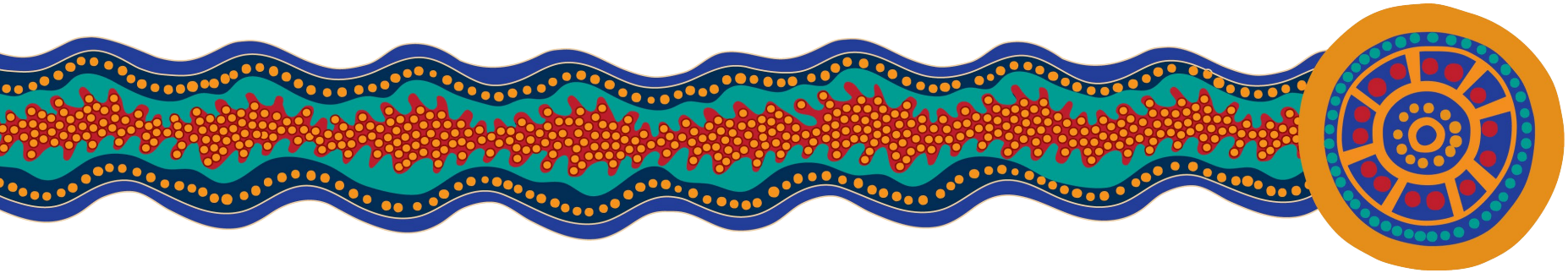
QLD
National Primary and Acute Care
Data Linkage Project

Privacy-Preserving Record Linkage and the Separation Principle

Sankha Bandara and Raymond Daniel

21 November 2024

National Primary & Acute Care
Data Linkage Project



The Queensland Government respectfully acknowledges Aboriginal and Torres Strait Islander peoples as the Traditional and Cultural Custodians of the lands on which we live and work to deliver healthcare to all Queenslanders and recognises the continuation of First Nations peoples' cultures and connection to the lands, waters and communities across Queensland.

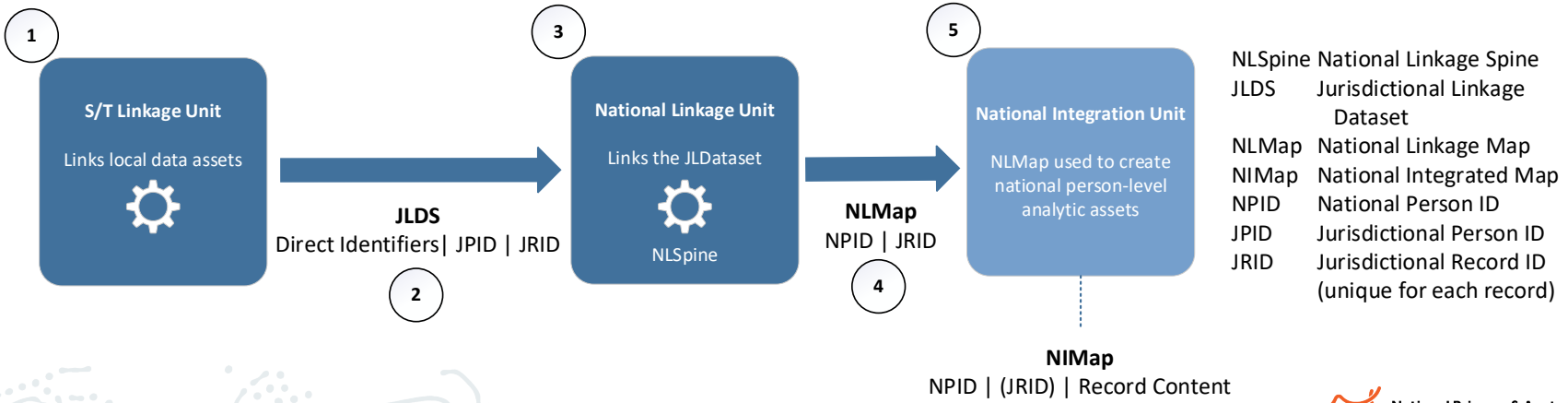
About this presentation

- Concept of distributed record linkage
- Privacy Preserving Record Linkage
- Separation Principle
- Addressing security concerns

Centralised linkage system

Currently, a centralised national linkage system is used to deliver cross-jurisdictional linkage projects and enable national analytic assets.

