ENERGY SECURITY BOARD

Data services delivery model

Consultation paper

December 2022

Anna Collyer

Chair

Australian Energy Market Commission and
Energy Security Board

Clare Savage

Chair

Australian Energy Regulator

Daniel Westerman

Chief Executive Officer

Australian Energy Market Operator

Contents

List of abbreviations

ABS	Australian Bureau of Statistics
ACCC	Australian Competition and Consumer Commission
AEMC	Australian Energy Market Commission
AEMO	Australian Energy Market Operator
AER	Australian Energy Regulator
AREMI	Australian Renewable Energy Mapping Infrastructure Project
ARENA	Australian Renewable Energy Agency
ASX	Australian Stock Exchange
CDR	Consumer Data Right
C4NET	Centre for New Energy Technologies
CER	Customer energy resources
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DAT	Data Availability and Transparency
DCCEEW	Department of Climate Change, Energy, the Environment and Water
DER	Distributed Energy Resources
DISER	Department of Industry, Science, Energy and Resources
ECA	Energy Consumers Australia
ESB	Energy Security Board
EV	Electric vehicles
LV	Low voltage
MSATS	Market Settlement and Transfer Solutions
NEAR	National Energy Analytics Research Program
NEL	National Electricity Law
NGL	National Gas Law
NMI	National Meter Identifier
NSW DPIE	New South Wales Department of Planning and Environment
PV	Photovoltaic
SA DEM	South Australia Department for Energy and Mining
VIC DELWP	Victoria Department of Environment, Land, Water and Planning

1 Executive summary

The Energy Security Board (ESB) was tasked by Ministers in late 2021 to deliver the Data Strategy. Designed to unlock data to meet changing needs in the energy transition and support the long-term interests of energy consumers in a digitalised future, the Data Strategy seeks to address a wide range of regulatory, technical, and organisational challenges, targeting access to a range of datasets, across several workstreams.

This consultation paper focuses on facilitating better access to one important subset of energy data: data already held by the Australian Energy Market Operator (AEMO), that policy makers, planners, and researchers need to improve consumer outcomes in the energy transition. These data sets have significant value, particularly as they include data on consumer meters and consumer energy resources (CER). This data is critical to better understand how different consumer behaviours and needs are changing and being impacted by the energy transition, informing forecasting, investments, new services, and consumer protections.

Legal changes are being advanced to reduce regulatory barriers to sharing AEMO data through the Initial Reforms workstream.[1] However, the Data Strategy identified that, without also resolving practical constraints on safe data sharing, including resources, processes, and capabilities, these reforms would have limited impact. The growing scale and complexity of modern datasets requires expanded skills, systems, and processes to facilitate safe access and create value and insights. Development of clearly defined data services will give effect to the improved data access enabled via the *Initial Reforms* proposals.

Energy agencies already provide some data services for data users. User experience with these services highlights limitations in current resources and gaps in supporting processes across the initial identification of datasets, negotiation of access to data, systems to facilitate data access, and capabilities for analysis.

Case studies have identified many instances where gaps in services have created barriers, constraining, or delaying access to the data-driven insights needed by policymakers, planners, and researchers. In many cases access to data proved impossible or took years to achieve and stakeholders frequently invest in trials, surveys, or other expensive alternatives to seek similar data.

This consultation paper addresses the question: **how should data services be provided and governed in the energy sector to meet stakeholder needs in policy, planning and research and to evolve flexibly as new data needs emerge?**

The data services under consideration are:

Data Dictionary	User requirements	Custodian-ship	Contracting	Data access	De-identification	Compilation/ analysis	Visualisation/ insights
Publishing and maintaining a simple catalogue of available data (a light touch transparency measure to allow users to identify data)	Engaging actively with data users on their needs – providing early advice to scope and manage risks in proposed on work, as well as understanding collective user requirements.	Working with data holders to curate data to ensure its suitable for sharing and analysis. Could include sharing guidelines and formats for data custodians and "data hosting" services.	Negotiating agreements and data protections for specific data sharing and services, as well as management of published standard terms	Support systems and processes for safe sharing of data with a wide range of users (e.g.: curating data through secure data labs, dashboards and portals).	Ensuring that privacy and sensitivity requirements are met before data is shared. Including development of robust and transparent de-identification processes.	Expertise to develop insights and value from data. Includes linking datasets and a range of computational approaches.	Supporting ways to present and access insights, Including through publication and tools.

---

[1] As part of developing the strategy, ESB undertook a review of existing data regulatory arrangements in the energy sector, led by KWM and Galexia. The Legal Report provided specific legal recommendations adopted in the strategy, including initial reforms to reduce key regulatory barriers in the short term, development of common data sharing guidelines, and more fundamental reforms to develop a new fit-for-purpose energy data framework. This Consultation paper should be read in conjunction with the Legal Report. The *Initial Reforms* proposals are currently being considered by Energy Ministers.

Services to date have demonstrated that the right governance model is critical to build and maintain these resources, processes, and capabilities to ensure they:

- focus on responding to and meeting stakeholder needs to inform policy, planning and research for public benefit, and ultimately creating benefits for consumers
- develop flexibly and adapt and innovate over time in response to changing needs and opportunities.

This paper assesses the following five potential delivery models and governance options against evaluation criteria and draws on case studies of similar arrangements from the United Kingdom and other sectors. The models are:

- Giving a data services mandate to AEMO to leverage internal capabilities.
- Adding new dedicated data services capabilities to AEMO, with governance from stakeholders.
- Giving a data services mandate to the National Energy Analytics Research Program (NEAR), an existing energy data research collaboration.
- Creating a new independent data services body with a focused mandate.
- Building data service capabilities in all existing agencies to work collaboratively.

Some of the models show greater benefits in the short-term, particularly the first three models which leverage existing capabilities to facilitate greater access to AEMO data by the other market bodies, jurisdictions, and policy makers. Other models demonstrate some potential for longer term benefits but require more complex development of new organisational structures, which would require more detailed design and cost-benefit analysis and significant time to implement.

To facilitate stakeholder consideration and views of more extensive service provision models, providing the opportunity to realise additional benefits at some additional costs, the paper sets out proposed priorities for delivery over the short and medium term.

Based on these findings, a phased implementation approach is proposed, which would initially develop the "Dedicated unit within AEMO" model as a short-term solution, but with a fixed review at 2-3 years to consider the need to adapt the model to meet further needs. This approach targets short-term services to access initial priority data sets (held by AEMO) that provide the greatest immediate value during the transition. This would leverage benefits from proposed *Initial Reforms* and build on existing capabilities to develop services needed quickly and efficiently. This phase would develop initial infrastructure needs, such as the data dictionary and de-identification methods, but would be designed with a view to adapting to changing needs over time and allow for trials of different service approaches. The review would consider learnings from services developed to date, establish what is working and what is not, and assess whether longer-term needs require aspects of other governance models considered in this paper, such as the independence of separate or different body, or greater engagement across a network of organisations.

The Energy Security Board invites stakeholders to share their views on the approach for developing energy data services by responding to the questions in this consultation paper (see **Appendix A**). Submissions on this paper are due by **13 February 2023**. A stakeholder webinar will be held to present on these options on **7 February 2023**.

2 Context

2.1 The ESB Data Strategy

The ESB has been tasked by government to implement the Data Strategy to unlock data as an enabler in the energy transition.[2] The Strategy plays a critical role, integrated with the broader energy reform program. It provides overarching consideration of the energy sector's existing and future data needs, supporting the needs of consumers, industry, and policy makers in the energy transition.

In July this year, the ESB released a consultation paper on Initial Reforms recommended in the Data Strategy.[3] These Reforms seek to reduce regulatory barriers to allow greater access to existing data held by AEMO for public-good purposes, such as policy, planning and research. This is the first stage of reforms to provide short-term benefits in the market transition. Barriers will be further reduced through the development of Common Guidelines to support data access and a broad set of reforms to develop a New Energy Data Framework, fit for purpose in a digitalised future.

The Data Strategy highlighted that regulatory reforms and new data rights alone would have limited impact without supporting services and capabilities to facilitate this access. Resource constraints and gaps in supporting processes remain key barriers to existing data requests. The growing scale and complexity of modern datasets requires expanded skills, systems, and processes to facilitate safe access and create value and insights.

This paper discusses the need for new data services capabilities proposed in the Data Strategy. It explores a range of alternative governance models to support the development of these services and ensure that stakeholders can safely and effectively access growing datasets to create value for consumers. This paper is limited to discussing capabilities and governance needed to facilitate access and benefits from existing data available to market bodies and to be able to grow data capabilities flexibly over time working collaboratively with stakeholders across the sector.

The Data Strategy recognises there is a much wider set of data gaps already in place and that data needs will continue to change and develop as the market transitions and digitises. A range of key data gaps has been identified, with activities developed to explore options to address each of the top five priority gaps, with work underway by ESB and market bodies to deliver feasibility studies on options to address these needs. These priority gaps focus on supporting customer benefits from integrating consumer (distributed) energy resources, including data on:

- growth of electric vehicle charging infrastructure
- network visibility for market planning
- bill transparency to support consumer protections
- over-voltage impacts to support investment in smarter grids
- improving customer metrics to understand changing needs, behaviours, and impacts.

How these datasets could and should be assessed and managed will vary widely and, depending on these solutions, datasets may or may not be facilitated by data services. This data services workstream does not seek to pre-empt this work and these datasets are explicitly out of scope at this stage. The data services workstream aims to deliver a flexible and scalable model which can respond to changing needs over time.

---

[2] Energy Security Board (2022), Data Strategy; Department of Climate Change, Energy, the Environment and Water (2021) – Energy Security Board Data Strategy.

[3] Energy Security Board (2022), Data Strategy – Initial reforms consultation paper. Note, the *Initial Reforms* proposals are currently being considered by Energy Ministers.

Data services provide resources to facilitate access but cannot alone resolve all problems of data access. In many cases, data gaps will not be resolved without more direct regulatory or technical interventions and targeted consideration is needed as to the benefits of doing so. In the wider context of the Data Strategy, data services aim to improve access to existing datasets and facilitate the *Initial Reforms* already under consideration. However, further priority datasets need to be addressed separately with an assessment of the policy implications and delivery options.

Table 2 in Chapter 5 provides a further breakdown of how the Data Strategy is addressing the range of data gaps.

3 The purpose of this consultation paper

3.1 Objective and scope

This consultation paper compares a range of service delivery and governance models developed for delivery of data services, for consideration by stakeholders and interested parties.

The initial focus for development of data services includes:

- access to existing datasets held by AEMO
- enabling the ESB *Initial Reforms* proposals, providing greater data access and value to policy bodies and public-good research.

Data services models however should also be explicitly assessed for their capacity to adapt, scale, and grow over time, responding to emerging data needs and priorities.

The data services delivery model should align with the Data Strategy objectives – namely, managing changing data needs in the energy transition and optimising the long-term interests of consumers in a digitalised future – and the new Energy Data Policy Principles:

Frameworks governing management and use of data across the energy sector should:

1. *Drive outcomes consistent with the energy market objectives and the long-term interest of consumers.*
2. *Ensure appropriate privacy and security safeguards are maintained.*
3. *Capture benefits of a transparent, innovative, and informed digitalised energy market.*
4. *Be fit-for-purpose, flexible and cost-effective for a digitalised market.*
5. *Be coherent with wider national reforms on data (such as the Consumer Data Right and the Data Availability and Transparency Act).*

This consultation paper discusses high-level governance models and sets out the strength and weaknesses of each model for stakeholders to respond to. The paper does not discuss implementation details for any of the service delivery models proposed, however, a phased approach for initial delivery of data services over the next 2-3 years is proposed for consideration in Chapter 10.

3.2 Approach

This paper is based on the ESB review of stakeholder needs and insights drawn from case studies from other industries and approaches in other countries. ESB has drawn on expertise from Accenture to support development of these models.

