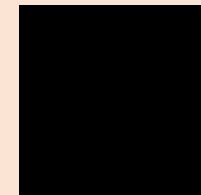
Citation & Impact Tracking

Data Stewardship Skills

Policies & Standards

Working with Data

FAIR Outputs



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

Data Owner:

Possesses and/or is responsible 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:

Manages *data as an asset* to help provide data users with high-quality data that is accessible in a manner consistent with the data governance framework, and to provide data generators with workflows, tools, skills, and resources for creating well-curated and appropriately accessible data sets.

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.

Who needs these skills?

Level	Competency	Description
1	Awareness	Background understanding - not a required skill though should be able to define and manage skill if required
2	Beginner (Foundational)	A novice understanding of the skill. You have exposure to the skill and understand basic concepts, but you lack experience.
3	Intermediate	You have experience with and can carry out the skill, but you don't understand advanced concepts. Ability to integrate capability into current work tasks.
4	Advanced	You have extensive and substantial training, practical experience and applied knowledge with the skill and understand advanced concepts.



Data Roles →	Researchers	Data Scientist	Research Software Engineer	Inst Senior Research Manager	Data Infrastructure Manager	Data Manager	Data Librarian	Data Archivist	Data Custodian
--------------	-------------	----------------	----------------------------	------------------------------	-----------------------------	--------------	----------------	----------------	----------------

Landscape Skills ↓

Data Stewardship									
Data Governance									
Policies & Standards	1	2	1	3	2	3	3	3	4
FAIR Principles									
FAIR Outputs	3	2	2	1	2	3	3	3	2
Discovery & Reuse	2	2	2	1	2	2	3	3	2
FAIR Tech Environments	1	2	4	1	2	2	3	3	1
Data Management									
Working with Data	3	3	3	1	2	4	2	2	4
Preserving Data	2	2	2	1	2	3	3	3	4
Data Generation & Use									
Data Methods	4	4	3	1	2	4	2	2	3
Data Infrastructures	3	3	3	1	2	3	2	2	2



Awareness

Beginner

Intermediate

Advanced

Data Science / Big Data Skills in SFIA

		Level of Responsibility						
		Low						High
		1	2	3	4	5	6	7
		Follow	Assist	Apply	Enable	Ensure, advise	Initiate, influence	Set strategy, inspire mobilise
Governance					Information governance			
					Strategic planning			
					Enterprise IT governance			
					Information systems coordination			
		Data management						
					Information assurance			
					Enterprise and business architecture			
					Relationship management			
					Business process improvement			
					Innovation			
Culture and capability					Organisational capability development			
					Organisation design and implementation			
			Methods and tools					
		Knowledge management						
			Measurement					
					Change implementation planning and management			
			Competency assessment					
		Learning delivery						
			Performance management					
			Systems development management					
Security and quality				Solution architecture				
		Requirements definition and management						
		Data modelling and design						
			Quality management					
			Conformance review					
			Quality assurance					
		Information security						
		Security administration						
			Business risk management					
			Continuity management					

Data Science / Big Data Skills in SFIA

		Level of Responsibility						
		Low						High
		1	2	3	4	5	6	7
		Follow	Assist	Apply	Enable	Ensure, advise	Initiate, influence	Set strategy, inspire mobilise
Governance					Information governance			
					Strategic planning			
					Enterprise IT governance			
					Information systems coordination			
		Data management						
					Information assurance			
					Enterprise and business architecture			
					Relationship management			
					Business process improvement			
					Innovation			
Culture and capability					Organisational capability development			
					Organisation design and implementation			
			Methods and tools					
		Knowledge management						
			Measurement					
					Change implementation planning and management			
			Competency assessment					
		Learning delivery						
			Performance management					
			Systems development management					
Data lifecycle management				Solution architecture				
		Requirements definition and management						
		Data modelling and design						
			Database design					
			Business analysis					
			Systems design					
		Programming/software development						
			Business process testing					
		Testing						
			Analytics					
Security and quality			Data visualisation					
		Information content authoring						
		Information content publishing						
		Database administration						
			Availability management					
		Storage management						
			Capacity management					
			Quality management					
			Conformance review					
			Quality assurance					
	Information security							
	Security administration							
		Business risk management						
		Continuity management						

This view illustrates the SFIA skills which are most relevant to the field of big data, data analytics and data science. You can find additional and complementary skill definitions in the SFIA reference manual or on the web site www.sfia-online.org



