Enhancing Australia's capability for secure and responsible sharing of human genomics research data

Marion Shadbolt[1]✉, Tiffany Boughtwood[3], Jeff Christiansen[1], Joe Copty[7], Mark Cowley[6], Kylie Davies[2], Matthew Downton[9], Kelsey Druken[9], Ben Evans[9], Clara Gaff[8], Andrew Gilbert[5], Christina Hall[1], Matthew Hobbs[7], Oliver Hofmann[11], Jessica Holliday[1], Warren Kaplan[9], Ross Koufariotis[10], Sarah Kummerfeld[7], Conrad Leonard[10], Angela Lin[6], Andrew Lonie[1], Heath Marks[2], Siobhann McCafferty[4], David Monro[9], Andrew Patterson[11], John Pearson[10], Bernard Pope[1], Shyamsundar Ravishankar[7], Florian Reisinger[11], Andrew Robinson[9], Victor San Kho Lin[11], John Scullen[2], Mustafa Syed[6], Kamile Taouk[6], Andrew Treloar[4], Jingbo Wang[9], Marie Wong-Erasmus[6], Scott Wood[10]

✉ marion@biocommons.org.au

1. Australian BioCommons 2. Australian Access Federation (AAF) 3. Australian Genomics 4. Australian Research Data Commons (ARDC) 5. Bioplatforms Australia 6. Children's Cancer Institute (CCI) and the Zero Childhood Cancer Program (ZERO) 7. Garvan Institute of Medical Research 8. Melbourne Genomics Health Alliance 9. National Computational Infrastructure (NCI) 10. QIMR Berghofer Medical Research Institute (QIMRB) 11. The University of Melbourne Centre for Cancer Research (UMCCR)

What is the Human Genomes Platforms Project?

The Human Genomes Platform Project (HGPP) is a nationally funded initiative that aims to leverage best practice technologies and global standards to accelerate FAIR human genomics data sharing in Australia. Involving Australia's largest human genomic sequence generators, along with national computing infrastructure partners, the HGPP will break down silos and facilitate the deployment of a services toolbox across Australian genomic research institutes.
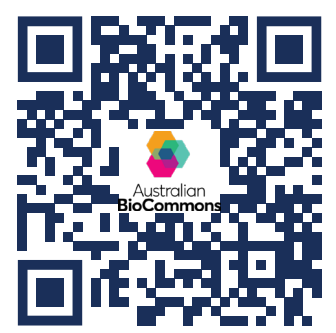
Virtual Cohorts

Currently, searching across national data repositories is challenging due to limited data sharing infrastructure support. The virtual cohorts sub-project is enabling cross-repository searching by facilitating implementation of data commons and lighting a series of Beacon v.2 instances, to allow cross-institution querying via a Beacon network. We are also working to create a web-based user interface for user-friendly searching across the beacon network.

DAC Automation

Currently, the Data Access Committee (DAC) request and approval process is manual across many Australian research institutes. As the quantity of human genomics datasets increases; tracking, auditing and managing this process is becoming increasingly challenging.