As part of this work, Accenture carried out interviews with four groups of stakeholders, see below:

- **Energy market bodies:**
  — Energy Security Board (ESB)
  — Australian Energy Market Operator (AEMO)
  — Australian Energy Market Commission (AEMC)
  — Australian Energy Regulator (AER)
  — Energy Consumers Australia (ECA)

- **Government departments and bodies, including**:
  — Former Commonwealth Department of Industry, Science and Resources (DISER)/Department of Climate Change, Energy, Environment and Water (DCCEEW)
  — Commonwealth Department of Treasury
  — Australian Bureau of Statistics (ABS)
  — New South Wales Department of Planning and Environment (NSW DPIE)
  — Victorian Department of Environment, Land, Water and Planning (VIC DELWP)
  — South Australia Department for Energy and Mining (SA DEM)
  — Australian Competition and Consumer Commission (ACCC)

- **Research and industry, including**:
  — National Energy Analytics Research Program (NEAR)
  — Centre for New Energy Technologies (C4NET)
  — Project SHIELD (funded by the Australian Renewable Energy Agency (ARENA))
  — Powercor

- **International energy data specialists and regulatory bodies, including:**
  — Energy Systems Catapult (UK)
  — Ofgem (UK)

Stakeholders identified potential datasets to unlock and data services to provide, as well as impacts and benefits of providing new data services by various bodies. Stakeholders provided views on the barriers they face in accessing priority datasets and the costs and feasibility of moving to a new data service model. Stakeholders reflected on the learnings of existing energy data service delivery models, such as the NEAR Program and C4NET. Insights were drawn from case studies from other industries and approaches in other countries.

The initial round of consultations identified a long list of models for future data service delivery. This paper considers five options for more detailed comparative assessment.

4 Why new data services are needed: barriers across the data services value chain

Improved access to data is essential to manage the energy transition

Australia's energy system is changing. With a rapidly shifting generation mix, strong growth of CER and an increasingly active role for energy consumers, maintaining a stable and efficient system presents new challenges. New energy technologies are increasingly dependent on use of data and digitalisation, driving the creation of new types of data, with volumes of energy sector data already growing exponentially.

This new and complex operating environment requires a fresh approach to data services and digitalisation. Decision makers across the sector have new and growing needs for data and frequently face barriers, creating risks for effective markets, planning, policy, and system operations, with subsequent risks to affordability, reliability, and security.

The ESB's Data Strategy highlighted that improving access to data and data management is critical to managing growing needs and risks in the energy transition.

Data held by market bodies is increasing rapidly, as more complex and variable services require more modern metering, monitoring and management systems. The significant volume of new types of energy data from CER, storage, smart appliances, electric vehicles, and flexible demand in our energy system increases presents an invaluable opportunity.

The Data Strategy found that decision makers across the energy sector frequently lacked data they need to make informed and effective decisions, particularly given the current rapid pace of the energy and market transition. The Data Strategy identified datasets held by AEMO, AER and other entities across the energy system that are currently difficult to access. If these datasets could be safely accessed, it would provide public benefits, including through improved planning, policy, and related research.

Rights to access data are a key barrier but reforms are underway

In Australia, data access has been a focus across sectors since the Productivity Commission's *Data Availability and Use* report in 2017, including through initiatives, such as Open Banking and the consumer data right, and the DAT Act (2022). The ESB's Data Strategy recommendations, endorsed by Ministers in December 2021, laid out a coordinated approach to addressing data gaps in Australia's energy system, implementing a recommendation from the Finkel Review in 2017. The objective was to balance the need to develop a transparent, innovative, and informed digitalised energy market whilst ensuring appropriate privacy and security safeguards are maintained.

The ESB reviewed existing data regulation in its [preliminary legal report](#) and established that it has not kept up with the needs of the sector.[4] Many regulatory barriers were due to inconsistencies and uncertainties in current regulation, or regulation that was not up to date with consumer need and benefits. These barriers are further exacerbated by liability arrangements and social licence concerns. Many of the existing institutions do not currently have clear rights to access the data they need or share data efficiently and in line with consumer benefits. As part of implementing the Data Strategy, the ESB is pursuing a range of regulatory reforms to clarify data access rights and better align them with both current needs and the future vision of a sector that exemplifies the importance of data as a public good and a critical mechanism for addressing future complexities in the energy, social and economic spaces.

---

[4] KWM and Galexia ([2020](#)) – ESB Data Strategy – Preliminary legal report.

Regulatory reforms alone are not enough – data services are critical to facilitate access and value.

Beyond regulatory barriers, the Data Strategy identified practical barriers including privacy risks, resource constraints and other reasons that make it difficult to open data access.

Even in cases where access rights exist or can be negotiated, the Data Strategy found that barriers remain through clear gaps in supporting data services – specifically resources, processes, and capabilities, which help unlock benefits from the data while maintaining information privacy. Releasing data, even where the rights to do so are clear, requires extensive resources to curate datasets, negotiate agreements, de-identify and link data, as well as maintain systems to manage and deliver large-scale datasets.

Similarly, gaining value from data usually requires not only access to data but expert analytical capabilities with in-depth knowledge of both energy market challenges and data science. These expert capabilities are scarce and difficult to build and maintain across most data holders and users in the energy sector, with experts already serving existing priorities.

Existing data services are resource constrained, limiting access and benefits

There are several organisations currently providing energy related data services, including AEMO, C4NET and NEAR.

Many data services (such as curation, negotiation of agreements, linking and de-identification) require parties to have direct access to the data, often in its raw, unprotected form, which can limit these services to the data holder. AEMO is a key data holder and frequently has unique access to datasets, with other data providers unable to supplement these needs or only able to provide subsets of related data. However, AEMO's capacity to provide data services is resource-constrained and many data service demands exist outside its current regulated and funded priorities. Research projects and data requests often struggle as resources from all parties are not available or able to be freed up to complete the project. Limits on these services also mean that most agreements and processes are bespoke, with AEMO and other market bodies, jurisdiction or researchers having to invest extensive time developing unique arrangements, adding cost and delays. This creates friction and inefficiency in outcomes from publicly funded research and trials.

Creating value from data increasingly requires specialist skills

Organisations that need data often lack in-house analytical capabilities to use these very large complex datasets, and often need to partner with data researchers or service providers. NEAR and C4NET are examples of bodies that were established in part to support these gaps.[5] While the work of both organisations has had impact, including in providing advanced analytics services, their access to data and resources remains constrained. Both NEAR or C4NET have struggled to resolve bottlenecks affecting their access to data, limiting their greater potential value to customers and stakeholders in many cases. Regulatory reforms to support greater access and sharing would have limited impact if data services resources and capabilities are not resolved.

Barriers exist across the data services value chain

Policymakers and innovators consulted for this paper outlined that to understand change in the market they need to be able to extract insights from energy data, to show the impact of new technologies, interventions, improve services and planning and better understand consumer behaviour to inform future policy decisions. However, the data services required to obtain these insights are currently limited: both data holders and data users frequently lack resources, capabilities, streamlined processes and mandate to deliver these services. These barriers impede the timeliness and relevancy of outputs which are often subject to further secondary disclosure restrictions. Furthermore, the analysis and insights can be limited by a lack of visibility and understanding of the raw data inputs, with policymakers and innovators often lacking the right contextual knowledge to target effecting analytical requests or draw informed conclusions.

---

[5] Centre for New Energy Technologies (2022); NEAR (2022)

5   What data services are required?

The Data Strategy identified that gaps in data services resources and capabilities were often as much of a barrier to data sharing as access rights. Data reforms underway (as part of ESB *Initial Reforms* proposals) aim to increase access by allowing AEMO to share data with trusted bodies. However, these reforms will have limited impact without resolving critical gaps in resources and processes needed to manage and gain value from these large and often unwieldy datasets.

Data services needed to unlock insights from data can be represented as stages in a data 'value chain'. These services are summarised in Table 1 and discussed in the section below. These represent the range of services that should support and could be available to accompany any dissemination of data (with services able to be delivered by more than one data holder).

Table 1 *Description of data services*

Data dictionary	User requirements	Custodian-ship	Contracting	Data access	De-identification	Compilation/ analysis	Visualisation/ insights
Publishing and maintaining a simple catalogue of available data (a light touch transparency measure to allow users to identify data)	Engaging actively with data users on their needs – providing early advice to scope and manage risks in proposed on work, as well as understanding collective user requirements.	Working with data holders to curate data to ensure its suitable for sharing and analysis. Could include sharing guidelines and formats for data custodians and "data hosting" services.	Negotiating agreements and data protections for specific data sharing and services, as well as management of published standard terms	Support systems and processes for safe sharing of data with a wide range of users (e.g.: curating data through secure data labs, dashboards and portals).	Ensuring that privacy and sensitivity requirements are met before data is shared. Including development of robust and transparent de-identification processes.	Expertise to develop insights and value from data. Includes linking datasets and a range of computational approaches.	Supporting ways to present and access insights, Including through publication and tools.

Many of these services could only be provided by the data holder, as they have unique access to the data which must often be assessed, de-identified or manipulated to protect it prior to sharing. But currently data holders often have limited capacity or mandate to undertake these activities for the benefit of policymakers, planners, and innovators, or to understand the needs of other potential data users, due to scarce resources and competing roles and responsibilities. Data users and alternative service providers cannot reduce the burden on data holders to enable safe sharing since they cannot have visibility of the data prior to being manipulated. Funding is difficult prior to establishing if the data is appropriate to the need, and funding alone may not resolve the issue if the problem is scarcity of expert resources or lack of supporting processes.

5.1   Publishing information about datasets

An initial hurdle in accessing any data is knowing if and where it exists. Stakeholders said researchers were not able to establish what data was currently held and how they could access it, and that this was limiting the potential for research and value for consumers, as well as the wider ability to resolve data gaps. Current arrangements depend on data holders to voluntarily invest time to proactively release and maintain information on datasets they hold, which may not be an internal priority, particularly where barriers remain to releasing this data in most cases.

The creation and maintenance of a *data dictionary* would provide an accurate and regularly updated list of data elements, including their location, availability, granularity, sensitivity, mechanism, and cost to retrieve. A data dictionary is an important feature of any future data service delivery model. It not only allows data to be located but also is part of ensuring that datasets across data holders remain compatible and coherent and data gaps and priorities can be identified and addressed. It would link to any common standards for these datasets to ensure interoperability.

In the Australian energy sector, this could build on a range of existing data standards and be a relatively simple task initially, focussed largely on ensuring relevant information is all in one place and digestible. The data dictionary would not seek to bring data together, change its ownership or the way it is maintained. This is intended as a light-touch transparency measure as a first step to improving access and understanding for data users.

For example, the data dictionary could initially focus on priority AEMO data and build on existing data exchange requirements (including AEMO's business-to-business exchange procedures) and link to interoperability arrangements and standards under development as part of the CER Implementation Plan and the Consumer Data Right. Wider data sets could be included over time as appropriate, and the body tasked with delivery of Data Services could be required to maintain information and ensure it covered useful datasets relevant to data user requests. This would provide greater transparency for all businesses and market participants into data held by the market bodies, government, and other businesses. As datasets such as the Distributed Energy Resources (DER) registry continue to grow and evolve rapidly, understanding emerging data capabilities will support innovation and a transparent market.

5.2   Data requirements gathering

Data requirements gathering includes assisting a data user with framing a business, research or policy problem and turning it into an analytical approach and a specific data query. Stakeholders were concerned about the limited ability of current data services providers to help users formulate a data query adequate to their policy question. Data users struggle without this support, as gaining insights depends on in-depth understanding of the data, its weaknesses and how it can be interpreted, as well as often advanced data science capabilities to understand opportunities to interrogate large and complex datasets. Many workstreams invest substantial resources into accessing and analysing data, only to discover analytical hurdles which cannot be resolved as the data was not fully understood at inception and is not appropriate to the task.

For example, consultations with a State Government energy policy department revealed that much of its recent engagements with first AEMO and later NEAR for data services involved lengthy negotiations on the ways the two providers were responding to the Department's policy questions. Challenges emerged in the data and methodology which limited the outcomes the department was seeking, yet these challenges could have been predicted, or alternative approaches sought with greater transparency or advice in the early stages when planning the analysis. These studies took many months to negotiate and further time and resources to undertake.

