

The National Data Network

AGSF

May 2004

The National Data Network..

... will link and share data so that the best use can be made of our national information resources without compromising standards of privacy or intellectual property

The ABS Mission

We assist and encourage informed decision-making, research and discussion within governments and the community, by providing a high quality, objective and responsive national statistical service.

ABS Corporate Plan Objective #1:

An expanded and improved national statistical service



Prof. Fiona Stanley

National Data Network

The National Data Network will link and share data so that the best use can be made of our national information resources without compromising standards of privacy or intellectual property.

Why Do We Need a National Data Network?

Australia has a rich collection of population surveys, cohort studies, population databases and administrative records that are all held separately, so that their national potential is not being realised. These can be linked together to answer important questions about the wellbeing and life chances of Australia's children and young people.



Health Data Integration

- **Infrastructure Development – Data Integration and Metadata.** This area will develop advanced infrastructure facilities for publishing, locating, retrieving and analysing data in a large scale, diverse, distributed information system.

- **Data Linking, Cleaning and Management.** This area will address the data acquisition model of HDI where protocols, algorithms and software are produced to allow clean, linked data to be extracted, analysed and delivered from multiple databases.

- **Security and Privacy.** A key design criterion of HDI permits data custodians the ability to retain real-time control over their data resources, including location, format, storage optimisation, quality control and access control.

- **On-line analytical techniques** to allow the integrated databases to be analysed and explored.



The challenge...

- **Researchers and Policy analysts** want easy and unrestricted access to data.
- **Agencies** increasingly want to share/integrate data with each other
- **Administrative data sources** are under-utilised for research/policy development
- **Custodians** cannot just give data away. They have legal obligations, “customer” expectations, and budget restrictions which must be met. So, they need to be able to expose data in a way that allows them to be confident that their obligations will be honoured.
-

The challenge (cont.)

- **Data Sources** are not well documented, this makes it difficult for users to know whether data is “fit for purpose”
- **Agencies** lack infrastructure for managing data. They need user-friendly tools to manage meta data, design data collections and support data access.

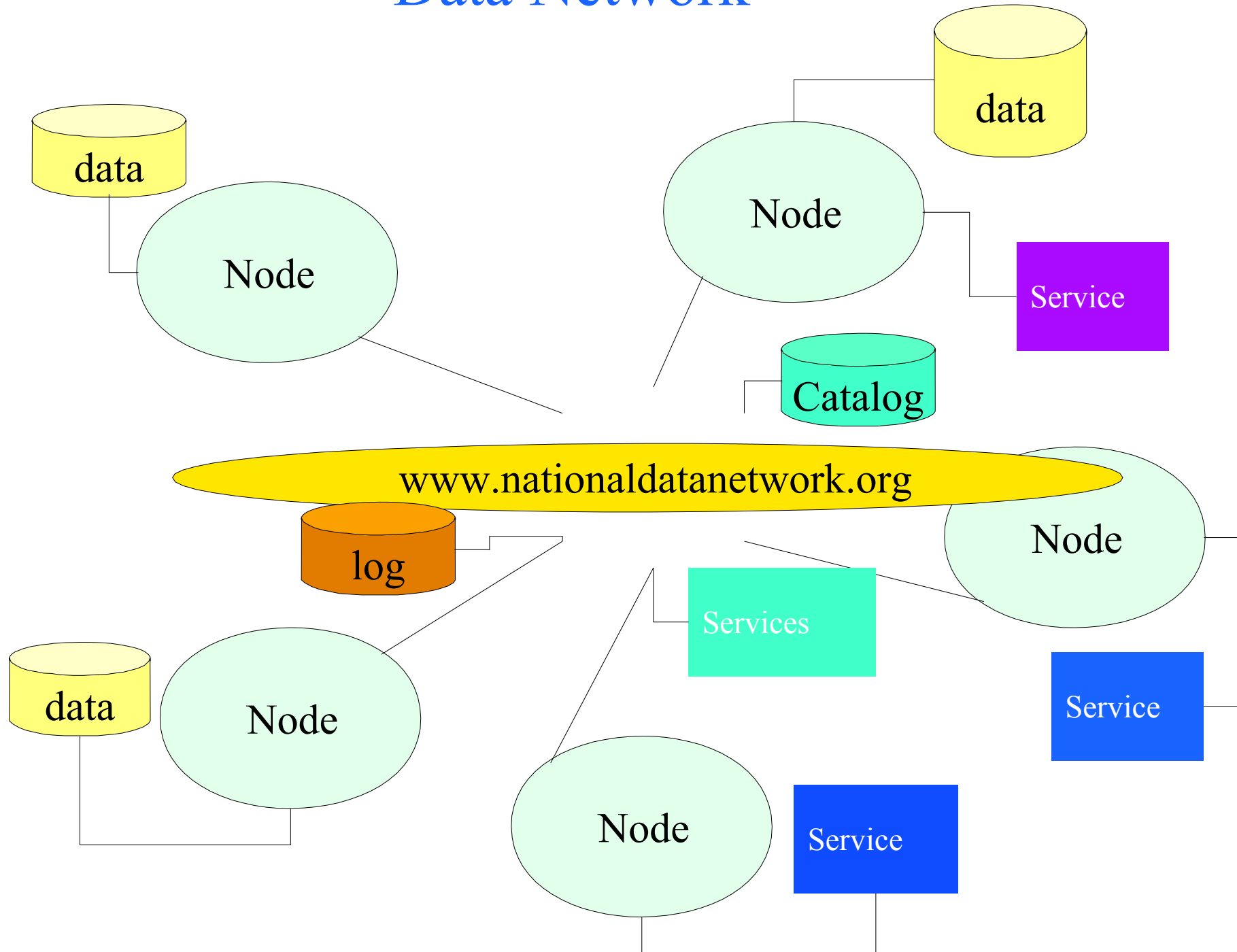
The “National Data Network” ...