ARDC Competency Profile - Data Roles		
Title of Data Role		
Data Repository Manager/Data Curator/Data Archivist/Librarian - Data Services		
Purpose of Data Role		
Develop and provide scalable and sustainable research data management services, infrastructure, procedures, advice and training that support faculty, researchers and students in the discovery, management, use and preservation of data. Also, to enable active lifecycle curation and re-use of FAIR digital assets (including objects, datasets and artefacts) generated during research activities, projects and contracts.		
Accountability Areas		
Key Responsibilities	Tasks	ARDC Australian Skills Landscape - Skills - Learning paths
Policy, outreach and advocacy	<ul style="list-style-type: none"> Generally raise research data awareness and advocate for the benefits of research data management Design and deliver training and advice to target communities on: <ul style="list-style-type: none"> Research data services Good data management practices Open access and data sharing Relevant research data policies, including Funders' policies and requirements and data publication requirements of publishers/journals Build community around research data management 	DG - Institutional Policies DG - Funders' & Publishers' Policies DG - Research Integrity
Data Management	<ul style="list-style-type: none"> Assisting researchers with data management plans and DMP tools Data cleaning and validation (verification) Data conversion Best practice for data structures, types, formats and file naming conventions Understand research practices and workflows (workflow controls) 	DM - Planning for Data Management DM - Cleaning & Validating DM - Structuring Data FAIR - Open Formats DM - Workflows

	<ul style="list-style-type: none"> Disciplinary norms and standards for data management Version control 	
Data discovery - description and documentation	<ul style="list-style-type: none"> Metadata standards and schemas Domain ontologies, vocabularies, etc. Identifiers, data linking and data integration techniques 	FAIR - Metadata FAIR - Semantic Resources FAIR - Persistent Identifiers FAIR - Provenance
Data deposit/publishing	<ul style="list-style-type: none"> Data repository and storage platforms Prepare data for deposit Appraisal, selection, and ingest into repository system Assigning identifiers Data citation Data licensing and intellectual property Data security 	FAIR - Repository and Discovery Portals DG - Trust Certification DM - Appraising, Selecting & Disposing of Outputs FAIR - Persistent Identifiers DM - Citing & Tracking FAIR - Licensing DG - Intellectual Property FAIR - Terms of Access
Archiving and preservation	<ul style="list-style-type: none"> Refresh digital media and migrating data Secure storage and access Link validation and other checks Processes for recording 3rd party data access requests Data retention, embargo and disposal processes 	DM - Moving Data DM - Accessing & Storing FAIR - Standardised Communication Protocols DM - Retention & Discovery Infrastructures DM - Appraising, Selecting & Disposing of Outputs DM - Managing over Long-term

References:

- Tammaro, A., Matusiak, K. K., Sposito, F., & Casarosa, V. (2019). Data Curator's Roles and Responsibilities: An International Perspective, *Libri*, 69(2), 89-104. doi: <https://doi.org/10.1515/libri-2018-0090>
- Schmidt, B. & Shearer, K. (2016). Joint Task Force on Librarians' Competencies in Support of eResearch and Scholarly Communication. Librarians' Competencies Profile for Research Data Management. https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf

Learning Path

Persistent Identifiers (Draft)

Persistent unique identifiers provide a means of long-lasting identification of digital objects that are global, standardized, and widely used in the digital environment and can provide information on the object, regardless of where the object is located. Persistent unique identifiers include DOIs, ARKs, Handles, and ORCIDs. Assigning persistent unique identifiers to data helps to provide a method to locate data in the vast amounts of research data generated on a daily basis. DataCite is one initiative that provides an opportunity for research organizations to assign DOIs to their datasets. Assigning DOIs allows for a particular dataset to be persistently identified so that it can always be located and cited appropriately. While information about a digital object may change over time, including where to find it, its DOI name will never change so it can always be found. Librarians can help researchers assign persistent unique identifiers to data by providing insight on the process of assigning and applying the identifiers. <https://nmlm.gov/data/thesaurus/persistent-unique-identifier>

Time to complete

TBD

Date created

5th October 2020

Curated by

Australian Research Data Commons (ARDC)

Description

Persistent identifiers (PIDs) – for people (researchers), places (their organizations) and things (their research outputs and other contributions) – are foundational elements in the overall research information infrastructure. They enable these entities to be uniquely identified, to create reliable links between them.