This is a common issue, with many other recent examples seen when policymakers do not have expertise in or visibility of the specific datasets, and data providers do not possess the specific energy sector, economic or social policy knowledge to understand how best to answer the question. Opportunities to answer critical questions are often missed, or significant investment undertaken in approaches which are unlikely to work. Lack of data requirement gathering capability in current service providers has frequently created barriers to efficient policy making and planning.

The Victorian C4NET provides an example of the value added up-front advisory services. C4NET have added value to many research processes by facilitating careful up-front scoping of the problem, reducing risks and barriers, and increasing the value of the resulting insights.

Example 1 Centre for New Energy Technologies (C4NET)

C4NET was initiated by the Victorian Government to promote innovation using Victorian energy data, given the state's unique coverage of smart meter data. C4NET is a membership-based model that engages with the network businesses and a range of research and government bodies to advise on and support data sharing for innovation. It provides advisory consulting services for energy data users and has established trusted relationships with its member distribution businesses, researchers, and policymakers to facilitate data access and data contracting between parties. It also advises on data sharing agreements, using experience gained across many projects to reduce uncertainty and streamline negotiations with common terms.

C4NET has provided a clear example of the value of expert advice in scoping the initial problem in the context of data available, using its role as the intermediary between data holders and data users to ensure policy questions are being framed in the right way to obtain access to the most relevant data. Without this service, data queries are open to misinterpretation or are not meaningful since they do not answer a precise question or consider limitations in the data. This increases risks to outcomes and whether the problem can be solved. In stakeholders' experience, it also creates delays, unnecessary red tape for service providers and complicates the user data journey, particularly when combined with the lack of a centralised data dictionary which describes datasets, their limitations and where they are held. C4NET's role as an intermediary has had strong success in managing these risks and increases the likely value of the final insights.

C4NET also provides a good example of the resources needed in facilitation. C4NET works with the network businesses to release data, often by embedding its own experts into the network businesses to provide the resources and expertise needed within their internal systems to appropriately prepare and release the data.

But C4NET's data access and expertise are largely limited to partnerships with a number of network providers. They are limited in their capacity to resolve many of the wider concerns and provide greater services without enhanced access to datasets, such as those held by AEMO, and resources to scale.

Any new data services model should learn from the C4NET arrangements and experience. Data services should also be set up in a flexible manner to support these kinds of existing experts to increase their impact.

5.3   Data custodianship, contracting for data access and data accessing

The most common barrier raised by stakeholders was the issue of data access and data access arrangements which were very slow and cumbersome and often resulted in an inability to access data at all. Many data requests from policymakers are rejected due to the inability of agencies to share data, even with trusted partners and protected environments.

Other barriers raised by stakeholders covered issues around data custodianship and data contracting. Stakeholders indicated that many data custodians lack sophisticated processes and internal capabilities to support data quality and curation to allow for analysis and sharing. Some data users consulted indicated that negotiating data access and sharing through data contracting can take over two years of legal negotiations, and even then, has often failed to progress. This is often due to a lack of common terms and uncertain rights around data access.

There are a range of access arrangements that are associated with this service type with their respective barriers considered in more detail below.

Data custodianship services includes physically storing, managing, and maintaining data to an agreed quality level (e.g. AEMO stores the full dataset of electricity standing data and meter read data). Custodianship of data that is not held by the entity that did the initial data collection can result in difficulties with managing data quality, interpretation, and scope. For example, meter data held in the Market Settlement and Transfer Solutions (MSATS) has many flags that can be used differently by different meter providers, making them inconsistent and hard to interpret. Data custodians need to be resourced to help users understand the data nuances, communicate them, and manage the risk of misinterpretation.

Often data that has been collected for a specific operational purpose is not designed for sharing or wider analysis and can need significant curation or interpretation to create value. By increasing the ability of data holders' to share data, there is value in providing advice and support to put the data in a form where it can be shared usefully and safely. This can be resource intensive if not a regular part of managing the data set and can become a barrier to ad hoc data sharing.

Data contracting services entail developing a legal agreement between a custodian and data query owner to enable access to data and any subsequent data service. However, uncertainty in regulatory requirements around data access creates risks for the data holder and limits incentives to negotiate. A lack of standardised, common data sharing terms and supporting processes can also contribute to make data contracting arrangements a long, drawn-out process. For example, interview feedback from a state department highlighted lengthy timeframes in developing contracting arrangements with AEMO, then further time negotiating separately with NEAR/CSIRO. Negotiations for the NEAR arrangements also took 2-3 years for CSIRO to gain access to the data sought, and this access was still heavily limited.

These delays were largely due to limitations in the NEL, being addressed as part of the ESB "*Initial Reforms*" proposals, and will be assisted by the development of common guidelines and standard terms for data sharing (planned for early 2023). However, to support increasing the efficiency of these contracting processes, will require expertise familiar with the arrangements to lead the negotiations and understand issues specific to each data user and the scope of the service. Guidelines and standard terms will also need to be maintained over time.

Often there are additional complications regarding constraints and requirements put on data sharing by the data holder to limit risk, such as limiting secondary disclosure even if disclosure occurs with consent and data is de-identified and protected. Sometimes inconsistent regulations mean that the same data from another source can be shared. Consultations revealed that some data users have reverted to using web scraping tools to scrape data off agency and regulator websites where agencies are not able to share the data in other formats. This can also lead to duplicated data gathering costs as data users recollect the same data from consumers.

Data access services require the systems and skills needed for handling big data, including protected systems or data labs, and expert skills to curate and prepare data for release. This requires material resources within the data holder. C4NET often manages this constraint by deploying their own staff to sit within a data custodian's organisation to assist with the preparation of data before it leaves the system. Beyond energy data, the Australian Bureau of Statistics (ABS) uses protected data labs to facilitate data access for approved users, however individual projects still require manual data curation behind the scenes and remains resource intensive.

AEMO also create a similar safe data lab for NEAR, but the level of resourcing required to curate the data provides an example of the value added in deploying staff to sit within a data custodian's organisation to assist with the preparation of data before it leaves the system.

Example 2: National Energy Analytics Research program (NEAR)

NEAR provides a key example of an attempt to meet data service needs, and the many governance challenges it faced.

NEAR is a $20 million Commonwealth-funded joint venture between AEMO, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Commonwealth government (Department of Climate Change, Energy, Environment and Water (DCCEEW)). It was set up to respond to critical gaps in energy data needed to improve demand forecasting, planning and policy, particularly in the energy transition where consumer choices and drivers of demand are changing rapidly. Data sets previously linked energy usage (meter data) to the drivers of energy usage (such as technology uptake, surveys on consumer behaviour, building data or data on commercial activities) to support understanding and analysis of rapidly changing demand

The partnership aimed to resolve this by bringing together analytical services from CSIRO with data available to AEMO (particularly meter data). It was intended the joint-venture approach could resolve data-sharing requirements and provide solutions to restrictions on AEMO in the NEL. However, these constraints remained challenging with negotiation of data access and wider joint-venture arrangement creating major delays in the venture.

NEAR's current access arrangements to AEMO data necessitates the use of a secure environment in which NEAR staff can view and analyse data but AEMO staff need to de-identify and aggregate any data sent beyond the environment. These restrictions can constrain data linkage and a range of potential analysis and also requires AEMO personnel and resources to be engaged for bespoke data requests. To address this, NEAR funded a permanent resource inside AEMO to reduce constraints on AEMO staff who work concurrently on other projects and priorities.

NEAR has been able to achieve a range of value add to AEMO datasets and a large volume of foundational research using a wide range of data sets, including many outside of AEMO such as network data and bespoke surveys. It has demonstrated the potential of many types of data analytics - for example analysis meter data to extract the proportion and trends in heating and cooling demand, which remains a critical driver of demand. However, delays and barriers to data access have limited its ability to address some of the intended services, including facilitating better data access for other data users and stakeholders.

NEAR also demonstrated range of other challenges in governance covering data access concerns with wider parties. For example, another state government recently worked with NEAR to analyse the impacts of some of their programs, to avoid challenges they had faced requesting analysis from AEMO directly. However, difficulties were experienced with data sharing agreements and providing some of their own data to CSIRO in a way that would enable it to be linked. The ABS has also faced difficulties working through NEAR, as data linking still needs to be done within AEMO's system (meaning the ABS could not share their related data for linking). Wider research programs still need to request CSIRO to conduct analysis for them, which frequently does not meet their needs.

NEAR also highlighted some structural concerns in data services. As a joint partnership supported by a limited grant, priorities and workflow were negotiated by the parties into annual workplans to be implemented by CSIRO. There was a strong focus on demonstrating the potential benefits in energy data research, particularly to the joint partners to support ongoing funding. This aligned well with longer-term research priorities but lacked flexibility and responsiveness to support more short-term queries. It also drove a strong focus on the partners needs, with limited development of services for wider stakeholders or growing underlying data infrastructure.

NEAR has demonstrated that energy data research has huge potential to create benefits in the market but that collaboration approaches alone cannot resolve key barriers. More fundamental energy data reforms and access arrangements are needed to support a wider range of stakeholders.

Stakeholders cited the highest number of barriers in this data service category.

Lack of clarity in legal requirements and lack of resourcing and incentives for data custodians to share data have resulted in limited risk appetite for ongoing data contracting and data linking arrangements. This, in turn, contributed to a proliferation of inefficient, repetitive, and duplicative data requests with long processing times and burdensome administrative arrangements.

This includes issues such as constraints on the use of data, such as secondary disclosure. Frequently data access agreements (including from AEMO) limit secondary disclosure, limiting provision of value-added data services. This is a key issue to resolve if data services are to be provided outside of AEMO.

Stringent privacy and confidentiality rules, as well as broader organisational concerns relating to risks associated with data protection, often complicate access to raw data and data contracting. This is particularly true for market bodies like AEMO that are statutorily obliged to preserve confidentiality of personally identifiable data. Recent highly public cybersecurity incidents are likely to increase caution on this issue.

Stakeholders were particularly concerned with the limited provision of data access and contracting services in relation to datasets they identified as priority.

A key benefit from the development of Data Services will flow from increasing expertise in managing these identified data sharing risks effectively and proactively, reducing the tendency for legal processes to be unnecessarily cautious or blunt. This will require management of data holder risks (being considered in *Initial Reforms*) but would involve a shift in the culture and approach to data sharing. It would likely be appropriate for the body tasked with developing Data services to have objectives in both building this expertise and help to facilitate this cultural change.

5.4   Data de-identification

Many priority datasets held by market bodies include protected data of individual residences or businesses, such as MSATS metering data and the DER register. These data sets are stored linked to a physical location and unique National Meter Identifier (NMI). Energy use data patterns are also complex and unique, like a fingerprint, and so need advanced methods to de-identify them without losing important characteristics which may be needed for analysis.

Importantly, however, market bodies do not store information on individual account holders (i.e. name and phone numbers etc), only data linked to a location. Most meters support multiple individual residents or are related to business sites. Therefore, meter data is not usually *personally* identifiable and not captured by privacy laws, except when linked to billing data held by retailer service providers.

However, while not technically 'private' these datasets can reveal information about the customers at the meter, as usage patterns can be linked to information about their residence and equipment when they are regularly home and likely scale of their bill. Related metering data can reveal which commercial provider they are with and when they last changed. Sensitivity around these datasets makes them protected legally under the NEL/NGL (even if not under privacy law) as well as a key source of social licence concerns and risk-averse decision making. There is always concern that they could be 're-identified' by linking to address information. Complexity of energy usage patterns also makes them unique, like a fingerprint, and also useful for re-identification. These concerns are particularly relevant in the context of heightened public interest surrounding data access and protections in relation to recent large scale data leaks (e.g. Optus, Medibank).

Many research purposes do not need to identify individual meters beyond a general location, but they do often require disaggregated meter-by-meter data to analyse differences between users and understand potential drivers of energy use. This can require meter data to be linked to other data to analyse the impact of different drivers of energy use – such as whether the site has particular alternative energy sources like solar or gas, was involved in a particular program or intervention, has particular equipment or commercial activities, or is in a particular climate location or household demographic etc.