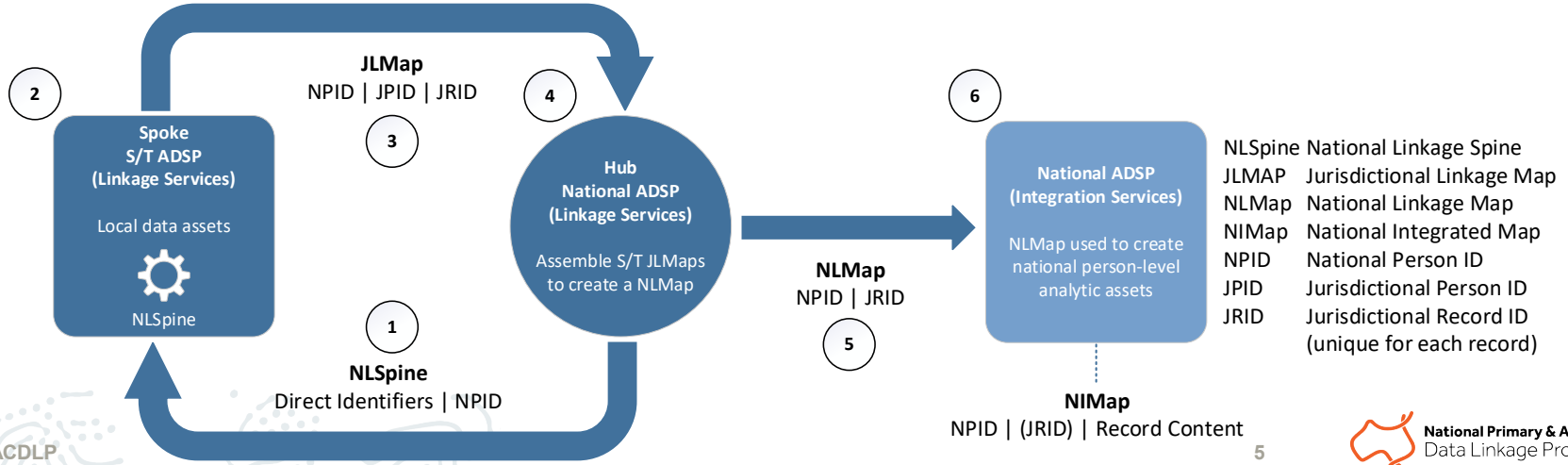
- Deliverable: create a National Linkage Map to enable person-level integration of content data.
- One-way data flow – jurisdictions' direct identifiers sent to the Australian Institute of Health and Welfare.
- Identifiers used to enable AIHW to link to their National Linkage Spine and create a National Linkage Map.



Distributed linkage system (Hub-and-Spoke)

Population Health Research Network National Master Linkage Key project is (re)distributing the linkage effort – State/Territory jurisdictions adopting a larger role.

- Data cycle – Accredited Data Service Providers receive the National Linkage Spine, link local data to the spine, and provide maps not containing direct identifiers back to the AIHW.
- The AIHW uses deidentified Jurisdictional Linkage Maps to assemble the National Linkage Map.



Distributed linkage system benefits and challenges

Benefits
Linkage system efficiency • Reduced duplication of effort • Better resource allocation • More even distribution of effort • Clear linkage requirements
Linkage quality • Greater consistency across jurisdictions • Enables group changes across the full linked period (i.e. historical records)
More timely and frequent linkage
Technical capability for faster and more efficient cross-jurisdictional linkage projects

Challenges
Having more sensitive richer data, there's a higher risk to privacy (e.g., re-identification of individuals) and data security
Technical complexity and dependencies
Initial costs and infrastructure upgrade requirements
Ongoing funding and maintenance
Public and stakeholder readiness

Privacy Preserving Record Linkage

Privacy Preserving Record Linkage (PPRL): A set of techniques that allows for the linking of records across different datasets without exposing sensitive personal information. Prevents persons involved in linking records to see personal identifying information.

