Custodian-Controlled Data Repository (CCDR)

Supporting the timely, easy and cost-effective access to high-quality linked data

Jan 2025

SA·NT **DataLink**

PHRN Population Health Research Network

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

The Australian Government provides financial support to SA NT DataLink through the National Collaborative Research Infrastructure Strategy (NCRIS).

Previous Data Acquisition Model
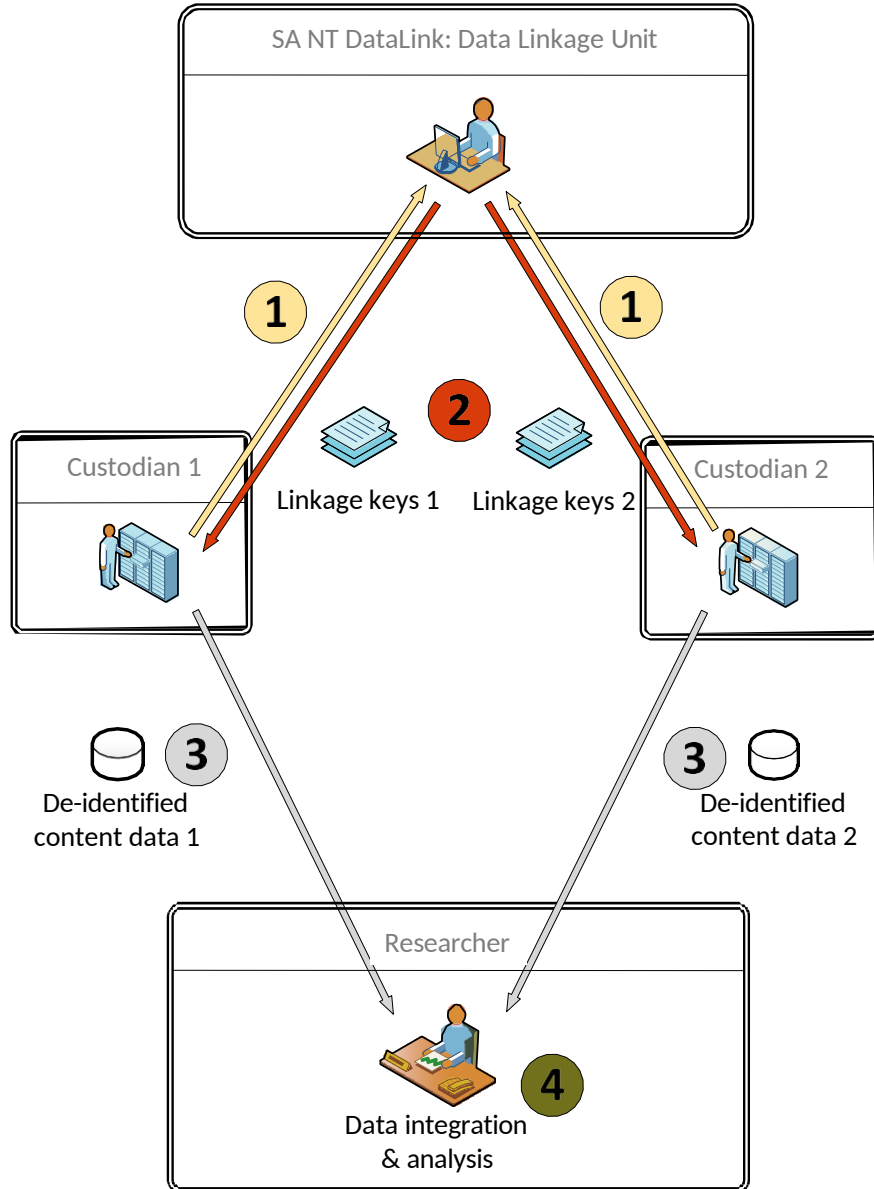
1. Data Custodians provide personal identifiers

2. Data linkage unit produces linkage keys

3. Custodians de-identify data

4. Researcher undertakes integration and analysis

Benefit of Custodian-Controlled Data Repository (CCDR)

Problem:
Reduced Data Custodian resources with increased work loads preventing quick data delivery

→

Solution:
Improved timely delivery and access to approved de-identified data via CCDR

Building on Existing Local & National Infrastructure

- **'Unchanged' data / research governance:**
  - Custodians remain data curators
  - Researchers only receive de-identified project data with custodian and ethics approvals

- **Centralised capacity:** De-identified data jointly managed with SA NT DataLink's Data Integration Unit staff

- **Centralised data:** De-identified data safely stored in Australian Government-approved infrastructure