There are many ways to analyse and share this kind of data while maintaining its protection. Data de-identification services can enable personally identifiable or confidential data to be shared safely – through methods such as obfuscation, aggregation, or randomisation. Since many research purposes only require de-identified data, these methods could open up new avenues for research with many wider bodies.

However, these methods are difficult to apply in complex unique data sets such as metering information, often destroying the value of the data researchers are seeking to analyse or being able to be reversed or 're-identified', e.g. this is particularly the case where research is being undertaken into better understanding drivers of consumer behaviour. De-identification services therefore require significant expertise and agreed and tested data science methods to help assuage stakeholders and reduce risk. These methods remain on the forefront of data science and are ever changing as re-identification methods also improve. AEMO and the wider energy sector currently lack common agreed methods to use with confidence, or the data science resources to invest in developing them and keeping them up to date.

Provision of this service in the current data sharing model has been identified by stakeholders as a major barrier to data contracting and accessing.

Additionally, stakeholders highlighted their experience of dealing with a deeply ingrained culture of strict data protection, whereby data is only released in extreme circumstances, as opposed to a culture where risks are appropriately triaged and minimised to an acceptable standard. Data custodians also outlined concerns regarding re-identification of previously de-identified, sensitive data. De-identification services therefore require agreed and tested data science methods to help assuage stakeholders and reduce risk.

Stakeholders in the energy data ecosystem have tried to fill in this niche. NEAR (at AEMO's request) is currently working on an advanced de-identification methodology to improve the anonymisation of metering data and increase confidence in sharing with other agencies. But these capabilities require maintaining ongoing advanced expertise in this space, which is arguably beyond AEMO's current mandate and cost-recovery ability. Other bodies, such as C4NET, signed agreements with relevant data holders that embed C4NET analysts in their daily operations, assisting with data de-sensitisation services. Neither of these two bodies, however, provide data custodianship services: they do not collect or store any data themselves but rather rely on data obtained from third parties, and so are limited in their capacity to resolve these issues.

5.5   Data linking, analysis and presentation

Data analysis services involve cleaning and organising data, linking, or combining it with other datasets, performing analysis on it (such as regression and trend analyses, statistical analyses, segmentation and clustering, machine learning for patterns, etc), and combining it into a form useful for answering a specific research or policy question. These services can range from simply querying data to complex analyses, which require superior computing power. This is where data becomes valuable – in its ability to provide insights.

Most datasets however cover only part of a problem, with integration across a range of datasets usually key to expanding benefits. For example, linking electricity metering data to some of the attributes of the energy users (e.g. their location, equipment, building characteristics, gas use, a survey of the consumer demographics and decisions, or their commercial contracts and price incentives) can provide insights into what drives demand and consumer impacts. But linking or integrating protected data remains challenging, as attributes of the data needed to link or integrate it generally are the same attributes which make it identifiable and personal, making them difficult to share. This can mean that linking and integrating data sets is limited to the original data holder prior to data sharing, with data de-identified once linked.

Similarly, aggregation is often undertaken by the data holder as a pre-requisite to sharing as part of de-identification. But aggregation without a focus on the analytical problem being solved often destroys the value of the data. For example, smart meter profiles aggregated by a local region immediately destroys the critical diversity in neighbouring users, aggregating consumers with and without CER, with and without gas, with different home efficiencies, different primary appliances, and different demographics.

With the increasing scale and complexity of datasets, interpreting and interrogating this data often requires advanced systems and skills, which are scarce in many data users. This is leading to a growing need for analytical services such as linking data sets together or common methods of analysis such as creating control groups to measure impact. Current service providers, such as C4NET and NEAR, are limited and resource constrained. Wider research entities with the data analytic skills currently lack access to the data in granular form.[6] AEMO provides some data compilation, combination, aggregation, and analysis services, yet this service offering is limited by funding models. AEMO offer these services on a cost recovery basis based on ad hoc requests. Due to the ad hoc nature of requests, AEMO uses existing resources to maintain cost efficiency. This approach limits the ability to meet requests due to resource contention and competing priorities, with limited investment in growing the availability of these skills.

For example, AEMO has recently delivered analytics services for many jurisdictions, including NSW, SA, Vic and the Commonwealth. In all cases, work was materially delayed and scheduled around availability of relevantly skilled resources in AEMO, many of whom are involved in delivery of core AEMO functions such as regular forecasting requirements and can be unavailable for many months of the year.

Data presentation services assist data users with presenting data in a format required for the end state, such as graphical visualisation and compilation into dashboards or tables. This is a key element in making data widely accessible and useful, for example, publishing aggregated metrics for a wide range of market users. Stakeholders highlighted the limited and ad hoc character of current data presentation and visualisation services, which places the burden to provide these services on the energy data users. NEAR and C4NET are the main bodies providing data presentation, visualisation, and dashboard creation services to their clients, yet these services are limited and delayed by processes required to overcome data access arrangements with energy data holders. The processes required to overcome data access arrangement is confounded by the limited resourcing that is available to support access to commonly needed datasets.

AEMO provides a wide range of reporting and data publication, but is primarily focused on regulated or market requirements, rather than policy and research audiences, leaving significant gaps to those stakeholders. AEMO also provides presentation services along with analytics for a fee to state governments on an ad hoc or ongoing basis. For example, AEMO has an agreement in place to deliver regular tailored reports for SA government, with delivery often timed around resource availability. AEMO has recently advanced this regular reporting to a useful, self-serve dashboard arrangement. The SA government's needs and requests for this report have also recently expanded to support greater analysis of local grid behaviour under growing solar penetration, and like other jurisdictions, these growing needs are likely to continue to expand beyond the scope of current services.

Another key source of energy data presentation services is the Australian Renewable Energy Mapping Initiative (AREMI), which was implemented on National Maps and funded by ARENA, supporting locational renewable energy planning through aggregated layers of GIS data and energy sector data. This was a unique and highly useful data source to many. However, when it moved from development to operational state AREMI there was no body with an ongoing mandate to fund and maintain this function.

---

[6] Note that changes proposed via the *Initial Reforms* are seeking to support further access for these parties.

5.6   Interaction with Initial Reforms

These gaps to data services also interact with regulatory challenges identified in the Data Strategy. *Initial Reforms*[7] are being developed to address these regulatory barriers but will depend on and be complementary to Data Services.

The main regulatory barriers identified are around risks for data custodians and data contracting. The overall objective of the *Initial Reforms* is to permit greater access to, and use of, data held by AEMO in a secure way. It is envisaged that a small number of modest changes to the NEL, the NGL and the Regulations will improve the ability of AEMO to securely and confidently share useful data with trusted bodies. Specifically, AEMO will be permitted to disclose protected information to what are described within the *Initial Reforms* proposed as "Class A bodies" and "Class B bodies" and that other complementary changes be made such as imposing conditions on data access and attaching civil penalties to breaches of those conditions.

It is envisaged that the initial legislative reforms will be accompanied by non-legislative controls. It is expected that AEMO may:

- publish standard terms and conditions that apply to each Class B Body that receives protected information from AEMO
- maintain a register of data that it has shared to Class B bodies (and potentially Class A bodies). The register is likely to include a publicly accessible part and a part that is not publicly accessible
- may make written guidelines to support the interim data reforms with principles and processes relating to, for example, data release, data management and curation, technical matters and standards and emerging technologies.

5.7   Summary of Data Services required to address barriers identified

Overall, data users are faced with a fragmented data service delivery model where data custodians lack analytical resources, and organisations that could provide de-identification, analysis, and other services often lack access to priority datasets. Any future data delivery model should aim to address these problems by facilitating a smooth, end-to-end data services offering, closing existing gaps in services to allows data holders and users to satisfy their data needs in the most efficient way possible.

Table 2 below provides a summary of the data service needs identified and shows how these services interact with *Initial Reforms* to address the relevant barriers.

Consultation questions:

Q 1: Are there any priority data services missing from the analysis?

Q 2: Are there other barriers that inhibit data services not identified here? Can you provide examples or case studies of these barriers or are there examples of where data services are addressing any key barriers?

---

[7] Energy Security Board (2022), Data Strategy – Initial reforms consultation paper

Table 2 Summary of Data Services required

A lifecycle of data services: from finding to accessing to analysing to sharing data

Data Services	Data dictionary	Data requirements gathering	Data custodian	Data Contracting	Data Accessing	De-identification	Compilation/ analysis	Visualisation/ insights
Description of services required	Publishing and maintaining a simple catalogue of available data (a light touch transparency measure to allow users to identify data)	Engaging actively with data users on their needs – providing early advice to scope and manage risks in proposed on work, as well as understanding collective user requirements.	Working with data holders to curate data to ensure its suitable for sharing and analysis. Could include sharing guidelines and formats for data custodians and "data hosting" services.	Negotiating agreements and data protections for specific data sharing and services as well as management of published standard terms	Support systems and processes for safe sharing of data with a wide range of users (e.g.: curating data through secure data labs, dashboards and portals).	Ensuring that privacy and sensitivity requirements are met before data is shared. Including development of robust and transparent de-identification processes.	Expertise to develop insights and value from data. Includes linking datasets and a range of computational approaches.	Supporting ways to present and access insights, Including through publication and tools. .
Initial reforms (work together with data services to address barriers)			**Empower AEMO as custodian to share data with trusted bodies.**	**Protection requirements Common data sharing agreements Register of agreements.**				
Current barriers and gaps identified	Limited visibility of existing data and few transparency requirements – data users often don't know what data exists and have few ways to find out.	Data users frequently don't have sufficient information / context to determine if or how data available could answer their questions.	Data is generally collected for internal purposes (operations, services). Many data custodians lack processes/resources to manage additional data quality and curation needs for sharing and analysis.	Lack of common terms and uncertain rights means data sharing agreements often takes years and significant resources to negotiate.	Many data holders and users lack systems and internal capabilities to share data safely – particularly "big data".	Lack of agreed methods for de-identification of many core data sets (e.g. meter data) makes safe sharing often impossible/ uncertain.	Limited expertise in linking or common methods (such as control groups to measure impact). Linking of existing datasets could add major value.	Limited resourcing to support open access to commonly needed datasets (e.g. portals for policy makers).

6 Data Services delivery models – proposed scope

Earlier in this paper we have discussed practical barriers which limit current data sharing and have set out a series of data services needs. These are set out in Table 2.

In considering possible governance models to deliver these services, it is important to clarify the scope of these delivery models, and data services, data sets and parties expected to be covered – and not covered. This section provides discussion of the approach intended for development of data services capabilities from any of the proposed models (over time). While development of relevant data services will be based on identified priorities and demand for such services and informed by assessment of the relative costs and benefits associated with delivery; the discussion below is intended to provide greater clarity of what features will and will not likely be progressed as part of any future model.

The ESB recognises that the full range of potential benefits will not be unlocked from the outset, and there will need to be effective prioritisation of services developed as part of building capabilities and processes with initial implementation of any new delivery model.

Features to be included in development of any delivery model:

1. Data services needs identified which address gaps to facilitate a chain of access and benefits, including: Data catalogue; requirements gathering; custodian; contracting; access; de-identification; compilation and analysis; visualisation and insights.

2. Each delivery model should seek to provide the expertise and resourcing to best facilitate the full range of data service needs identified, so as not to leave gaps and barriers. Services may be supported across a range of providers in any given delivery model, but all aspects need to be covered to resolve gaps in capabilities.

3. Data services should be a genuine service model with clear targeted priorities informed by data users (e.g. policy makers, jurisdictions), and able to respond and adapt to service requests from data users over time. A phased approach, starting small and focused on current data and current data needs (e.g. leveraging access to AEMO data and facilitating *Initial Reforms*) will support keeping these needs focussed up front. Service development can grow over time to include wider coverage as new priorities and needs emerge, informed by demand, potentially evolving between different models as they grow. Fee for service models to ensure costs incurred are apportioned appropriately by data users will need to be developed to support effective service delivery.

4. Delivery models should be adaptive to consider and respond to how changing data user needs may be identified, prioritised, resourced and delivered over time. Mechanisms will be needed to consider collective data user needs as well as individual data user requests. Growing expertise and capacity in data users is also important, so mechanisms to engage data users in priorities should consider opportunities for ongoing engagement with experts.