Learning Path - Persistent Identifiers

Haak, L., & Brown, J. (2019). Persistent identifiers: the building blocks of the research infrastructure. *Insights*, 32(1), 9. DOI: <http://doi.org/10.1629/insig.457>

Do you want to know more about Persistent Identifiers (PIDs) and the role they play in building open infrastructures?

This learning path covers the essential principles of Persistent Identifiers (PIDs) and the requirements needed to start issuing and delivering PIDs for research outputs, and contributors, facilities, organisations, etc.

Level (competency)

This learning path is advanced - depending on knowledge requirements for your role, the learning path can be undertaken in its entirety or to the level of competence required.

What you will learn

A "Persistent Identifier" is a unique and permanent digital reference to an object. Persistent Identifiers are useful for a wide range of research outputs and contributors. The various types of Persistent Identifiers are used to identify and manage research outputs and Persistent Identifiers are used by researchers, librarians and research managers for the management of Persistent Identifiers systems and practices for implementing Persistent Identifier systems.

Who should take this learning path

This learning path is primarily for anyone associated with the creation and management of data, those in Data Stewardship roles (Data Managers, Data Archivists, Data Librarians), Data Service Researchers, Research Software Engineers and Research Managers.

Competency Level ↓	Data Stewards	Data Services Developers	Researchers	RSEs	Research Managers
Beginner	✓	✓	✓	✓	✓
Intermediate	✓	✓	✓	✓	✓
Advanced		✓			

Learning Path - Persistent Identifiers

1

Learning Path - Persistent Identifiers

2

Learning requirements and/or prerequisites

Start your learning journey

Gain a basic understanding. Big-picture view.

OpenAIRE - what-is-a-persistent-identifier

ID Forum - Knowledge Hub

Getting started with PIDs

- The Power of PIDs - Video
- Why Use Persistent Identifiers?
- New Types of Persistent Identifiers

OpenAIRE - Preservation Coalition (DPC) - Persistent identifiers - Digital Preservation Handbook

OpenAIRE - Persistent Identifiers (PID) - The Whys and the Hows

OpenAIRE - Data citation | RDNL - Essentials 4 Data Support

What to learn more?

Gain a basic understanding. Move on to the how and why.

ID Forum - Knowledge Hub

PIDS for Librarians and Repository Managers

- Working with some PID Providers
- Why are PIDs important for Librarians and Repository Managers
- Case Study: DOIs and Historic Literature

PIDS for Funders and Policy Makers

- PIDs for Funders and Policy Makers

PIDS for Publishers

- Case study: Adopting PIDs for publications at National and University Library in Zagreb
- PIDS for Publishers

- PIDS for Researchers
 - Background reading materials on PIDs - collated by a student in digital humanities
 - Persistent identifiers - what's in it for researchers?

Also for Researchers

- How persistent identifiers can save scientists time - NCBI - Meadows, A., & Haak, L. (2018). How persistent identifiers can save scientists time. *FEMS microbiology letters*, 365(15), fny143. <https://doi.org/10.1093/femsle/fny143>
- OpenAIRE - Guides for Researchers - How can identifiers improve the dissemination of your research outputs? (Focus is on ORCIDs) <https://www.openaire.eu/how-can-identifiers-improve-the-dissemination-of-your-research-outputs>

CrossRef Curriculum - Persistent identifiers

Building expertise

Applying new knowledge into practical skills. Making your learning work.

The PID Forum - Knowledge Hub

- PIDS for Developers
 - APIs and Documentation

DataCite - Lessons Learned on Persistent Identifiers for Research Data

McMurry JA, Juty N, Blomberg N, Burdett T, Conlin T, Conte N, et al. (2017) Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biol* 15(6): e2001414. <https://doi.org/10.1371/journal.pbio.2001414>

Clump, J., & Huber, R. (2017). 20 Years of Persistent Identifiers – Which Systems are Here to Stay?. *Data Science Journal*, 16, 9. DOI: <http://doi.org/10.5334/dsj-2017-009>

OpenAIRE - Project outputs

Learning Path - Persistent Identifiers

3

Who is providing these skills?