- **Controlled access:** All access to data in repository controlled by permissions

- **Opt-in:** Use of new model by custodians is voluntary

- **Process quality:** Utilising consistent Australian Government data integration principles to provide high quality and safe data

SA NT DataLink High-Level Data Flow Diagram and Corresponding Security Classifications

The SA NT DataLink site at Level 8 SAHMRI Research Facility is divided into three physically segregated areas;

1. PROTECTED Area (DLU)
   - Accommodates DLU staff and all the ICT systems, equipment and stored electronic information classified at the protected level

2. Low Security Area (Client Services / Administration / Capacity Building)
   - Containing the Clients Services, Administration and Capacity Building functions, and Director's office, and common kitchen and shared meeting rooms on the floor.
   - Houses ICT equipment processing and storing For Official Use Only information
   - Door and lock between the Low-Security Area and the PROTECTED Area, with a sign-in process to control all visitors into the PROTECTED Area

3. Low Security Area (Data Integration Unit)
   - Containing the Research and Analysis functional capacity
   - door and lock between the Low-Security Area (Client Services / Administration / Capacity Building) and the DIU
   - 'OFFICIAL: Sensitive' data is stored, analysed and integrated using remote access to the Custodian-Controlled Data Repository (CCDR) through SURE.
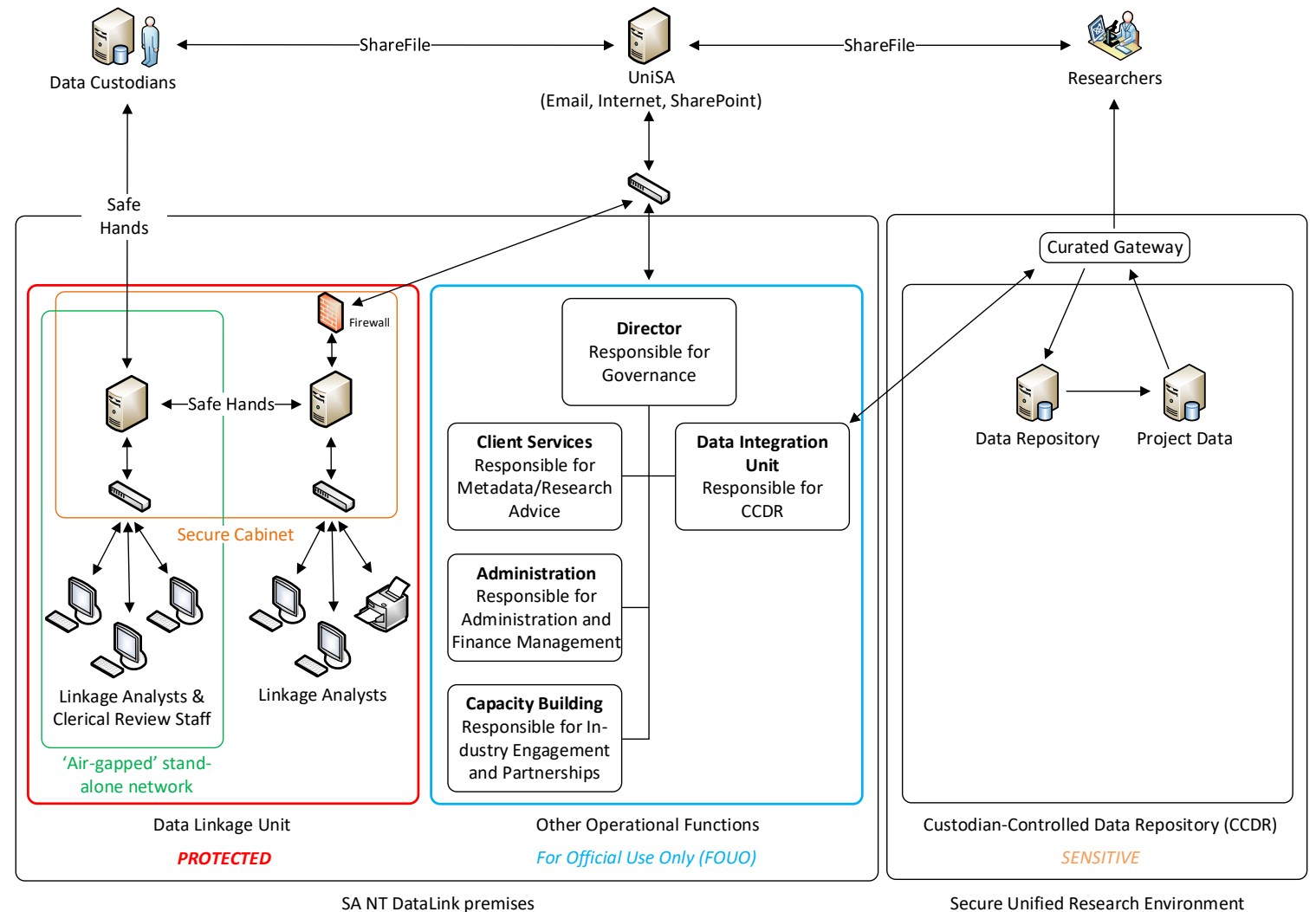
Current Data Acquisition Workflow
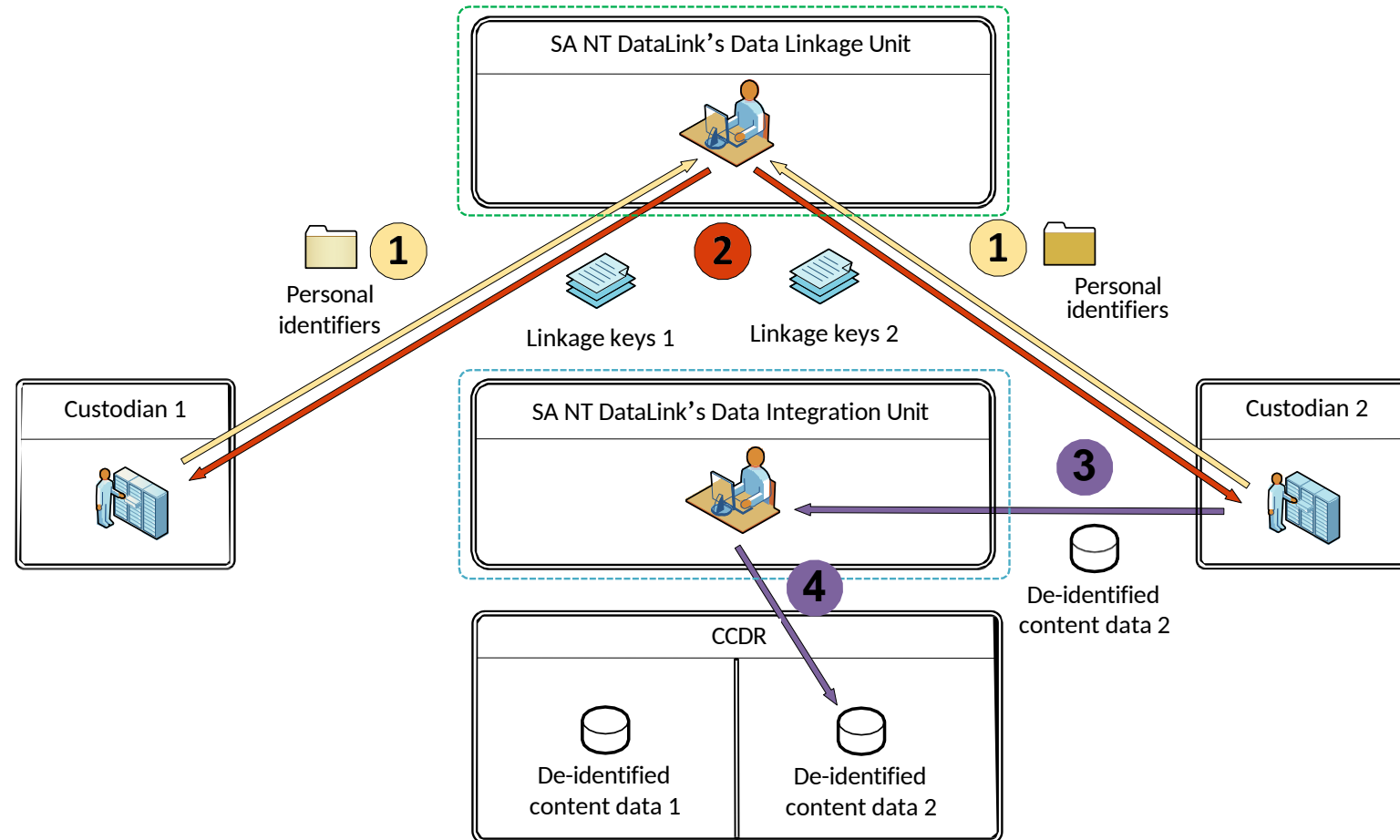
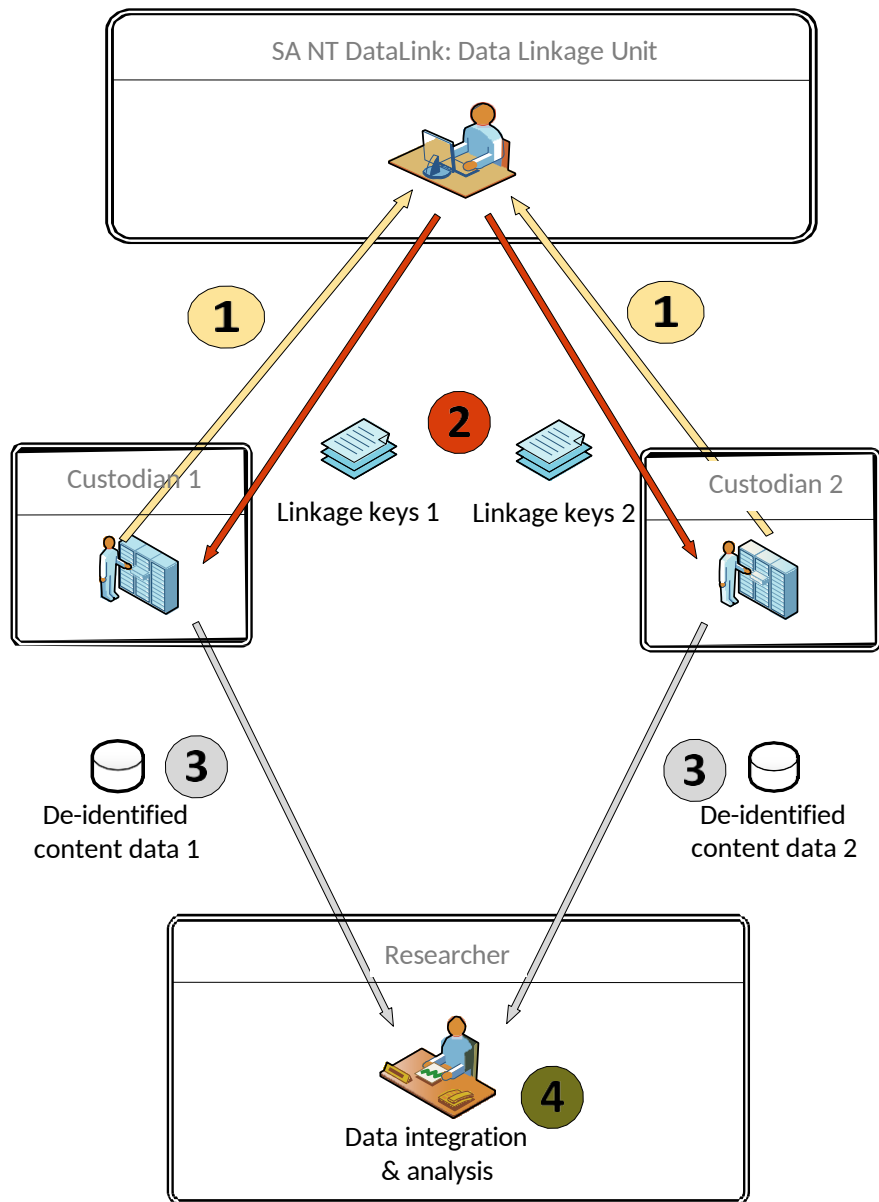1. Custodians provide personal identifiers