- Shared facilities and protocols which are well understood by a large body of custodians and analysts/researchers.
- Streamlined and consistent data access approaches
- Trusted facilities for protecting secrecy and confidentiality
- Will provide a range of services to both people and applications

The “National Data Network” ...

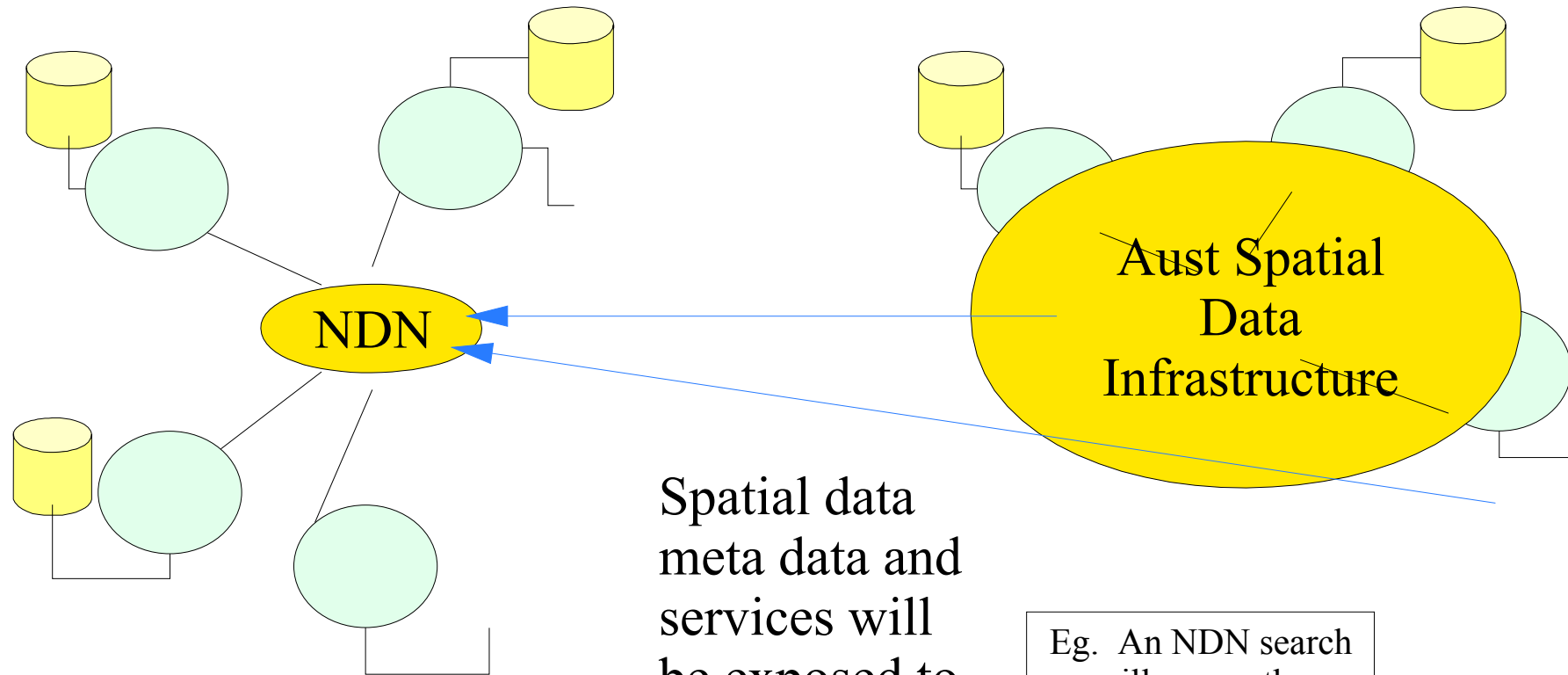
- Data Owners will remain in control of their data.
- Will exist as:
 - a collection of network “nodes”. Each (major) custodian will operate their own node on which their own data is stored
 - A hub which maintains a catalog of sources and services (and network administration services)
- Will service multiple sectors and jurisdictions – health, transport, statistics, education
- must “interoperate” with other networks
-