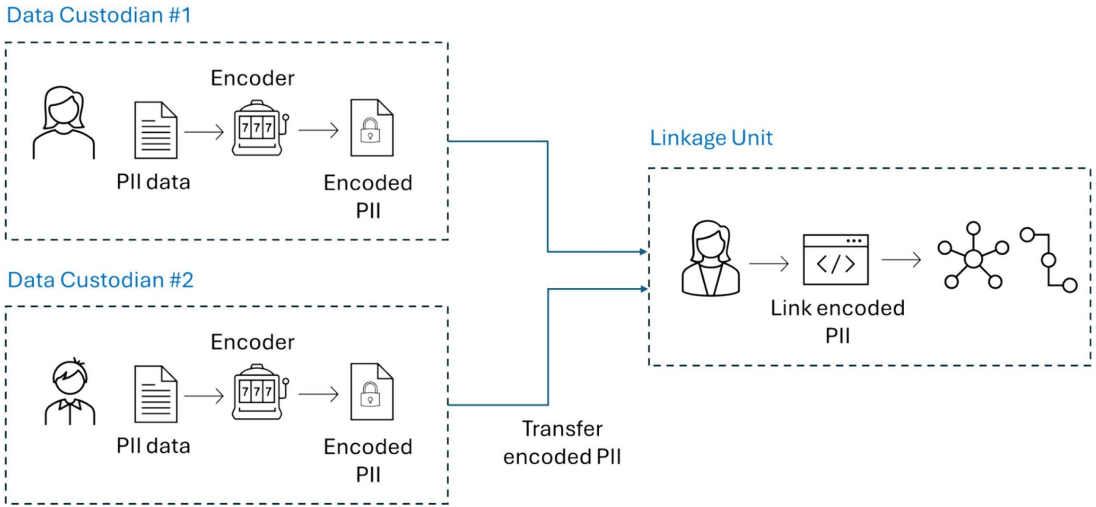
PPRL techniques allow us to:

- Protect privacy and sensitive data
- Link data without exposing personal identifying information (names, DOB, address etc.)
- Assures data provides that direct patient identifiers are not shared\reduced security risks
- Comply with legislation (*Privacy Act 1988* and Australian Privacy Principles)
- Scalable (able to handle large datasets)
- Improves trust of the public and stakeholders

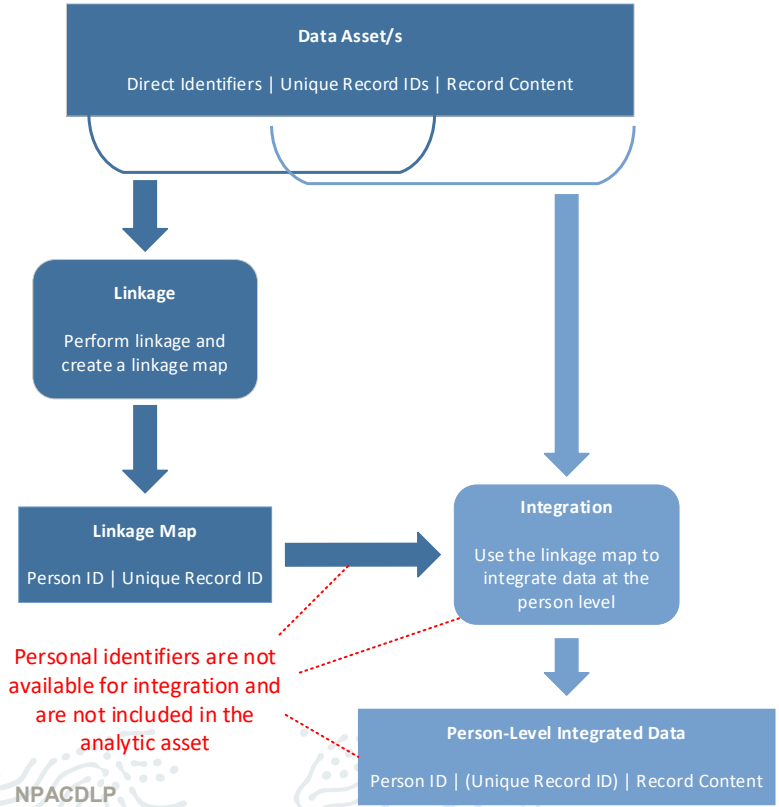
PPRL is used elsewhere in Australia.

PPRL Process

The typical PPRL data flow. Custodians first encode data before transferring to a 3rd party linkage unit, who identifies which records of encoded PII data belong to the same individuals – Source: *Randall, et al. 2024*



About Separation Principle



Separation principle: The concept that to ensure privacy protection, the tasks of linking records and working with person-level statistical data should be handled separately (separation of personal identifiers and content data).

- Linkage functions work with personal direct identifiers to create deidentified maps.
- Integration functions work with deidentified maps and content data.

Separation principle allows us to:

- Prevent persons working with data to view both the personal identifying information (e.g. names, DOB, addresses) together with the content data (e.g. clinical data).
- Prevents association of a person's identity with their medical and social history.
- Achieved through separation of roles and access levels.