2. Data Linkage Unit produces linkage keys

3. Custodians can provide de-identified data into DIU via
   - ShareFile
   - Other Secure file transfer methods
   - Safe hands

4. DIU uploads de-identified data into CCDR

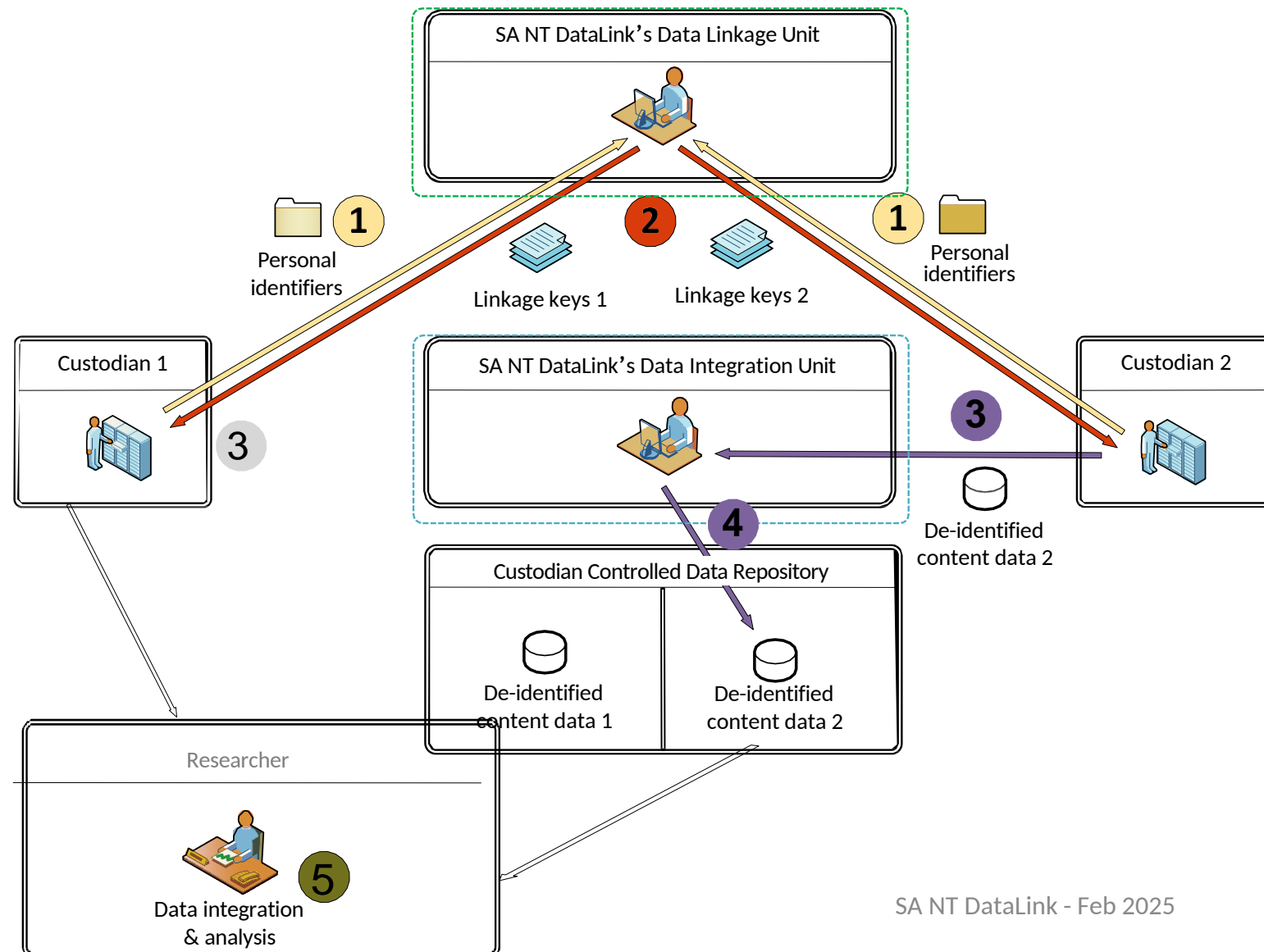# Option 1 – only personal identifiers provided to SA NT DataLink

1. Data Custodians provide personal identifiers

2. Data Linkage Unit produces linkage keys

3. Data Custodians receive project specific linkage keys. They extract and de-identify content data and send it to researcher

4. Researcher undertakes integration and analysis

Option 2 - both personal identifiers and de-identified content data provided to SA NT DataLink

1. Data Custodians provide personal identifiers

2. Data Linkage Unit produces linkage keys

3. Data Custodians provide de-identified content data into Data Integration Unit to be uploaded in Custodian Controlled Data Repository (CCDR)

4. Data Integration Unit stores de-identified data in the CCDR and releases to researcher on behalf of data custodian for approved projects

5. Researcher undertakes integration and analysis

SA NT DataLink - Feb 2025

8

# Current Data Acquisition Workflow Cont.

Data custodian (DC)
- provides demographic data and record IDs to DLU via ShareFile (1a & 2a)

DLU
- Generates record specific linkage identifiers (RSLIDs) (3)
- Sends RSLIDs & record IDs to DCs via ShareFile (1b)
- Send RSLIDs to DIU via ShareFile(2b)