5. Delivery models should consider services to different classes of data users who could support consumer benefits, and different mechanisms and resources needed to meet their needs. Examples include:

   a) Bodies who will have expanded trusted access under the ESB "*Initial reforms*" proposals, including relevant government bodies and Australian university research bodies, and will need this new access to be facilitated.

   b) Bodies who may have some access to relevant data but may lack aspects of analytic capabilities needed to gain insights, particularly from increasingly complex "big data" sets.

c) Bodies who may not have the rights to access or internal capacity to protect data but may benefit from data transformed into more useful protected outputs: e.g. regional aggregates for local government or community groups, or portals supporting aggregated insights supporting improved market services.

d) Bodies who may be data holders but lack the internal capacity to curate data for analysis or facilitate safe access and may seek support for hosting capabilities.

6. Development of data services are limited to facilitating benefits from existing data sets voluntarily provided, including:

a) Data already held by bodies with a public-good mandate (such as market and government bodies), which could support greater consumer benefits if shared safely.

b) Data already available in some form where consumer value could be increased with analytical services (such as linking or more useful aggregations).

7. Data sets held by and available to these bodies are not fixed and will likely continue to change rapidly with data user needs, innovation and reforms in the market. Delivery models will need to scale and adapt overtime, or even evolve into alternative delivery models, with capabilities transferred to different parties.

8. Delivery models should consider opportunities of different parties in facilitating advice and access:

a) Some services may need to be at least partially facilitated by a data holder, such as aspects of data curation, negotiation of access agreements, and activities required to protect the data before release, such as de-identification or identified-linking.

b) Some services could be provided externally by expert advisors, such as: advice on data requirements and analytical methods, management of data dictionary, standards or registry, and advice on common sharing agreements and compliance.

c) There are specific services, such as third-party protected matching, where support from an independent party may be necessary.

d) Data exchange systems and access services could also be facilitated remotely from data holders. For example: portal services facilitating access to data which is not held centrally, such as national maps and AREMI.

Issues out of scope for development of any delivery model:

1. Issues of data rights and governance – i.e. changes in who owns, manages or has access rights to different data sets.

a) Where data sets are not currently accessible to government or market bodies, any change would require an explicit benefit case and policy mandate. Development of models for delivery of data services are not intended to address accessibility of such data sets.

b) The policy mandate for these data sets may be considered by alternative reform workstreams and, if made available, could be supported by data services in the future.

2. Data sets not currently accessible or voluntarily provided to market bodies and governments are not within current scope of data services, even if identified as a data gap within the Data Strategy, as issues of data rights and governance are out-of-scope. Excluded data sets are set out in Table 3 below, but in summary contain:

a) Data held by commercial bodies (such as market participants and service providers), which may provide consumer benefits if available in some protected form to the market or regulators, but there are constraints or limited incentives to release. A benefits case and policy mandate would be needed.

b) Data not currently captured or managed by any party, requiring primary collection or monitoring investment. A benefits case and investment decision would be needed.

c) Data that may not be in the interests of consumers to share (even while remaining protected), such as sensitive data held by market or regulatory bodies which could impact market or security outcomes.

3. Data services which seek to compete with or duplicate existing capabilities which may be better promoted in the market, such as:

a) ***Data science research capabilities*** which exist widely across universities and research bodies and can promote in-depth research into new methods, approaches and models. Data services should be designed to enable and grow, rather than compete with, these capabilities, by facilitating access for or collaborating with these bodies and promoting sharing of intellectual property. Any analytics provided by a data services delivery body should be focused on more immediate needs: e.g. protection, linking or aggregation of data sets, and responsive provision of short-term, targeted insights like trends/metrics and reporting, or direct user queries. There may be also sensible roles for data services to operationalise or collaborate with new research capabilities developed by others, where access and data protection remains an issue. For example, operationalising new advanced methods in de-identification, linking or aggregation, or supporting protected data feeds for new forecasting models or visualisation tools.

b) ***Commercial data services*** which add insights, tools or value to data that may already be available and is not limited in the market (and where monopoly access to data does not risk limits to consumer benefits)

Data gaps being addressed via Data Services

As discussed above, the Data Strategy identified a range of gaps and constraints, many of which need different approaches for resolution. Table 3 below provides discussion on this range of data gaps, highlighting where Data Services are targeted, noting this is a sub-set of relevant data sets.

Table 3: Scope of different classes of data gaps within the Data Strategy, indicating the role of Data Services

Data gap	Examples of data (non-exhaustive)	How are these datasets considered by Data Services and the Data Strategy?
1. Data already held by AEMO. which could support greater consumer benefit if shared safely, but constraints on sharing remain. *(Primary focus of this consultation paper)*	• Data held by AEMO (protected data), such as meter and DER data	**Included in Data Services** • **Initial Reforms and Guidelines** - reduces barriers to safe sharing with trusted parties. Sharing is allowed but not mandated.
2. Data already held by bodies with a public-good mandate (beyond AEMO), which could support greater consumer benefit if shared safely, but constraints on sharing remain. *All sharing would be voluntary on the basis of consumer benefits*	• Data held by market bodies – some types of AER, AEMC or ECA data, such as survey outcomes. • Jurisdictional data from programs or reporting. • Data from publicly funded research with obligations to share, such as ARENA. • Data held by other government bodies, such as local planning bodies, transport or infrastructure, or central bodies like the ABS.	**Included in Data Services** • **Data services** – facilitates and improve safe sharing arrangements and promote value from existing data held by public bodies and voluntary providers.
3. Sensitive data held by market or regulatory bodies which could impact market or security outcomes, or data that should not be shared in the interest of consumers.	• Confidential data provided to AEMO for forecasting or security management, with potential to impact markets. • Data gathered through ACCC compliance powers.	**Excluded from data services** • Data holder retains discretion to identify data as 'too sensitive to share' – where it is not in the interest of consumers, the market or security.

		• AER data acquired for targeted compliance/monitoring of a market sensitive nature.	• Data holder to consider any data aggregations or analysis appropriate and beneficial to release. • Consider alternative approaches for related data needs or insights.
4.	Data held by commercial bodies, which may provide consumer benefits if available in some protected form to the market or regulators – but there are constraints or limited incentives to release.	• Local network performance and planning data. • DER performance data. • Billing/retail data. • Market contracts. • Gas usage and billing. • Individual data – bills/tariffs/usage. • Some aspects of meter data, such as voltage. • Equipment uptake or installation data, such as appliances or EV chargers.	**Excluded from data services** unless/until there is a future policy decision to include them. • Access and obligations around these data sets need explicitly policy consideration and cost benefit analysis. This is not the role of the data services workstream. • Some of these data sets are being considered through **ongoing targeted reforms** – including Consumer Data Right, AEMC's Metering Review, ESB CER Implementation Plan and Commonwealth revisions to AER monitoring. • Some of these data sets have been identified to be considered as Data strategy **Priority projects*** Including: **Bill transparency, Network Visibility, Over Voltage**
5.	Data not currently captured or managed by any party, requiring primary collection or monitoring investment.	• ~60% of homes without smart meters. • Majority of LV network performance which is unmonitored. • DER/EV without smart metering, monitoring or communications. • Data which could be gathered by installers – such settings or equipment characteristics from EV chargers/DER. • Data on energy users requiring primary survey – such as views, motivations, past behaviours, existing equipment.	**Excluded from data services** – unless/until there is a future policy decision to include them. • These datasets do not yet exist and would require investment based on explicitly policy consideration and benefits analysis. This is not the role of the data services workstream. • Some may be considered by **ongoing targeted reforms** – including AEMC's Metering Review and ESB CER Implementation Plan. • Some may be considered as investment decisions by commercial parties, or where supported by AER or government programs such as ARENA. • Some have been identified by the Data Strategy for policy consideration as **Priority projects*** Includes – **EV Charging standing data, Network Visibility, Over Voltage and Consumer metrics**
6.	Data already available in some form but consumer value could be increased with investment or linking.	• Commercially available datasets, such as DER, buildings or industry, satellite data, weather data, banking data, etc.	• **Data Services** - Could be included in value-added data services. Would require funding due to a specific data user need/outcome.

*Priority project options or ongoing targeted reforms, considered in consultation with stakeholders, could result in data moving to one or more of the other categories - including options for support from Data Services, e.g. through facilitating shared industry dashboards, portals or aggregates for data held by a range of parties. However, these issues are not intended to be resolved via development of any Data Services delivery models.

Consultation questions:

Q 3: The ESB welcomes feedback on the features proposed for data services delivery models. Are there other considerations that should also be taken into account?

Q 4: What are stakeholder views on the appropriate scope for data services in the short to medium term?

7 New data services delivery models

This section discusses options for potential governance models that could support delivery of data service needs, consistent with the proposed features set out in Chapter 5.

Five potential options for future data service delivery have been identified through stakeholder discussions to date, building on a range of existing examples and needs. These have been described and assessed at a high level for wider stakeholder consideration.

7.1 Potential future data service delivery models

This section sets out possible governance arrangements and discusses each model's high-level strengths and weaknesses.

Table 4: *The five models proposed in initial consultations*

Model	Approach	Governance
Resourced AEMO	Funding and resources for AEMO to provide additional data services.	Existing AEMO internal governance arrangements – remaining at AEMO's discretion to respond to stakeholder requests.
Dedicated unit within AEMO	A discrete data services unit within AEMO – set up with dedicated funding, staff/skills, and steering committee of data users and stakeholders to set priorities.	Steering committee comprising a range of priority data users and wider data holders, including: energy policy and planning bodies (such as AEMC, AER, ESB, DCCEEW, state governments); representative expert researchers; relevant industry service providers and data holders.
Re-fund and re-shape NEAR	NEAR receives renewed funding and takes on a reshaped role and mission.	A joint partnership arrangement bringing together priority data users and data holders (including new partners beyond the current partnership group of AEMO, CSIRO, DCCEEW).
New independent entity	A new independent entity to facilitate data access and provide consulting-style data services.	New incorporated entity with board comprising priority stakeholders (including AEMO, AEMC, AER, ESB, DCCEEW, state governments).
Data services network	A coordinated network of capability built across key stakeholders.	Network across stakeholders with resources required in each body (incl. AEMO, AER, AEMC, state governments, DCCEEW) supported by new data curation authority. Participation would be driven by specific requirements or obligation.

Some of the advantages and disadvantages of each model option are summarised in Table 5.

Table 5: *Advantages and disadvantages of potential data service delivery models*

Model	Benefits	Challenges
Resourced AEMO *Funding and resources for AEMO to provide additional data services.*	• Direct access to AEMO-held data. • Builds on significant existing IT infrastructure and AEMO skills (including NEAR capabilities as a NEAR partner).	• Critical security and market operation roles means existing functions drive strong internal priorities – which compete with resources to service external stakeholder needs. • Competing priorities can make it difficult for some stakeholders to influence and

Model	Benefits	Challenges
		provide input into data service decision-making.
		• Limited drivers to innovate outside of internal needs.
		• Limited drivers to facilitate access to datasets not held or required by AEMO.
Dedicated unit within AEMO _A new dedicated unit: clearly separated within AEMO, with dedicated funding; dedicated internal leadership, staff and skills; and stakeholder steering committee to set priorities._	• Direct access to AEMO-held data. • Builds on existing IT infrastructure and AEMO skills (including NEAR capabilities as a NEAR partner). • Involves stakeholders, including data users, as part of prioritising resources. • Dedicated leadership/responsibility allocated to develop capability and report to steering committee. • Dedicated staff resource to help build out data service capabilities (ideally non-AEMO insourced data experts to assist with new skills and a cultural shift in managing data).	• Limited drivers to facilitate access to datasets not held or required by AEMO. • Priorities still partially influenced by internal AEMO needs.
Re-fund and re-shape NEAR _NEAR receives renewed funding and takes on a reshaped role and mission._	• Can build on lessons learned, capabilities and assets developed in NEAR's first iteration, including projects linking meter data, network data, building and appliance data and EV data, and its data deidentification capabilities. • Access to CSIRO technical expertise. • Existing data access facilities with AEMO.	• Current NEAR focus on annual workplans and joint venture partners drives in-depth research but creates limited flexibility and responsiveness. • Not good at servicing wider stakeholders, operational data flows, discrete data requests and managing negotiations of data agreements. • Difficulties gaining access to raw data may persist – and constraints on secondary data sharing remain which would prevent many activities. • Depends on data custodians to ingest data host limited data. • Priorities set by joint venture funding partners (DCCEEW, AEMO, CSIRO). Would need to expand to wider joint venture partners (which could be complex), otherwise wider data users may have concerns about priorities and access. • Joint venture arrangements may remain cumbersome even if expanded.
New independent entity _A new independent entity to facilitate data access and provide consulting-style data services._	• Could be designed with a clear leadership and greater focus on data outcomes and consumer benefits, with new mission to maximise capability building and innovation. • Focus data skills and best placed to develop and manage data infrastructure (e.g. data dictionary). • Could provide wider sophisticated hosting, linking and coordination services for a range of datasets, including those held across wider policy and market bodies and industry.	• Requires significant costs and resources to set up. • Lacks ability to draw on AEMO's systems create greater up-front costs and delays. • Is not a primary data holder (starts with no data); could require lengthy negotiations to access data. • Would need to be designed to collaborate with and support wider data service providers rather than compete, to ensure wider growth in expertise.