Data Network



NDN and Aust Spatial Data Infrastructure

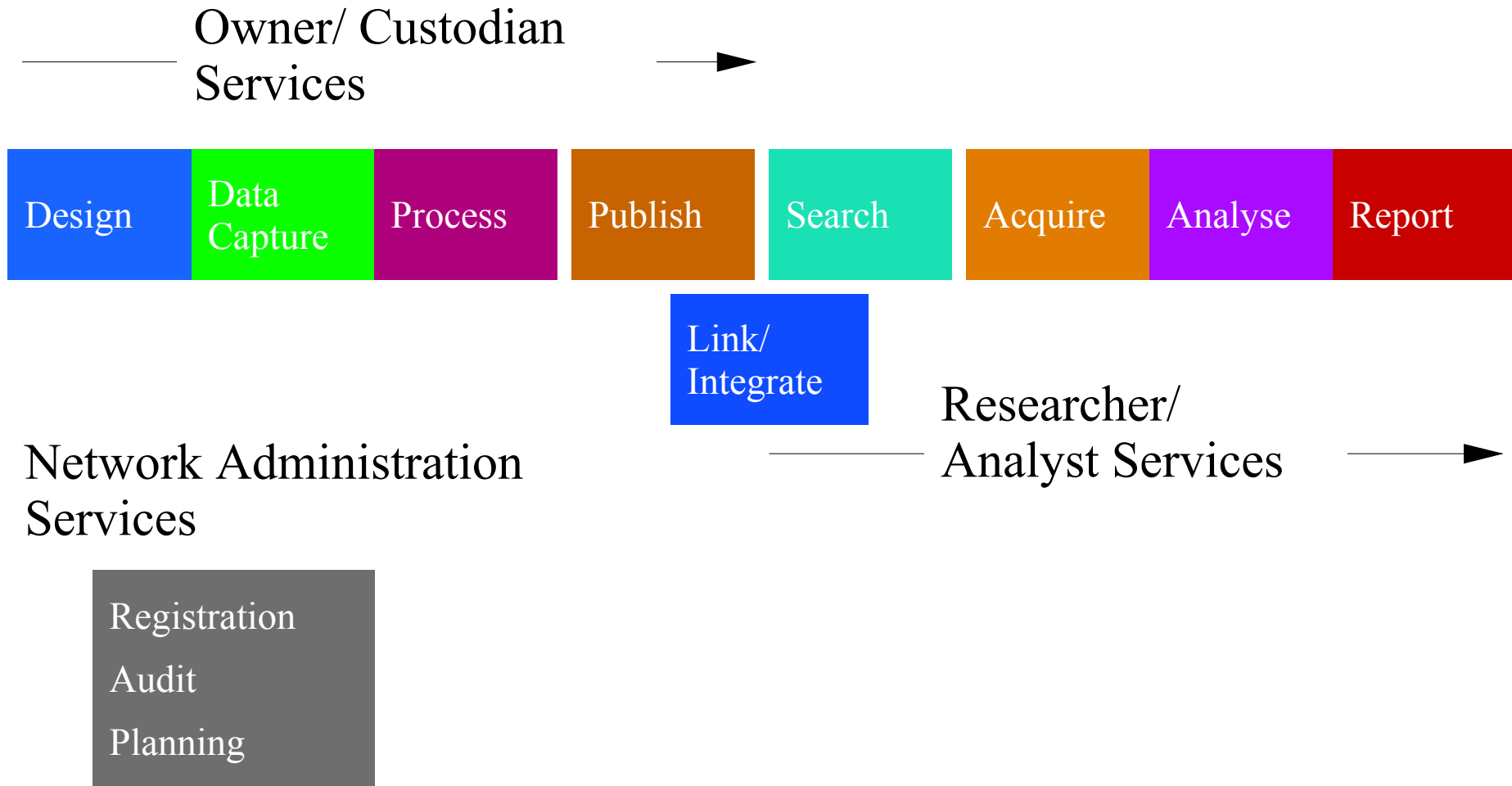
- must be interoperable



Spatial data
meta data and
services will
be exposed to
NDN clients

Eg. An NDN search
will access the
ASDD and return
meta data for Spatial
Datasets

Data Network “Service Framework”



Data Network Services

<ul style="list-style-type: none"> ● Auto coding ● Assisted coding ● Character recognition ● Collection Control ● Respondent Management ● 	<ul style="list-style-type: none"> ● Confidentialise ● Document ● Expose ● Define Access Policy ● Archive ● ● 	<ul style="list-style-type: none"> ● Acquire test data ● Qualify for access ● Extract ● Subscribe ● ● ● 	<ul style="list-style-type: none"> ● Document data source issues ● Document Findings ● Expose Findings ● ● ●
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------



<ul style="list-style-type: none"> ● Data Item Definitions ● Classifications ● Standard Questions ● Sample selection ● Form Design ● 	<ul style="list-style-type: none"> ● Edit ● Seasonal Adjustment ● Estimation ● Imputation ● Aggregate ● 	<div style="background-color: blue; color: white; padding: 5px; text-align: center;">Link/ Integrate</div> <ul style="list-style-type: none"> ● Match ● Link ● 	<ul style="list-style-type: none"> ● Search ● Request ● 	<ul style="list-style-type: none"> ● Supervised Analysis ● Analyse ● Graph ● Tabulate ●
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------

Getting Started..