DIU
- Receives de-identified datasets with RSLIDs via ShareFile (1c & 2c) and uploads to SURE via External Curated Gateway (4)
- Staging received de-identified content data and assign to corresponding schema (5a)
- Performs quality checks and confidentiality review on received de-identified content data (5b)
- Integrates reviewed datasets to existing data collection for future use (5b)
- Provide feedback to the data custodian and/DLU on data quality/RSLID issues if required (6)
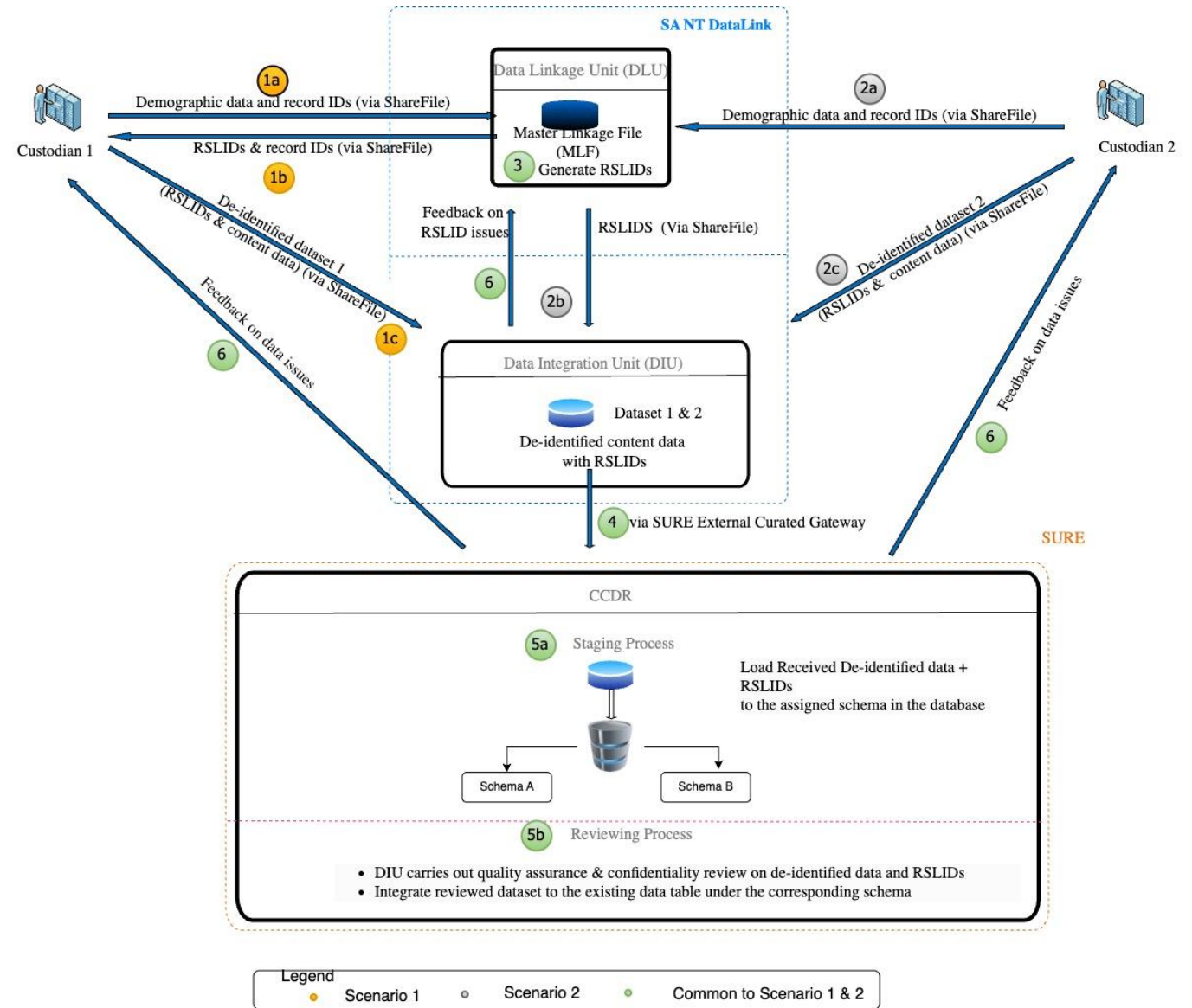
Linkage Project Data Extraction