Based on partner requirements, this sub-project has established a pilot instance of the Resource Entitlement Management System (**REMS**). Each partner has conducted pilot testing and we are working to ensure required features can be implemented.
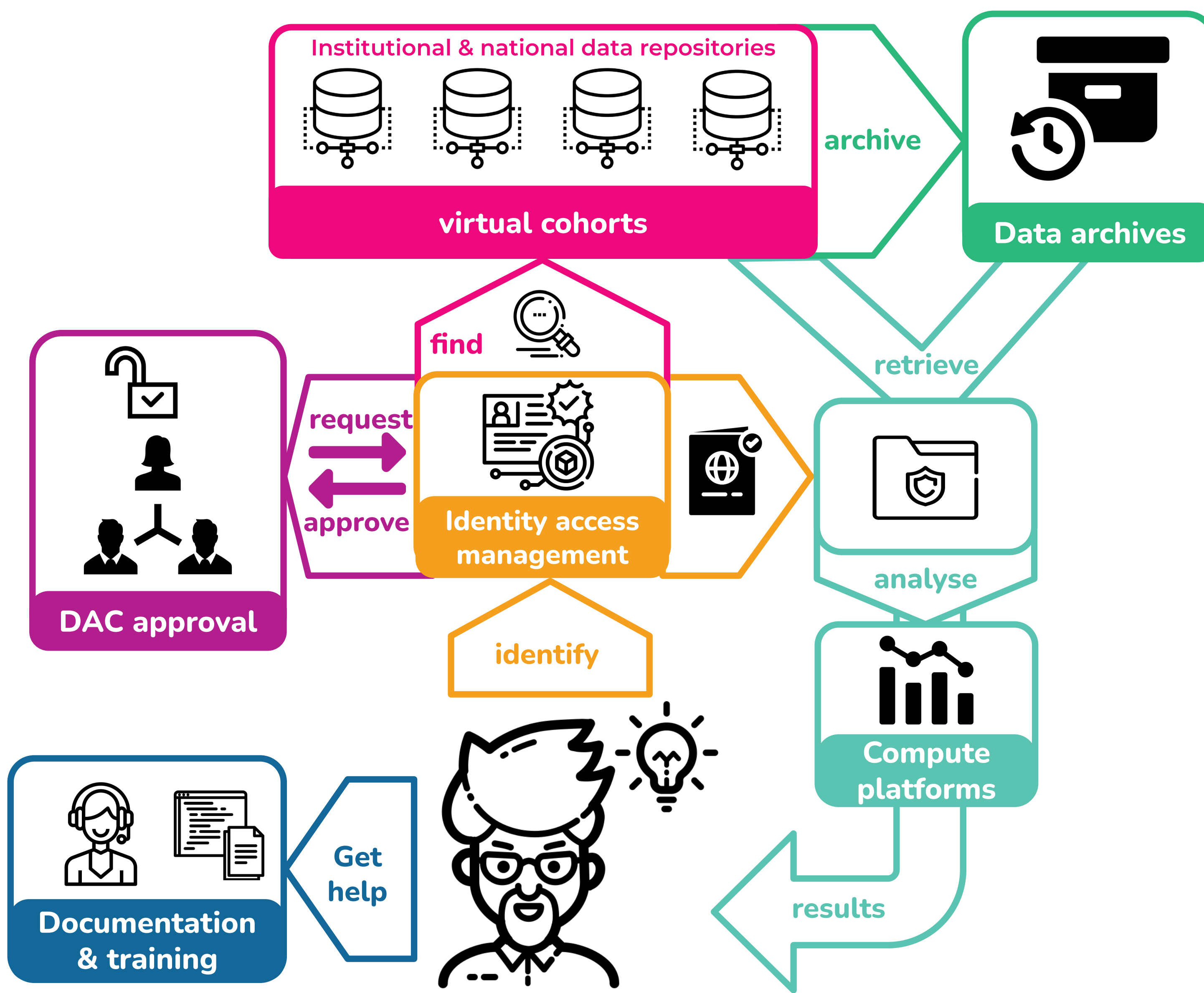
[Meta]data Archiving

Researchers need to archive human genomics data but submission to international controlled access repositories is challenging. The Data and Metadata Archiving sub-project is investigating the feasability of establishing a Federated EGA node in Australia.

To date, we have deployed and are testing an instance of the LocalEGA software at NCI.

Our main output will be a detailed feasibility report that explores use cases and makes implementation recommendations for this important national service.

Diagram labels

Institutional & national data repositories — virtual cohorts — archive — Data archives — retrieve — find — request — approve — DAC approval — Identity access management — identify — Get help — Documentation & training — analyse — Compute platforms — results

Documentation, Communications & Training

The documentation, communications and training sub-project is ensuring all newly-implemented technologies are adequately documented and disseminated to the broader community. The publication of news articles, reporting outputs to Zenodo and code sharing on GitHub, enables access to key documentation. Upcoming webinars and training events will facilitate uptake and adoption.

Federated Identity & Access Management

The Federated IAM sub-project is exploring systems that can be used across Australian institutes to confidently ascertain a user's identity and manage their access rights. We are assessing the suitability of CILogon an appropriate way to establish a National Researcher identity for life sciences researchers. This will enable researchers to use their existing, institutionally-managed identity to gain access to a range of national life science services and resources.
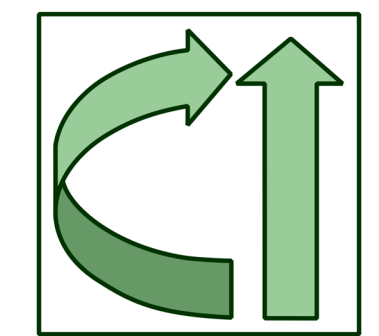
Icons from the Noun Project: Computing, search, cloud by Flatart, database by Start Up Graphic Design, Analyse by Taylan Sentürk, computing by Phonlaphat Thongsriphong, identified by Tippawan Sookruay, group by Gregor Cresnar, group by Oksana Latysheva, Data File by Blangcon, Unlock by Arthur Shlain, archive by Adrien Coquet, support by Komkrit Noenpoempisut, documentation by lastspark, Scientist by Maxim Kulikov, Immigration Approval by Ary Prasetyo