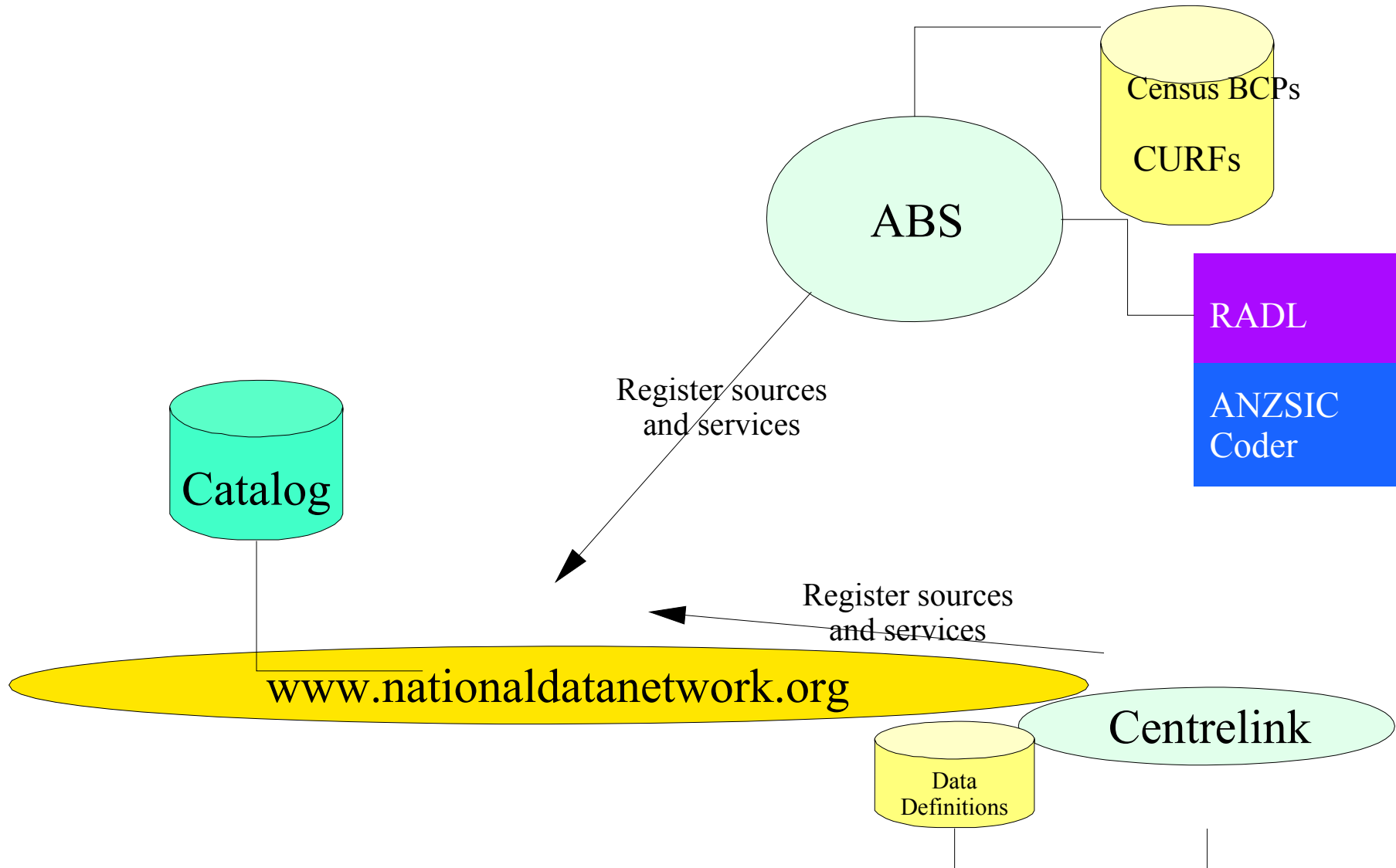
- Start with development of services at the interface between custodian and researcher:



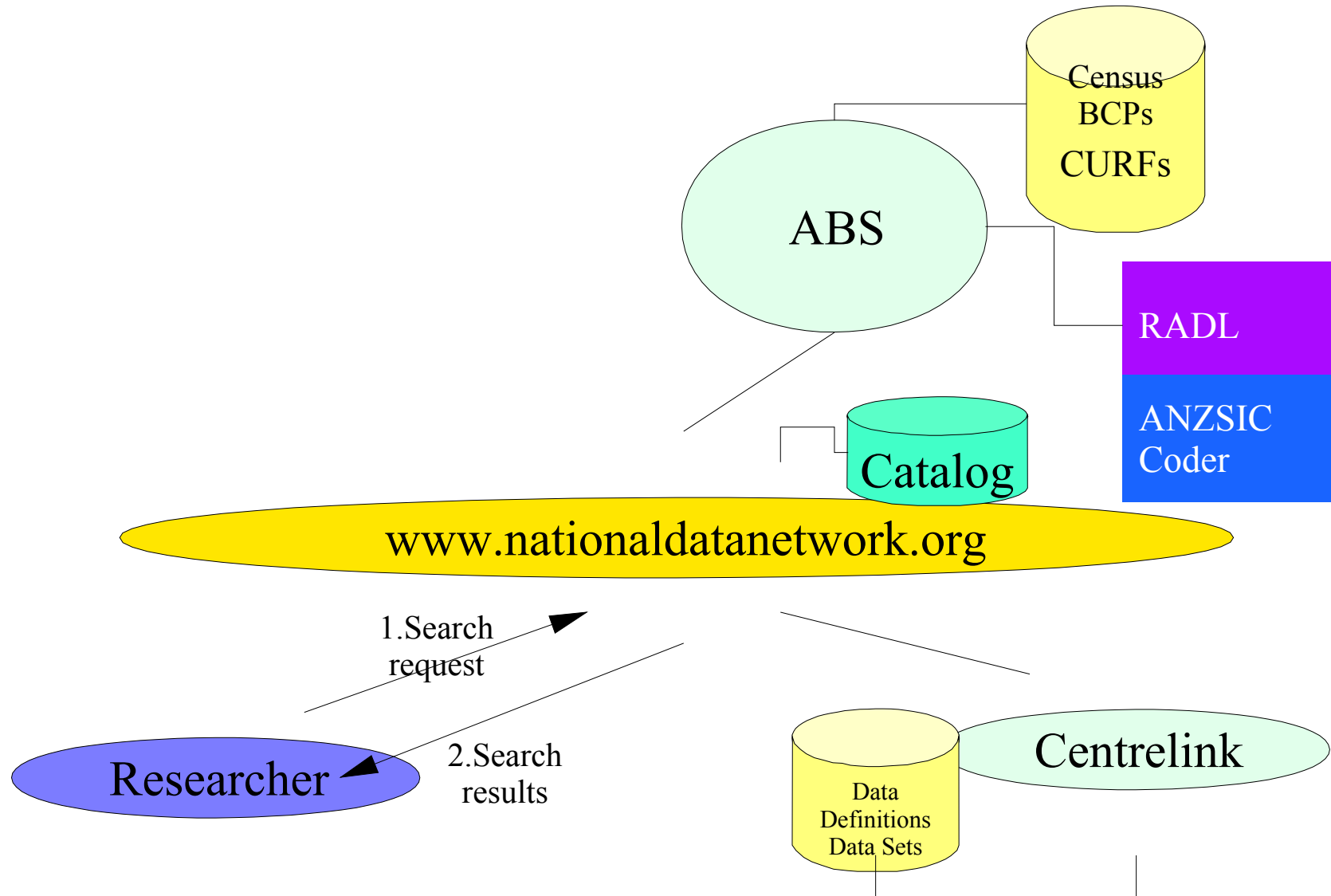
- Develop “base” standards for documenting data sources and “access rules”