- A linkage study usually requires approval from:
    - Ethics committee/s
    - Custodian/s
    - SA NT DataLink Director
- Upon receiving SA NT DataLink Director's final project approval,
    - ❖ CS provide cohort specifications to DIU, and they will send cohort identifiers to DLU(Optional)
    - ❖ DLU provides the following project specific internal identifiers and DCIs to DIU via ShareFile:
        - PSLKs (project specific linkage keys)
        - PSRIDs (project specific record identifiers)
        - RSLIDs
        - Data custodian instructions (DCIs)
        - other linkage keys if required
- DIU,
    - ❖ Uploads project specific internal Identifiers + DCIs to CCDR via SURE External Curated Gateway
    - ❖ Performs validation checks
    - ❖ Matches RSLIDS
    - ❖ Checks approved data elements based on DCIs
    - ❖ Creates project data extract/s without RSLIDs
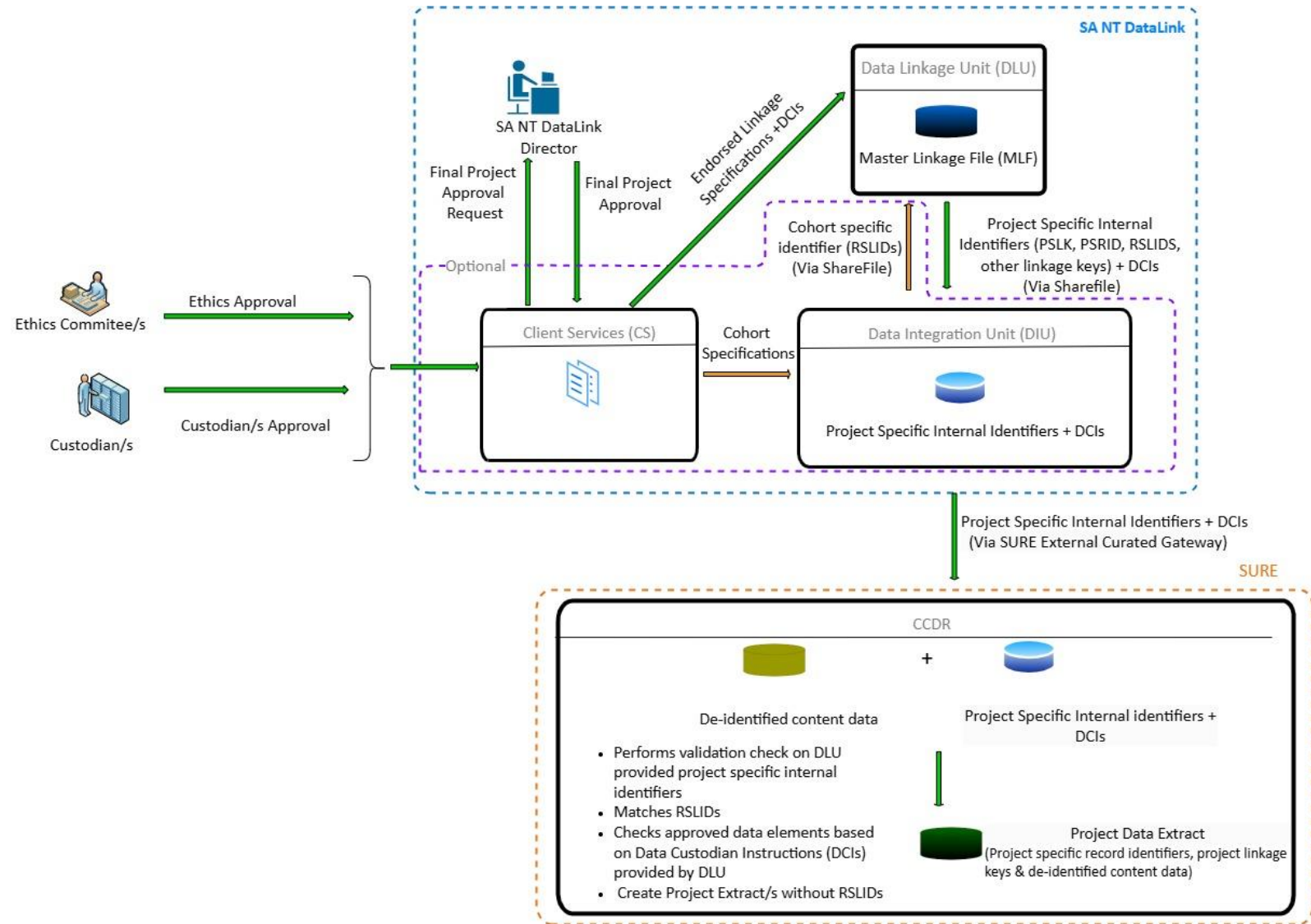
Linkage Project Data Delivery

- DIU uploads project extract/s containing project specific record identifiers, project linkage keys and de-identified content data via SURE Internal Curated Gateway

- DIU logs in into External Curated Gateway and downloads project extract/s data to the local computer

- DIU sends locally saved project extract/s to the nominated recipient mentioned in the DCI using one of the following methods:
  - ❖ Option 1 - DIU directly uploads project data extract/s to researcher's SURE Workspace
  - ❖ Option 2 - DIU sends project data extract/s via ShareFile or Researcher requested secure file transfer method