Model	Benefits	Challenges
	• Less constrained by competing priorities and historic cautious culture over data sharing than existing organisations. • New role may have synergies or be able to support other emerging roles – such as enforcement of data agreements or new standards, and roles related to emerging operational data needs such as cyber security or interoperability.	(For example, supporting access for research bodies and consultants).
Data service network *A coordinated network of capability built across key stakeholders*	• Greater expertise will be built across the sector – with new data teams. required in most data holders and data users. • Appropriate if seeking to share the burden across a wider group of data holders. • Greater market for independent data services and innovation. • Likely future model to be employed in the UK (see case study below).	• A new coordinating authority function would still be required – to manage data dictionaries, standards and access requirements. • New systems and skills needed in many organisations – duplication of some costs. • More complex data access arrangements/constraints time/costs in setting up access. • High compliance costs - limited incentive for many data holders to participate so new obligations, standards and enforcement would be required. • Many new governance layers involved. • Data users would need to attract, build, and maintain internal capabilities, which may be difficult for many policy agencies seeking data science experts. • Small, dispersed data science teams may have less capacity for building internal learning and innovation in the short term than a larger skills group.

Example 3: What could the AEMO dedicated unit look like?

To provide clarity on the difference between "Resourced AEMO" and a "Dedicated Unit" model, this case study outlines what a "dedicated unit" might look like. Importantly this is not the same as the recently established "AEMO Services", which from a data-regulation perspective is actually more like the "New independent entity" as it has been established with strong separation of data arrangements.

Priority data services: Consultation to date suggests initial priority services may include:
- A public facing portal providing: data dictionary, relevant dashboards (e.g. Interval Metering and DER dashboard), ability to self-serve curated public datasets (includes AEMO and non-AEMO data sets), and access to standard terms and conditions of services.
- A protected service for government and research bodies (Class A and B) to access information on-demand in accordance with their access agreements. This could be subscription-based or government funded.
- Dedicated team to support the services and work with stakeholders to deliver on priorities for the service. This team should include new expertise brought into AEMO, to assist with a new service focus and cultural change in managing data risks and avoid diminishing existing expertise needed to support core functions across AEMO.
- Resources to support fee-for-service data analysis/linking or access to data labs and broader advice on energy data.

- Capacity to curate the data and develop a range of standardised analytical methodologies – such as different de-identification and aggregation approaches.

Note that priorities and content of the service will be determined through the Governance model and would evolve over time in response to needs.

Governance: The intent of the new Data Services is to support policy, research, and the broader industry through the energy transition. The governance on priorities and investment should align with the service clients. Within AEMO the service would report directly into AEMO executive leadership team, however, the mandate for AEMO would be that the service is delivered to externally agreed priorities within the funding envelope. Models for establishing the priorities could be a combination of:
- Standing committee with a range of wider data users and data holders - across government, industry and research party representation.
- Formal regulatory governance body (Reliability Panel, IEC etc).
- Public consultation.

Given that the nature and initial extent of the service is relatively modest, the suggestion is that a combination of public consultation with an overarching standing committee would be a good balance between broad opportunity and the advantage of a standing group of interested parties to provide ongoing input and engagement.

Funding: Based on the scope of services the funding could be a combination of market fees (for aspects of services supporting market benefits), subscription funding or direct contributions from major data users (such as governments) and fee-for-service funding for specific work. The market fee would be an independent market fee set in partnership with the industry Financial Consultation Committee that currently works with AEMO in setting industry fees. The independent fee structure supports clear financial reporting processes.

The market fee and subscription-based funding envelope will determine the core resourcing and supporting digital assets. Fee-for-service work would be done on a cost recovery basis.

Resourcing and Digital Assets: The intent would be to start with a team of 3-5 core members covering leadership, energy expertise and data analysis and management capabilities. Set-up projects such as an initial data dictionary and base subscription service could be delivered with vendor support. Larger fee-for-service workloads could be met through a vendor partnership. The initial set of data sharing agreements will be delivered with the ESB data strategy project on initial data reforms. The digital assets supporting the service would leverage AEMO's existing public website, portals, and data platforms to avoid duplication of larger AEMO source datasets.

Consultation questions:

Q 5: Are there other data service delivery models that could be considered?

Q 6: Are there better governance models for the AEMO dedicated unit proposal, outlined in the example?

Q 7: Are there other benefits, challenges and implementation issues that should also be considered? Are there any cost considerations that haven't been explored in this paper?

8 Assessment criteria

To support an initial assessment of the five identified data service delivery models, the ESB proposes two qualitative criteria: *impact* and *feasibility*.

8.1 Impact: how might each model respond to current and future challenges and stay innovative and impactful over time?

- **Set-up time and governance structure:** Stakeholders have expressed a preference for a model that delivers a tangible short-term result and has a clear mandate to support data access, has measurable performance metrics, and is overseen by a body representing stakeholders.
- **Ability to address current barriers:** This criterion focuses on a model's ability to address barriers to access/contract for centrally and non-centrally held datasets identified as priority as well as provide other data services that stakeholders typically refer to as instances of barriers, notably de-identification, understanding data location and availability and data requirements gathering services.
- **Compatibility with broader policies and ability to be 'future proof'**: This includes factors such as scalability (a model's ability to add new users, datasets and data services over time) and innovation (a model's ability to foster innovation and adopt new approaches). This criterion also captures to what extent a model is compatible with the direction towards transparency and open data principles and the New Data Framework proposed in the ESB's Data Strategy.

8.2 Feasibility: what level of resources is required to set up and operationalise the model?

- **Cost and financing model:** This criterion measures the approximate total direct costs of setting up a model (including required technology, staffing, and contracting), ongoing funding needs estimate, as well as the availability of funding sources and feasibility of ongoing cost recovery.
- **Incremental regulatory burden:** This criterion assesses the extent of regulatory change required, e.g. data sharing mandates, any governance- and accountability-related regulations, and any other regulations necessary for the model set up. This criterion also captures whether a model increases or decreases the administrative burden on data users.

Consultation questions:

Q 8: The ESB welcomes stakeholder views on whether the proposed criteria are appropriate to assess alternative data service delivery models?

9 Preliminary assessment of delivery models

This section sets out an initial evaluation of the five potential models against the proposed criteria discussed in Chapter 8.

These models have been assessed qualitatively in regard to the status quo and to each other. While consideration has been given to the set-up time and feasibility to deliver data services in the short-medium term, the approach taken has been to assess each of the proposed models in respect of their potential to deliver outcomes and broader benefits over time.

The ESB is seeking stakeholder views on merits of each of these models, and its assessment of how each will achieve proposed outcomes over time.

9.1 Status quo

Future models have been compared as they might transition from the current state of energy data services, as discussed in Chapter 5.

This "status quo" includes:

- A range of data services already provided by AEMO for policy and research needs, but constrained by a lack of clear mandate or funding, regulatory restrictions and resource constraints for internal expertise. This means services are currently modest, can be prioritised on areas of relevance to AEMO's functions, and can be limited or delayed by resourcing constraints. Risks remain high for AEMO as a data holder, so data sharing conditions also remain cautious.

- Additional data services support provided by parties such as NEAR and C4NET – both of which contribute valuable data research skills and expertise, but service delivery remains constrained by access to data and funding.

- Limited access to data for public policy research through industry partnerships and trials – Governments provide significant support to fund trials, particularly focussed at the present time on research or trials to support better understand the needs of the energy transition. Access through these partnerships can also cause major delays and constraints, with inconsistent advice on data sharing allowed, limited confidence in de-identification, and constraints on secondary use. Some of this research is seeking data which could otherwise be more efficiently provided through AEMO, if AEMO were less constrained. Many publicly funded research projects have obligations to share data but lack of standard consents, data formats and data management arrangements limit the usefulness of this data.

- Data users can invest significant effort and costs in seeking access to data and negotiating for data services, often resulting in limited outcomes and delays due to constrained services available. Data users also frequently duplicate effort, investing in capturing data that is duplicated elsewhere. This means opportunities to improve policy planning and consumer outcomes can be limited due to lack visibility of data available.

9.2 Set up time and governance

Resourced AEMO

Based on consultations with AEMO, increasing resources available to the current data service capabilities within AEMO could be relatively quick and take up to 6 months to implement. As a central data holder, AEMO currently provides some data services. However, to date, these have been limited as few data users have been allowed able to obtain raw, NMI-level data from AEMO due to regulatory constraints and privacy concerns, despite the high interest and research and policy benefits of that data. The function could leverage

AEMO's current IT infrastructure and existing AEMO capabilities and dataset knowledge - and could initially utilise internal seconded staff with skills on a temporary assignment basis.

Dedicated unit within AEMO

Similar to a "resourced AEMO", a new dedicated energy data services function would take up a relatively short period of time, approximately six months to set up, but possibly a little longer than the first model. The function could also leverage AEMO's current IT infrastructure and existing AEMO capabilities and dataset knowledge but would need a little time bring in a new team of dedicated resources and dedicated leadership. Although this function would sit within AEMO, the entity would have a degree of independent governance, reporting through a steering group or advisory board of wider data interests. This committee would be comprised of data users, other data custodians, other energy market bodies, and government officials, and it would take a little time to establish and develop its arrangements. AEMO has a range of functions that have similar dedicated advisory groups (see Figure 7).

Refund and reshaped NEAR

Based on consultations, refunding, and expanding the NEAR program could be relatively fast given the NEAR institutional arrangements already exist, including prescribed access - taking up to 6 months to implement. Currently however much of the NEAR's data access arrangements are managed internally within AEMO, with NEAR accessing curated data through a data lab. For NEAR to take on this role (rather than burden AEMO – which would continue to look like one of the previous models) NEAR arrangements would need to be altered to allow more direct access. This could be similar to the "independent entity" model below but would avoid the need to set up a new organisation. Material time may also be needed to resolve a new funding model (as NEAR previously depended on Commonwealth funding) and re-negotiate the terms of NEAR UJV. However, it is possible that some services could begin in parallel given there are existing operational resources.

New independent entity

An independent model would take longer to set up than any of the previous models due the lack of existing arrangements and the complexity of setting up a new body and determining its governance and institutional arrangements as well as any regulatory changes required to access data or provide data custodianship services. This model could also face initial barriers in acquiring sufficient funding and delivering the critical services as quickly as stakeholders expect. The body would start by holding no data and would need to negotiate new data access arrangements and build supporting systems. Time would be needed to set up a Board which would govern and oversee the entity.

Data services network model

The longest model to stand up would be the data-services network model as it would rely on each stakeholder within the energy data ecosystem, including data users, to build up their own data services capabilities. It would take time to set up incentives structures or regulatory obligations need to drive this change, as benefits would be split across data holders and users. It would also take time to set up appropriate data protections across the parties and a new coordinating authority with dedicated resources to map data, data service offerings and common data sharing arrangements across the energy data landscape.