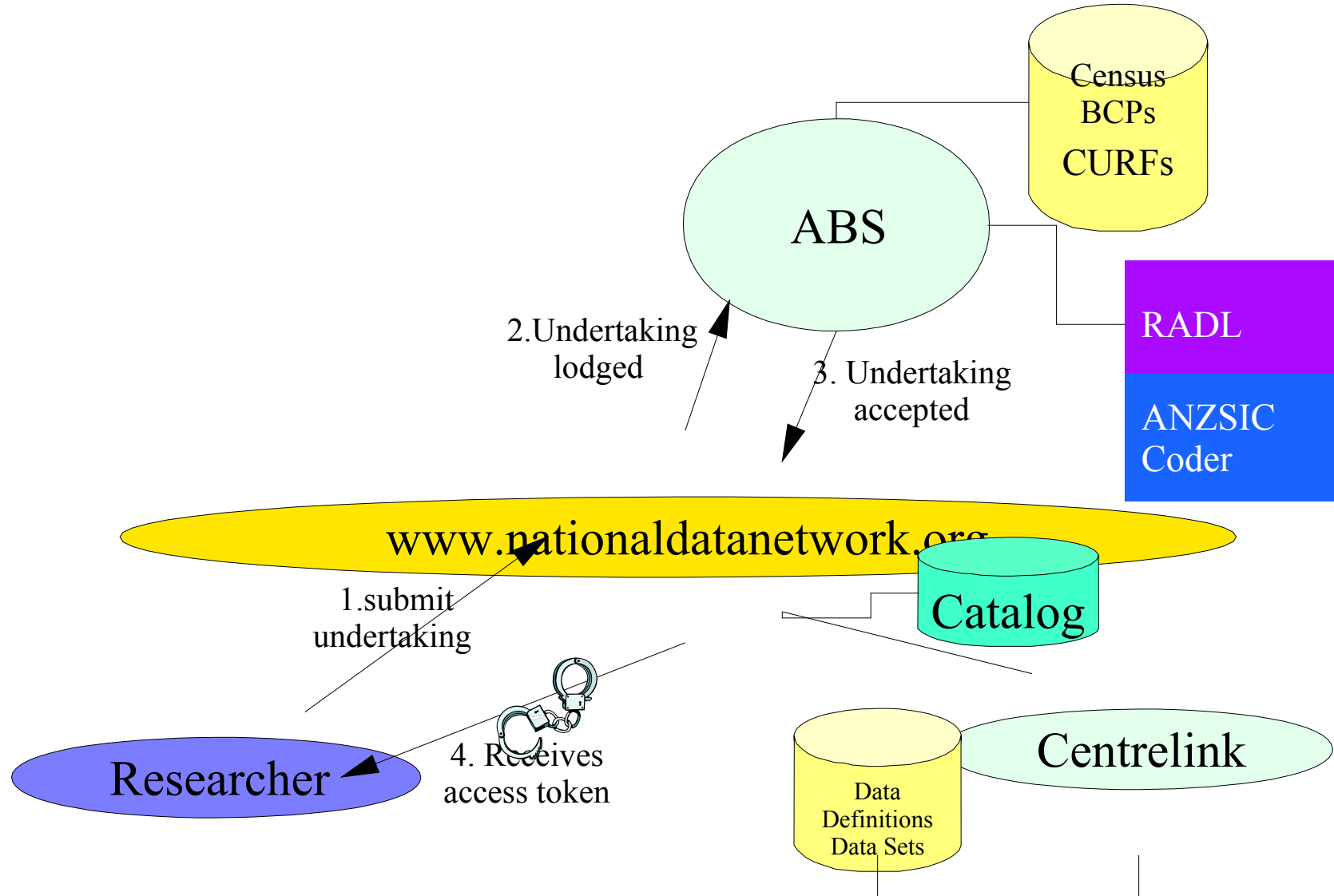
Data Network – resource registration



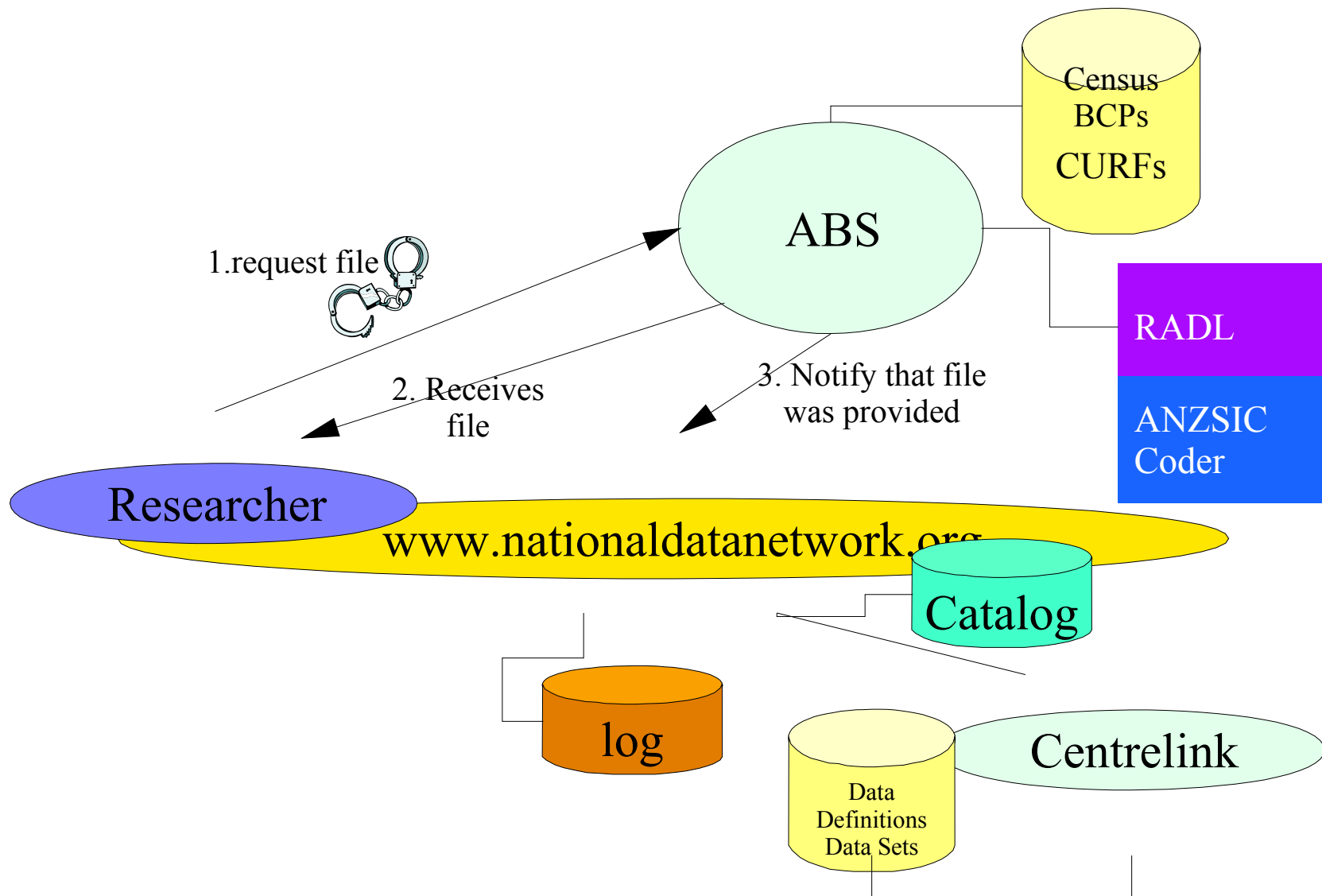
Data Network - search