- DIU permanently delete locally saved project data extract/s using secure file eraser tool
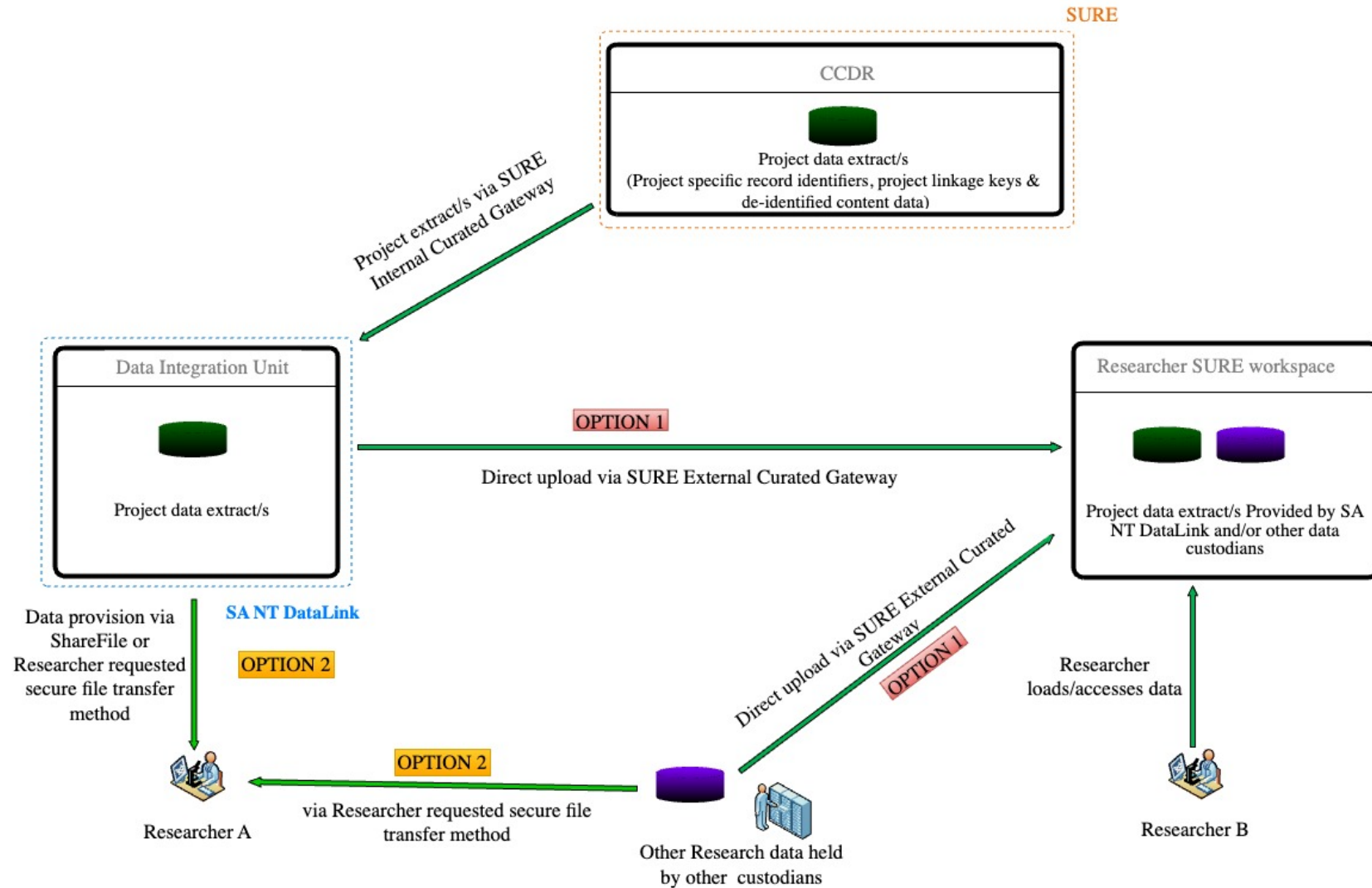
Benefits

☑ **Less complex:**

- Custodians: less frequent provision of de-identified data
- Researchers: less custodians to liaise with for data delivery

☑ **More timely:**

- Researchers: less delays through support of data delivery by dedicated SA NT DataLink staff

☑ **More resource-efficient:**

- Custodians: less staff time required for data delivery

☑ **More validated**

- Custodians & researchers: additional data quality review

**Please contact us if you have any questions
or require more detailed information on workflows and business processes
around the Custodian-Controlled Data Repository**

Data Integration Unit, SA NT DataLink
Health.SANTDL-CCDR@sa.gov.au
www.santdatalink.org.au