A FAIR Data Sharing Framework for Large-Scale Human Cancer Proteogenomics

Islam M[1,2], Christiansen J[3], Mahboob S[4], Valova V[4], Baker M[4], Capes-Davis D[4], Hains P[4], Balleine R[1,4], Zhong Q[1,4], Reddel R[1,4], Robinson P[1,4], *Tully B[4]*

[1] *The University of Sydney, Camperdown, Sydney, NSW, 2050, Australia*
[2] *Intersect, Level 13/50 Carrington St, Sydney, NSW, 2000, Australia*
[3] *Queensland Cyber Infrastructure Foundation Ltd, Axon Building 47, University of Queensland, St Lucia, Brisbane, QLD, 4072, Australia*
[4] *Children's Medical Research Institute, Westmead, NSW, 2145, Australia*

Background
The ACRF International Centre for the Proteome of Cancer (ProCan) at Children's Medical Research Institute (CMRI) is an "industrial scale" program specialising in small-sample proteomics analysis from human cancer tissue.

ProCan seeks to generate both a wide and deep analytics pipeline and requires an enabling data framework. The framework must accommodate initial analysis and proteomic profiling of a large number of tumor samples, along with the clinical and demographic information, subsequent multi-omics studies, and any previously recorded responses to treatment. The curated datasets will provide a valuable resource beyond their primary use and ProCan is committed to making its data accessible to collaborators and the wider scientific community.

Objectives
The objective of the project is to an establish efficient, reliable, secure and ethical data sharing and publication framework based on the best practice data sharing principles, such as the FAIR principle. The framework must address various challenges that stem from the scale and complexity of the program, and ProCan's focus on human-derived data and associated challenges presented in sharing these data while maintaining the privacy of any research participants.

Method
The project adopted a requirements-driven methodology and engaged with a wide range of ProCan stakeholders nationally and internationally. Together, various industrial-scale proteomics data management and sharing scenarios were explored such that robust and ethical sharing of the data would be achieved.

Results
The project developed a data sharing framework based on the FAIR principle that currently forms the basis of ongoing implementation work within the ProCan program.