

Digital skills for data-intensive science



Australian Research Data Commons

NCRIS

National Research
Infrastructure for Australia

An Australian Government Initiative

Dr Michelle Barker

michelle.barker@ardc.edu.au

Deputy Executive Director
Australian Research Data Commons

Chair of OECD Global Science Forum Expert
Group on Digital Skills for Science

What is the OECD Global Science Forum?

The overall objective of the Global Science Forum (GSF) is **to support countries to improve their science policies and share in the benefits of international collaboration.**

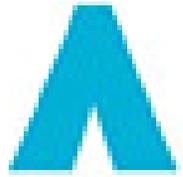
GSF provides a venue for consultations and mutual learning among senior science policy officials of OECD member countries. It carries out analytical work on high-priority science policy issues.

Specifically, the GSF serves its members in the formulation and implementation of their science policies by:

- exploring opportunities and mechanisms for new or enhanced international co-operation in selected priority areas;
- defining international frameworks for national or regional science policy developments; and
- addressing the science policy dimensions of issues of global concern.



1. What are the digital skills needs for science?
2. Are these skills distinct from other areas of the economy?
3. Do these needs vary across domains?
4. Are the needs being met? If not, where are the gaps?
5. What are countries, science agencies, institutions, communities doing?
6. Is this an area where trans-national cooperation, mutual learning at the policy level would be beneficial?



Australian Research Data Commons

NCRIS

National Research
Infrastructure for Australia

An Australian Government Initiative

The ARDC is a transformational sector-wide initiative, working with partners to build a coherent national and collaborative research data commons.

How can the demand for data scientist be quantified at research discipline, institution and national levels?

Estimated # Australian biology researchers: 30,000 (In 5 years → 31,500)		(and perhaps ~ 1 million worldwide)	
Estimated #: 1,000¹ (In 5 years → 1,500)	2,000² (→ 3,000)	7,000³ (→ 12,000)	20,000³ (→ 15,000)
<i>bioinformaticians</i>	<i>bioinformatics-intensive bioscience researchers</i>	<i>data-intensive bioscience researchers</i>	<i>biology-focussed bioscience researchers</i>
research into and application of techniques, tool development	where research is fully dependent on advanced or novel use of bioinformatics	where 'omics data analysis is a critical contributor to, but not definer of, the research outcomes	occasional users of bioinformatics web services where bioinformatics adds value to research outcomes determined by other means
<i>Eg. research generating new tool or statistical method; bioinformatics core facilities applying complex analyses</i>	<i>Eg. Genomic cancer research, population genomics/agricultural genomics programs</i>	<i>Eg. RNAseq analysis to identify upregulated genes in broader research program</i>	<i>Eg BLAST, Ensembl</i>

What is a data scientist?

- **Data engineer:** Operating at a low level close to the data, they are people who write the code that handles data and moves it around. They may have some machine learning background.
- **Data analyst:** This is someone who knows statistics. They may know programming or they may be an Excel wizard. Either way, they can build models based on low-level data. They eat and drink numbers that are related to their work but they're not interested in data as an abstract concept. Most importantly, they know which questions to ask of the data.
- **Data stewards:** These people think about managing and preserving data. They are information specialists, archivists, librarians and compliance officers. This is an important role. If data has value, you want someone to manage it, make it discoverable, look after it and make sure it remains usable

Kenji Takeda, [What is a Data Scientist](#), 31 January, 2013

Now research software engineers are routinely included

What sort of skills are we talking about?

- Data management in the form of curation skills
- Data integration including understanding of formalized semantics and ontologies
- Software skills and programming best practices
- Data visualization and mapping tools
- Collaboration and communication skills

Is developing skills enough?

What are the larger factors that enable success of existing skills initiatives?

- ✓ Institutional support
- ✓ Funding
- ✓ Government policy or roadmaps
- ✓ Proximity to technical infrastructure
- ✓ Being part of a community network, with resources and support

What is the role of research infrastructures in developing this workforce of the future?

- ? Institutional exemplars
- ? Curriculum development
- ? Best practice in recognition and esteem
- ? Influencing government
- ? Providing community networks
- ? Pioneering different physical spaces



Thanks!

michelle.barker@ardc.edu.au