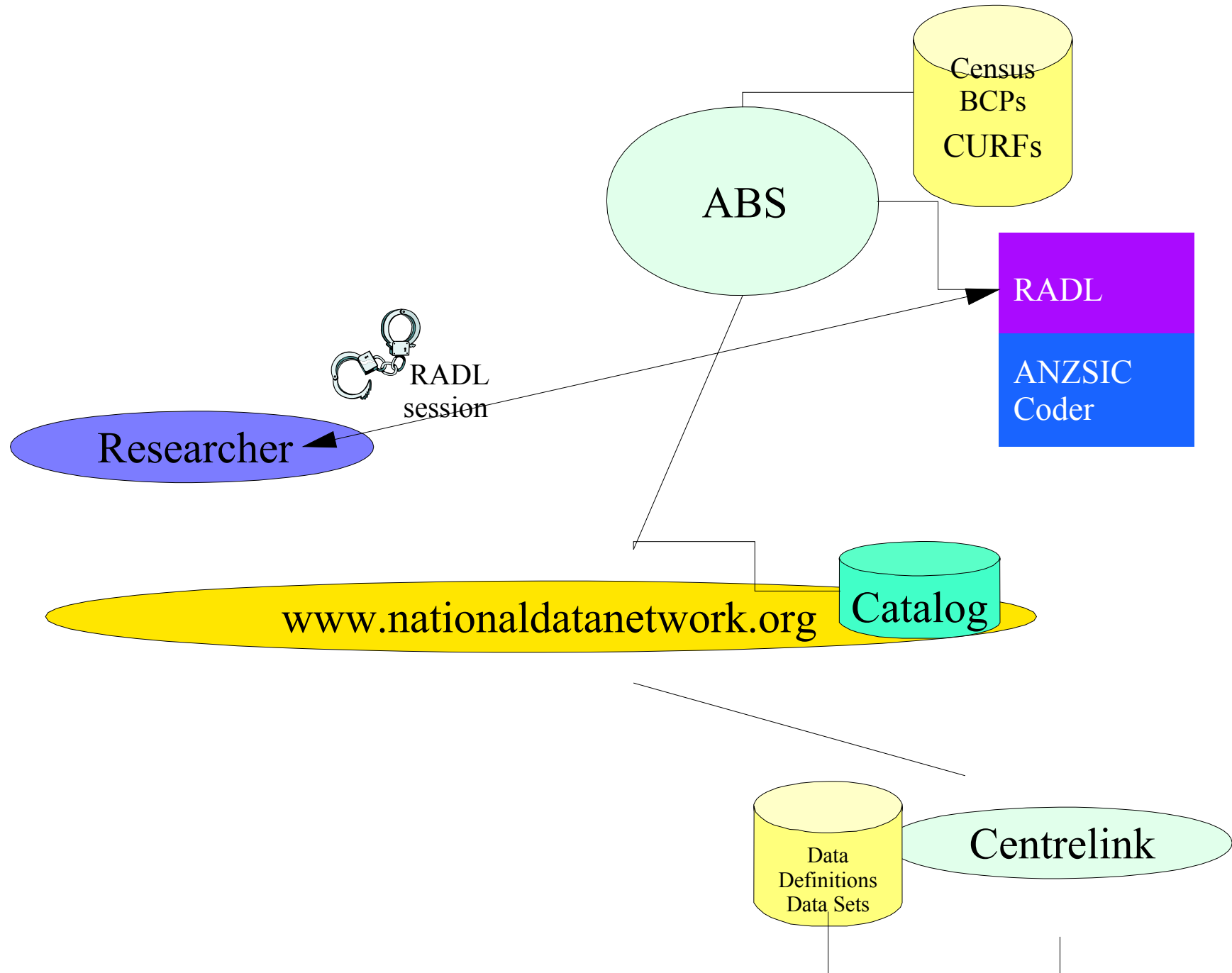
Data Network – undertaking process



Data Network – Access – via download



Data Network – Access – via RADL



Getting Started..