9.3   Ability to address current barriers

Resourced AEMO

Resourcing AEMO specifically to support data services would help address current resourcing constraints in the short term, with a likely expansion of services (funding arrangement discussed below). However, providing AEMO with more funding alone would not necessarily build a model focused on the priorities of wider data users and facilitate the creation of a best practice energy sector data delivery model in the long term. Without a separate governance structure ensuring resources are specifically dedicated to a data services team with clear goals, provision of data services would still compete with AEMO's core functions and necessary priorities. Without a clear transparency over user-driven priorities it may struggle to meet data

user expectations. As more data sources become available across the sector, AEMO alone may not be the best institution to provide centralised data services.

Dedicated unit within AEMO

The dedicated unit model aims to address some of the stakeholder concerns with AEMO's current ability to prioritise responsive data services and longer-term outcomes over urgent internal priorities – while preserving some independence and allowing AEMO to carry on with its main duties without diverting resources. The primary virtue of this model is it would combine the ability to relatively quickly provide data services related to existing AEMO data as well as having a semi-independent governance structure which would provide transparency and inform decisions in the best interests of the whole sector. To encourage quality and productivity of the data service offering and a greater drive to meet users' needs, the function would have a mandate to meet certain agreed service standards, research goals and be subject to periodic review. The model would build on existing IT infrastructure and AEMO skills but also provide additional dedicated resources to build out data service capabilities.

Refund and reshaped NEAR

The refunding and reshaping of NEAR could address the core need, particularly with direct access to existing analytical skills, experience with existing data sets and recent research. However, NEAR's UJV current drives a focus on longer-term research workplans and the priorities of the current partners. It would need structural reform to divert priorities and resources to capacity building and responsive services rather than research. It would also need to expand to wider joint venture partners (which could be complex), otherwise wider data users may have concerns about influencing priorities and ownership issues. NEAR's operator, CSIRO, is prescribed for AEMO to share data but currently still accesses data through curated systems and may need to expand its current access to reduce analytic constraints of not being a primary data holder.

New independent entity

The new independent body would be able to address current data services barriers by virtue of its ability to set new dedicated data-service related objectives and develop capabilities and legal arrangements to specifically focus on key data service challenges, unconstrained by existing approaches. It could be guided by the new aims and principles of the body in a light touch manner, with annual checkpoints and review periods to assess performance. A Board structure is proposed which could set fees and develop mechanisms to mitigate risks of anti-competitive behaviour by the body, given its centralised and potentially monopolistic position in the ecosystem, ensuring the body serve the interests of the energy data system users.

A new independent body could also be well placed to develop and manage new data infrastructure (such as the data dictionary) and develop a greater ecosystem of data service providers with wider sources of new energy data. However, as it would not be a primary data holder, and as the entity does not exist, it would take some time to set up this approach and would need to building skills and experience with the data sets.

A variation could be if AEMO or CSIRO spun-out part of its capability as part of seeding the independent new organisation. This could speed up and create efficiencies developing initial capabilities, but at the same time increase baggage of existing processes.

Data services network model

This model depends on building change and capability across a large number of organisations (as with CDR and the DATA Scheme, both of which have taken some years to establish). This would mean in the early stages a diverse range of existing internal barriers to address, across diverse organisational situations. Initial obligations/incentives to participate may take some time to address this range of internal challenges. Data protection arrangements with many organisations may be complex, initially depending on de-identified and aggregated data due to a wider range of bodies increasing risks. This approach may struggle to address many user needs in the near term. It would also still require effort in the short-term setting up a new centralised function and capabilities to streamline processes, manage standards, and provide incentives and enforcement.

9.4　Compatibility with broader policies and ability to be 'future proof'

Resourced AEMO

Increasing resources to the AEMO is likely to help address some current barriers but may not be as compatible with broader policies. The lack of external oversight and transparency could create a conflict of priorities (e.g. data services provision could become driven by AEMO priorities or priorities of specific funders). This would diminish stakeholder trust in the model's ability to balance its interests with the interests of data users. As more energy data sources also become available and more stakeholders involved, AEMO may not be the best institution to provide centralised data services; nor develop the necessary culture of data-sharing or innovation.

Dedicated unit within AEMO

The dedicated unit model would be an improvement on the resourced AEMO model as a semi-independent governance model could provide greater perspective on the holistic needs of the energy sector. However, in the medium to long term, the dedicated unit model may still prove less optimal than other model options due to more limited scalability and close affiliation with AEMO which may restrict the development of its own culture of data-sharing objectives and innovation. It would have fewer drivers to provide new services or facilitate access to data sets not held or required by AEMO. Robust arrangements for the stakeholder-led steering committee and in-sourcing new leadership and skills could manage some of these risks in the near term. Additionally, capability developed through this model in the short term could be transitioned to a more independent model in the medium term.

Refund and reshaped NEAR

Refunding and reshaping the NEAR program would leverage skills already developed in data analytics through investments already made in the program, as well as access to the wider skill and innovation culture of the CSIRO. However, there is a risk that the re-shaped NEAR will not overcome some of its past challenges included protracted arrangements for data access and the focus on solving 'big problems', rather than assisting data users with finding solutions to data reporting for day-to-day responsibilities and more basic data management needs. NEAR would also still require a wider set of joint venture partners than it currently contains which could increase the complexity of its set up and governance.

Noting that AEMO is a NEAR partner with full access to NEAR-IP and ongoing collaboration with CSIRO, some of the benefits of this model could also be captured by AEMO-led models. NEAR may provide more overall benefits by building on its advanced research strengths, supporting by new data services, to develop new capabilities, rather than diverting its resources to more narrow and operational data management services. New capabilities developed through NEAR (such as more advanced de-identification, linking or aggregation methods, or new data forecasting models or inputs) could be operationalised and supported through new data services.

New independent entity

The new independent entity without the baggage of an existing organisation and diverse priorities, would likely provide a more future-proofed, resilient, and adaptable model as it could be designed with a greater focus on future data needs and activities to maximise capability building and innovation. As part of setting up a new body, it would also be possible to provide clear objectives and foster a new culture focussed less on data protection at exclusion of other priorities and more on the principles of risk management, which is likely to be more conducive to an effective energy data management system.

Data services network model

The network model may be a long-term solution to fulfilling data services needs if it successfully incentivises the development of data service capabilities across a large number of organisations. In the long term, a network model might help create a greater market for independent data services provision that would benefit from multiple players, competitive pressure, and service delivery innovation. In the short term however, a data services network would have capability fragmented across many organisations, many of which may lack the critical mass to build expertise and trial innovative approaches.

Example 4. Challenges introducing a distributed data services network - UK Energy Digitalisation Taskforce – Open Data model

The UK model provides a useful case study of the greater complexities in a distributed or networks model.

The UK Energy Digitalisation Taskforce aims to unlock flexibility within the UK's energy system and realise the value of energy data assets. Its approach centres around long-term investment into the creation of federated data services model, in which data is treated as open by default. One of its key goals is to make datasets more visible via a single, searchable platform which connects requestors directly to custodians.

This is an ambitious and highly expensive undertaking, requiring investment across all data holders. The road to this model has been over five years in the making, which required the UK to ensure that goals of different stakeholders were firstly clearly aligned, with a shared purpose, direction and transparent governance which needed to be demonstrated in order to receive the scale of investment required.

The development of a distributed data services model was partly opportunistic, as there was a window of opportunity to integrate data-sharing and digitalisation obligations when operator and regulator licences were being renewed. While this approach has been effective at increasing data-sharing due to these imposed incentives, this is a heavy regulatory burden which has taken a long time to develop costly new arrangements. This model is also struggling to meet its current goals, as coordination between the bodies has been lacking making sharing and access difficult, as well as limiting ability to manage data across providers.

Therefore, the UK is now introducing a centralised body to provide data coordination, to streamline data-sharing standards, platforms and processes and manage compliance. Most distributed data sharing systems (like CDR and Open Banking) require onerous coordination, standards, and registries to ensure data remains interoperable and compliance is managed.

The UK undertook this approach starting from a different set of energy data arrangements than Australia. The UK lacks much of the existing data sharing infrastructure Australia already has. The UK has no central metering database like MSATS or related existing data exchange processes. Because of this, the initial open data obligations have focused on networks, requiring them to publish network performance data. Being able to analyse and access meter data remains a priority in the UK but was considered harder to implement as it would require data to be released by retailers.

For Australia to implement a similar data services network model, new regulatory requirements compelling networks and retailers to release data would be needed, similar to the UK. These requirements would likely take time to develop and impose a cost burden on these businesses.

Unlike the UK, Australia has alternative opportunities to build on existing data coordination (such as the central metering database and related data sharing requirements).[8]

9.5   Cost and financing

Resourced AEMO

Resourcing AEMO would likely have much lower costs than other models due to the ability to build on existing resources within AEMO. The model would require a small team of additional staff members (including business analysts, data engineers, visualisation engineers as well as management) and would require ongoing salary costs likely to equate to a few million dollars a year, with a small set up cost. Costs would scale with services, to some extent, so would depend on needs.

AEMO may be able to recover some of these costs as market fees, at least to the extent that data services supported market-wide benefits such as published outputs and system improvements. Any market fee would be an independent market fee set in partnership with the industry Financial Consultation Committee that currently works with AEMO in setting industry fees. The independent fee structure supports clear financial reporting processes.

Other costs could be recovered on a fee-for-service basis or subscriptions from more regular customers (such as governments). Initial capital funds may also seek direct contributions from government.

One of the challenges with this model however would be a potential lack of transparency of costs and fungibility with wider internal resources. Stakeholders may be concerned that wider AEMO resources are diverted to data services, or conversely that nominated data resources support internal priorities.

Dedicated unit within AEMO

A dedicated unit within AEMO would have similar low costs to the resourced AEMO model, leveraging existing systems and requiring a small new team (including business analysts, data engineers, visualisation engineers). Set-up costs are expected to be small with ongoing salary costs likely to equate to a few million dollars (initial proposals suggest 3-5 resources for the first 2-3 years). There may be some additional costs in this model, with greater clarity over the team being additional resources (rather than internally shifted or borrowed) and some resources needed to support the stakeholder steering group.

Funding sources would also be similar, with AEMO fee-based funding possible for some shared infrastructure and market outcomes (like published resources). Other funding would need to be sources from governments (for services, policy-focused resources like dashboards or research infrastructure) or service fees.

Unlike the resourced AEMO model, a dedicated unit model and funding allocations to different priorities would be more transparent and clearly defined, with less internal fungibility. The steering group would provide more rigour and clarity over how funding is sourced and allocated.

Refund and reshaped NEAR

This approach would build on existing resources under the current NEAR program requiring an estimated additional few million dollars for ongoing costs primarily related to salaries to deploy a data services team. However, the current model has depended on Commonwealth funding which has now ceased, with around $20 million in grants provided over six years to date. Alternative user-funded models would need to be found and reflected in the joint venture and priorities. AEMO may still be able to support some fee-based funding as a joint partner, but this would still be constrained to aspects of services with market-wide benefits. CSIRO

---

[8] Note that existing B2B data sharing arrangements are similar to the future the UK is moving towards, except with a more limited set of data users.

also has constraints in its cost/resourcing model focused on highly skilled research, which may make it less cost-effective for lower-skilled data management tasks and more basic analytics and reporting.

New independent entity

The new independent entity represents a more expensive solution compared to the dedicated unit model. The set-up costs would have higher system and capital costs and organisational overheads and are estimated to be significant (>$3 million) with ongoing costs totalling a few million dollars annually, but larger than AEMO options because of the need to hire additional back-office staff. In addition to the substantial set-up costs to build a new body and infrastructure capable of providing data custodianship as a service, ongoing costs will be required to maintain the talent and personnel and expand offerings as the energy data space evolves.

Sources of funding may be more limited, with less ability to draw on fee-based funding even for market-wider services such as publications. Dedicated funding would be required for set-up costs and would probably need to come from government.

Data services network model

The data service network model would likely have the highest set-up costs because many costs would be duplicated across many data custodians and users as each develops their own data service offerings, infrastructure, and capabilities. Fragments services may also have more limited scale efficiencies.