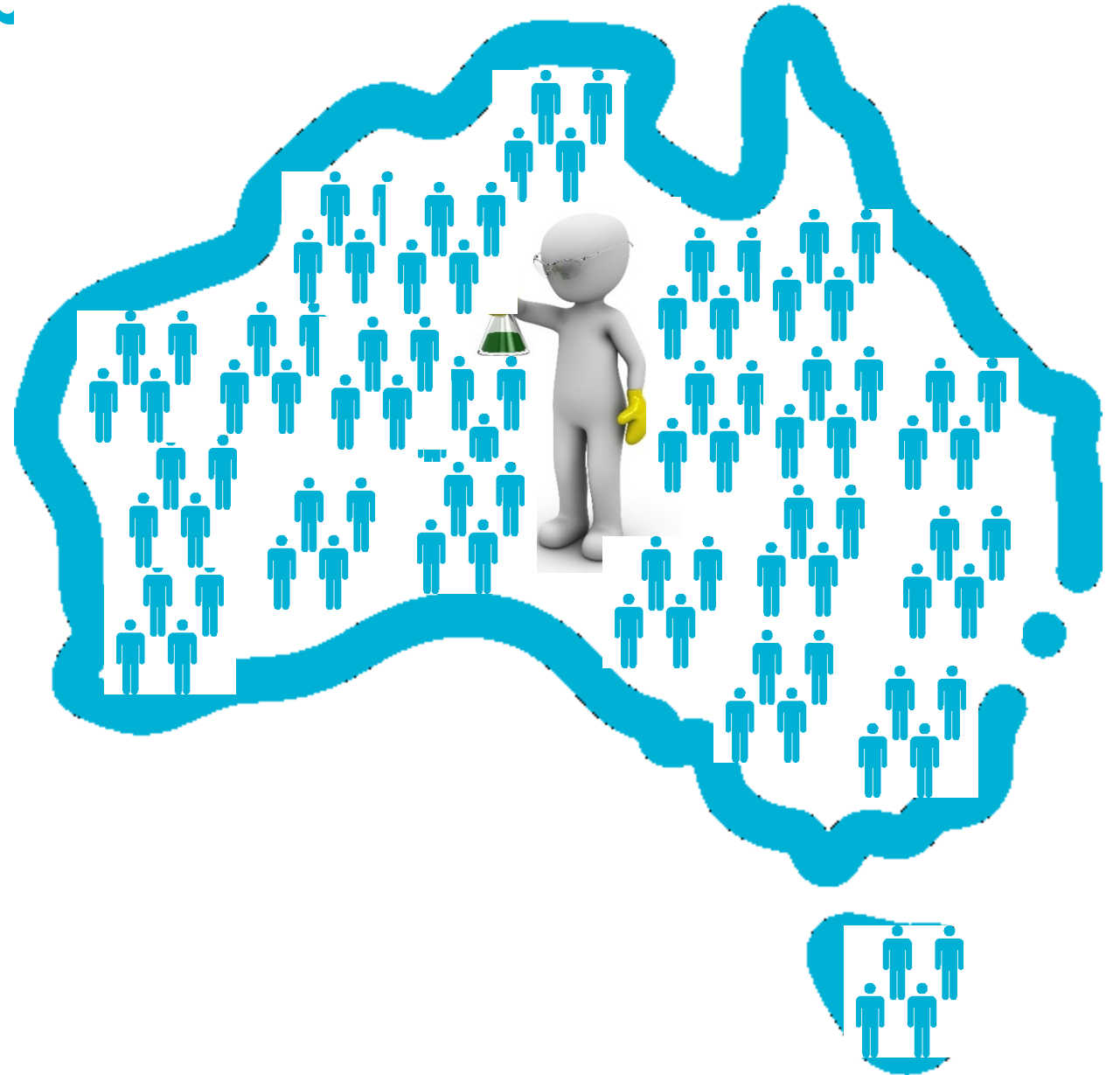
Activity - Skills Providers mapping

<https://tinyurl.com/y4hgt8l9>



Circling back to the beginning

Provide Australian researchers with a competitive advantage through data



Day 1: Monday 26 October - Next Session

Session 3.30pm (AEDT)

Community Networking – via
Runtheworld

ARDC's Annual Satisfaction Survey –
launched today! ardc.edu.au/survey

Hashtags

#ARDCSkills2020

Topic hashtags:

- #eResearchSkills
- #DataSkills
- #SoftwareSkills
- #Trainingresources

Group hashtag: #NeRDSkills