- Support 4 “access classes”:
 - “Unrestricted” (user can freely take the data and do anything with it)
 - “Approval Required” (Researcher must apply for access and agree/meet conditions specified by the data owner/custodian. Example conditions: payment, sign an undertaking..)
 - “Remote Analysis Only” (Researcher cannot acquire the data but, on giving required undertakings, can submit programs to analyse the data. Output subject to vetting by data owner)
 - “Specification Only” (Researcher cannot acquire the data but can submit specifications for interrogations/ tabulations. Scope of specs may be restricted)

Getting Started..

- Provide support for linkage:
 - Provide access to meta data and promote compliance with standards (dataset documentation includes description of linkage policy and process. Meta data will assist users to make “informed use”)
 - access to trusted linking service providers using approaches which protect privacy (eg. “Blind linking” (only identifiers are provided to an independent linking unit which creates chains of link leys))
 - Access to “smart” linking tools

How the Data Network will work...

- **Governing Body** endorses protocols, standards, acquires funds & commissions work
- **Administering Body** “administers” the network – promotion, audit, registration, performance monitoring
- **Custodians** document and expose data sources by registering them.
 - Registration includes documenting “access rules”
- **Service Providers** can register and provide services (which comply with standards)
- **Researchers/ Analysts** agree to comply with “access rules” and provide feedback
-
-
-
-

Getting Started..

Work has started on three “fronts”:

- governance, protocols, priorities, resources
- data source development
- infrastructure development

Getting Started..

Governance, protocols, priorities:

- Establish an **Interim Governing Board (or interim Governance Arrangements?)** and a broader “**member network**” which can be consulted and kept informed
- priority is to agree with key stakeholders on principles, priorities

Getting Started..

Data Sources:

- Enlist demonstration network “nodes” (Centrelink, Qld Gov, SA Gov, ABS, AIHW..)
- pick some “exemplar” projects, identify range of relevant data sources and work with custodians to “expose” the data sources using the Data Network Infrastructure
- Expose a range of “regional” data sources (eg. IRDB, Healthwhiz?)
- ABS Curfs (via RADL and download)
-
-

Getting Started..

- **Infrastructure:**

- Form **infrastructure development consortium** (this has started with ABS, CSIRO)
- Develop “demonstration” version of NDN system
-

NDN Services – internal resource as well as external

- NDN software can be installed inside and outside of firewall. So, NDN services can be used “privately” as part of the agencies infrastructure for managing their own data holdings.
- Use Open Source code wherever possible to reduce costs, promote portability and shared development
- This may improve the value proposition for some custodians as well as make it easier to meet standards and maintain quality.
-
-
-

Aim to have demonstrable system in 2004...

- National Data Network Website established
- demonstration nodes established with “relevant” data
- Services:
 - Search
 - Access services working (including Remote Analysis)
 - Linking support
 -
 -
 -
 -

Data Network– demonstration version

<ul style="list-style-type: none"> ● Auto coding ● Assisted coding ● Character recognition ● Collection Control ● Respondent Management ● 	<ul style="list-style-type: none"> ● Confidentialise ● Document ● Expose ● Define Access Policy ● Archive ● ● 	<ul style="list-style-type: none"> ● Acquire test data ● Qualify for access ● Extract ● Subscribe ● ● ● 	<ul style="list-style-type: none"> ● Document data source issues ● Document Findings ● Expose Findings ● ● ●
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------



<ul style="list-style-type: none"> ● Data Item Definitions ● Classifications ● Standard Questions ● Sample selection ● Form Design ● 	<ul style="list-style-type: none"> ● Edit ● Seasonal Adjustment ● Estimation ● Imputation ● Aggregate ● 	<p>Link/Integrate</p> <ul style="list-style-type: none"> ● Match ● Link ● 	<ul style="list-style-type: none"> ● Search ● Request ● 	<p>Supervised Analysis = RADL</p> <ul style="list-style-type: none"> ● Analyse ● Graph ● Tabulate ●
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------

Principles

- Who can use the network?
- What data will the network serve?
- What conditions can custodians impose?
- What conditions should the network impose?
- How visible/open should the network be?
- Who can be a node?
- Architecture/design principles
-
-
-