Addressing security concerns

Risk	Management
Data breaches and security risks: Linked datasets are at risk of breaches due to system vulnerabilities or external attacks.	Privacy-Preserving Record Linkage (PPRL) and Separation Principle to ensure secure linkage without exposing sensitive data.
Secondary use of data: Data may be used beyond its original intent (e.g., for commercial purposes or research) without explicit consent.	Establish formal Data Sharing/User Agreements to define acceptable use and ensure consent for secondary purposes.
Excessive data collection and storage: Storing more data than necessary increases risks of misuse and breaches.	Data minimisation: Only collect and store the minimal amount of data required for linkage (e.g., essential fields like identifiers). Separation Principle
Linking Records Across Multiple Datasets: Linking records from different datasets may expose sensitive information that wasn't previously accessible.	Use trusted third parties to perform the linkage securely, ensuring data remains protected and used for authorised purposes. Separation Principle.
Long-Term Data Storage: Storing linked data long-term may increase the risk of misuse or accidental disclosure if access controls or encryption aren't properly enforced	• Comply with policies and guidelines for how long data is stored and when it will be destroyed. • Conducting regular audits to ensure only necessary data is retained, and security protocols are current

Key takeaways

Distributed Data Linkage:

- Each linkage unit links its local datasets to a shared "spine file," creating a unique Master Linkage Key (MLK) for each jurisdiction. All states and territories are currently discussing about technical feasibility, privacy, policies, legal issues, and workforce readiness.
- Creating a national data linkage map in Australia has been challenging, but leveraging existing linkage facilities and developing interoperable systems can speed up and expand the process. A distributed data linkage model, which connects multiple units while preserving local data control, could standardise approaches, integrate infrastructure, and enable faster, broader data integration.
- **Benefits**: Improves health service planning, supports research, enhances privacy, and ensures compliance with privacy laws.

We intend to use

- **PPRL**: Techniques that link data across datasets without exposing sensitive personal information, reducing security risks and building trust.
- **Separation Principle**: Separates tasks of data matching and privacy protection, ensuring linkage units only access either content data or identifiers, but not both.

Security and Compliance: PPRL and the SP help manage security concerns and ensure compliance with privacy legislation (e.g., Privacy Act 1988, Australian Privacy Principles). While there are inherent risks, they are manageable. Comprehensive governance frameworks, data sharing agreements, and stringent safeguards are in place to ensure the protection of sensitive data.

References and additional reading

- Australian Government National Statistical Service. Separation Principle. Available from: https://statisticaldataintegration.abs.gov.au/topics/applying-the-separation-principle#:~:text=Under%20the%20separation%20principle%2C%20individuals,in%20analysing%20the%20integrated%20data

- Boyd J, Randall S, Brown A, Maller M, Botes D, Gillies M, Ferrante A. (2020) Population Data Centre Profiles: Centre for Data Linkage. *International Journal of Population Data Science*. DOI: 10.23889/ijpds.v4i2.1139.

- Chi Y, Hong J, Jurek A, Liu W, O'Reilly D. (2017). Privacy preserving record linkage in the presence of missing values. *Information Systems*.71:199-210. doi: 10.1016/j.is.2017.07.001. Available from: https://www.sciencedirect.com/science/article/abs/pii/S030643791630504X

- Christen P. Privacy and data linkage by Peter Christen (part 4/4) [YouTube]. 2021 [cited 2024 Nov 13]. Available from: https://www.youtube.com/watch?v=QeNs18edgvk&list=PLGhCBVcBzlGBD9p0AZh20mTPax3Kvmoow&index=5

- Petersen SJ, Lieberthal RD, Miller KJ, Vakil NH. (2023). Privacy preserving record linkage (PPRL) strategy and recommendations. *PPRL Linkage Strategies Report*. January 2023. Available from: https://www.nia.nih.gov/sites/default/files/2023-08/pprl-linkage-strategies-preliminary-report.pdf

- Randall SM, Ferrante AM, Boyd JH, Bauer JK, Semmens JB. (2014). Privacy-preserving record linkage on large real-world datasets. *Journal of Biomedical Informatics*. 2014;50:205-212. DOI: 10.1016/j.jbi.2013.12.003.

- Randall S, Brown A, Ferrante A, Boyd J, Robinson S. (2024). Implementing privacy preserving record linkage: Insights from Australian use cases. *International Journal of Medical Information*. 2024 Nov;191:105582. DOI: 10.1016/j.ijmedinf.2024.105582. Epub 2024 Jul 31. PMID: 39096591.

- Smith M, Flack F. Data Linkage in Australia: The First 50 Years. (2021). *International Journal of Environmental Research and Public Health*. Oct 28;18(21):11339. doi: 10.3390/ijerph182111339. PMID: 34769852; PMCID: PMC8583508.

- The ACT Health Data Linkage Team. Epidemiology Section. Data Linkage Technical Manual. Available from: https://www.act.gov.au/__data/assets/pdf_file/0008/2162978/ACT-Health-Data-Linkage-Technical-Manual_FINAL.pdf