RI related Challenges in Data Science

Large scale infrastructure especially GPUs for running deep learning experiments.

It's a challenge to capture and preserve the meaning of data so that it is usable, useful and used... is there any relevant infrastructure to support that?

More GPU resources for running Deep Learning experiments

Simple onboarding to understand a) when to use cloud infrastructure like Nectar and b) how to get started quickly and easily to test out and see if it's the right option for a project

a coordinating/supporting cross-discipline team to help researchers engage with relevant stakeholders in the industry, so their research has actual impacts for specific sectors

Fast moving area of data science, how to keep on top of thisHow to know about and access and learn new methods

lack of available and accessible datasets in health, limits in large computational cloud space, limited in collaboration platforms

Varying description of variables in datasets: even though it is the same area, that makes it very difficult to integrate datasets or compare them. Also, differences in how data is collected, especially in health care.

We currently operate a range of server-based pipelines that use machine learning approaches to process large numbers of images within very small intervals of time. This infrastructure can sometimes be highly intensive and volatile.

RI related Challenges in Data Science

Need for short accessible guides to demystify and de-jargonize topics relating to metadata, data dictionaries, voacbularies, research data infrastructure

real-time social media from a wider range of platforms beyond twitter

High power compute - nectar doesn't seem to cope at scale

How help with practices

Finding out who's doing what in Australia in different areas of data science

10

National Coordination to address challenges

National coordination about who is who where doing what in data science across Australia

National coordination about impact of data science

Help training hdr

Data science problem challenges specific to Australia

Coordination could help in reducing duplication of effort...e.g., by helping researchers contribute to existing data holdings, rather than starting from scratch.

Connecting industry with relevant academic/research expertise across the country

Ensuring there is an appropriate balance of investment in gathering new data, and making use of existing data holdings.

4