Initial funding required by each body would vary, depending on existing capabilities and the state of IT systems. AEMO would remain a major data holder needing to facilitate access, so few of AEMO's data services costs in other models would be avoided. Costs would then vary per data holder, with an additional $200k-1 million per data holder/data user. A coordinating body to manage the central catalogue, standards and compliance would also add costs, particularly as engaging a wide range of data holder would likely require obligations.

This cost would likely be borne by many individual data holders and users, with less ability to share costs. The split of costs and benefits across parties likely would require regulatory obligations or incentives to promote participation and investment, as seen with most distributed data models like CDR and the UK model. For example, CDR has imposed development costs on all retailers and AEMO, as well as maintaining enforcement and reporting capabilities (at ACCC) and a standards body. The DATA scheme also requires support of the Office of the National Data Commissioner to support capability development (such as standard agreements and training) as well as enforcement. While the DATA Scheme is voluntary, it is expected to drive uplift in data sharing investments across the Commonwealth driven by stakeholder needs.

9.6   Regulatory burden

Resourced AEMO

Increasing resources to the AEMO would not present extensive regulatory changes as it would be supported by existing reforms that are currently underway. ESB *Initial Reforms* proposals are currently being implemented to support AEMO sharing data with trusted bodies. These reforms would be sufficient to support most changes required, including giving AEMO a function to support data services, use data services and develop common data sharing agreements and registers. These reforms will be necessary under any of these service models, as under any of these models AEMO remains a major data holder who must play some role facilitating access.

In this model, limited drivers exist to facilitate access to wider datasets beyond AEMO's datasets, unless driven by separate targeted reforms (such as being considered under Data Strategy Priority Projects). Longer-term Data Strategy New Framework reforms could also support wider data holder access, likely in any scenario.

Dedicated unit within AEMO

The dedicated unit proposed is a light-touch internal governance arrangement with a steering committee providing advice to AEMO executive, based on clear objectives and requirements (see Figure 7). AEMO has committees overseeing different aspects of its operations which vary in their legal structure. This committee is proposed at least in the near term as largely advisory, to ensure wider data users and experts have full input to priorities and transparency.

In all legal senses AEMO will remain the legal entity and decision makers, and as such it would retain AEMO's data sharing rights and obligations[9], including the expanded sharing rights under the proposed *Initial Reforms*. This would avoid having to negotiate new data access contracts for key datasets.

More complex reforms would be needed in the medium term to support access to wider data sets. These could be considered through Data Strategy's New Framework and priority projects.

Refund and reshaped NEAR

The regulatory burden of refunding and in particular reshaping the NEAR program is more challenging. It is complex to reshape the joint venture and at the same time manage current risks to broad responsive services. Negotiations on key issues such as intellectual property ownership under a joint venture was a key hurdle in previous work to responsive user-driven services.

NEAR would also need to expand current data access arrangements, which depend heavily on AEMO resources. CSIRO is not a primary data holder but is now (since the NEAR agreements) a prescribed entity so should not need additional rights but may need to revisit current agreements. Negotiation of these agreements has previously been challenging; however, reforms may make this easier.

New independent entity

The independent body model will need substantive regulatory changes and new governance arrangements. They will need to develop a new legal structure with appropriate institutional arrangements, data protection obligations and immunity. Data protection obligations will need to sufficient for them to be granted Class 'A' status under ESB's proposed amendments to the National Energy Law (NEL) and National Gas Law (NGL), so that it can directly access protected data from AEMO and link it with other priority datasets. In addition to data from AEMO, negotiations would be required to acquire access to wider data sets voluntarily on appropriate terms. Regulatory requirements would also be needed to allow the new body to share data forward with third parties, as this would be central to their role.

Data services network model

The network model would require the most complex regulatory arrangements, with the need for extensive regulatory changes to both allow and oblige a wide range of data holders to share data, as benefits, costs and risks as split. This is similar to CDR and the UK model, who apply obligations on participants.

Regulation would also need to prescribe the functions of a new the coordinating authority. This body would need a range of enforcement capabilities, so may need supporting Commonwealth regulation / legislation as well as energy reforms. This would be substantive legal undertaking.

---

[9] Note that this is different to the recently established AEMO Services, who have more formal separation and do not inherit AEMO's data rights and obligations.

Low ▊▊ High

Table 6: *Summary model assessment against criteria*

This table below sets out an initial evaluation of the five potential models against the proposed criteria and the status quo

	Set-up time and governance structure	Challenge in addressing current barriers	Challenge to meet broader policies needs and adapt to be 'future proof'	Cost and financing	Regulatory and governance burden
Resourced AEMO	**Low,** Quick to set-up, ~6 months	**High**, hard to prioritise many foundational activities and capacity building over other urgent internal developments. May constrain responsive services and data user priorities.	**High**, competition from internal needs and priorities may constrain focus and limit investment in new capabilities and innovation.	**Low,** builds on existing resources but creates some resource risks. Would need to be clear which costs are AEMO fee-based versus user-funded services. Without dedicated resource, may lack transparency on cost allocation.	**Low**, supported by existing reforms underway, including expansion of AEMO's functions, but would need "New Framework" reforms to manage wider datasets. Would also need clarity over internal governance and objectives.
Dedicated unit within AEMO	**Low,** Quick to set-up, ~6 months	**Mid**, facilitates early access to AEMO data and provides a structure more responsive to external needs. May have a narrow focus in the long-term.	**Mid**, Dedicated resources and steering group will provide additional focus on new needs and capabilities. Less transparent and independent than some other models in the longer-term – but can incrementally grow into other models.	**Low,** builds on existing resources. In a dedicated unit model, costs would need to be explicitly identified and dedicated resources added. Fee-based versus user-funded services would need to be transparent.	**Low,** supported by existing reforms underway, including expansion of AEMO's functions, but would need "New Framework" reforms to manage wider datasets. Internal governance would need clarity on dedicated resourcing and oversight by a steering/advisory committee – but still relatively modest as AEMO has a range of similar committees/advisory functions.

	Set-up time and governance structure	Challenge in addressing current barriers	Challenge to meet broader policies needs and adapt to be 'future proof'	Cost and financing	Regulatory and governance burden
Refund and reshaped NEAR	**Low,** Quick to set-up, ~6 months	**Mid,** UJV creates challenges to change. Focus on long-term workplans and partner priorities constrains flexible and responsive services. Would need structural reform to divert priorities to capacity building and responsive services. May have ownership issues diversifying operations beyond CSIRO.	**Mid,** Access to in-depth skills but difficult to be responsive to a wide range of stakeholders due to funding and governance model. NEAR could likely add greater value in an ongoing research role, working with data services to support innovation rather than taking over the data mgt/operational role.	**Mid,** Builds on existing resources - but current model depended on Cwlth funding which has ceased. Alternate user-funded models would need to be reflected in the UJV and priorities. May be hard to diversify service providers, for example to source cost-effective lower-skilled areas like data management and maintenance.	**Mid,** CSIRO has recently been allowed prescribed access to AEMO data. But reshaping and diversifying the UJV with a wider number of parties would be complex. Key issues like IP ownership have proved complex in the current UJV model.
New independent entity	**Mid,** Moderate time to set up, >12 months	**Mid**, without the burden of incumbent arrangements could be designed to meet the barriers but would take time and require aligning costs and regulator.	**Low,** New organisation allows greater ability to define clear new goals, focusing on future needs and innovation.	**Mid**, some potential efficiencies in designing a new and dedicated structure, but high initial set up costs without cost-sharing with a larger-scale organisation and existing infrastructure.	**Mid,** regulatory requirements to set up a trusted new body with clear data protection and governance arrangements, as well as clear access to data.
Data-services network model	**High**, Longest to set up; set-up time will vary between bodies.	**High**, organisations would face diverse challenges in resolving short-term barriers The split of interests and benefits between data users and data holders would still make many current barriers a challenge.	**Mid,** Moderate compatibility with broader aims due to focus on improving capability for all stakeholders. Diversity may support l innovation, but innovation may also be constrained by lower concentration of skills and experience.	**High,** duplication of services, resources and capabilities across a larger number of bodies, each of which would lose some scale efficiencies.	**High,** regulatory change required to oblige a wider range of agencies and data holders to participate and support data service requests; governance would need to be agreed on.

Consultation questions:
Q 9: The ESB welcomes stakeholder views on the initial assessment of the strengths and weaknesses of each model presented here?

10 Phased approach to delivery

The initial assessment highlights that there may be different optimal models in the short and longer term. While we have not sought to set out implementation details for any of these models, the ESB recognise there is likely to be merit in taking a phased approach to delivery of future data services. This recognises the greater potential of some models to deliver value from data services over the short term, unlocking potential benefits earlier. A phased approach is also intended to enable key priorities to be developed first, build capabilities and support having initial models set up such that they can meet changing needs.

Therefore, as well as considering stakeholder views on individual approaches, ESB is interested in stakeholder views on a phased approach to develop models for delivery of data services over time.

In the short to medium term:

- Leverage the ability of the dedicated unit within AEMO to be formed relatively quickly, building on existing assets and capabilities to get short term benefits. This body could be tasked to establish many of the common processes, methods and platforms needed, and work with data users to develop capabilities and establish priorities as data needs grow.
- Leverage changes proposed as part of *Initial Reforms*, enabling access of AEMO data to a broader range of policy makers and jurisdictions (i.e. defined in the proposals as Class A and B bodies)
- As part of this phased approach, initial priorities for the first 2-3 years of delivery would include development of those data services which address gaps to facilitate a chain of access and benefits, including:
  - o Data dictionary; requirements gathering; custodian; contracting; access; de-identification; compilation and analysis; visualisation and insights.
- To support development of data services consistent with policy making priorities for the energy transition, under this model advisory group recommendations would be provided to energy senior officials for approval. The advisory group would include stakeholder, market body and industry representatives, with specific inclusion of innovators, research and parties with data or technology capabilities. This will support a tighter focus on key priorities for initial implementation, enabling core capabilities to be built progressively and learn / adapt from doing. Resourcing requirements to support this first phase of implementation would be similarly focussed.
- Design processes to flexibly consider future needs, taking a 'no regrets' approach such that data dictionaries, standard agreements and methods could be migrated to a range of other governance models as needs change over time.
- Provide opportunities to trial approaches and learn by doing in the short term.
- During this phase continuing data reforms under the Data Strategy would work with wider data holders on options to involve them and develop new data sets, as currently intended through the wider "new framework" reforms and priority data projects.

In the medium term:

A review of these arrangements would be built in at 2-3 years post implementation. This review would assess benefits and outcomes to date, progress, and learnings in common arrangements, and how data needs are evolving with the market. These insights would inform development of future data services delivery models (either within AEMO or in other distributed models). For example:

- *Independent body*: if the scale of data sharing needs has grown and there is value seen in greater capacity to innovate, but wider data holders do not want direct mandates, aspects of the dedicated unit model could be migrated out into a new more-independent arrangement to provide services, focused on priorities identified.
- *Data network*: if wider data holders and independent data service providers seek a more direct role, aspects of the dedicated unit model developing common processes could be migrated into a coordinating regulator (possibly combined with aspects of AER or another technical regulator's role). This body could lead reforms and work with wider data holders to create a more distributed services model.

- ***Dedicated unit model***: if data needs are being well serviced and needs have not outgrown this approach, it may be appropriate to continue with a dedicated unit model, but with a range of adjustments taking on learnings from the review.

It also likely that future data service delivery may be a combination of aspects of these approaches, with some datasets and services being provided centrally, and others being provided in a distributed way.

Consultation questions:

Q 10: The ESB welcomes stakeholder feedback on its proposed phased approach to delivery of data services.

Q 11: The ESB welcomes feedback in particular on how well models deployed for this first phase of delivery (e.g. AEMO dedicated unit model) might be able to evolve or transition to other future models.

Q 12: The ESB welcomes views on what might be priority features, services or data sets as part of this first phase.

11 Next steps

The ESB invites comments from interested parties in response to this consultation paper by February 13th, 2023. While stakeholders are invited to provide feedback on any issues raised in this paper, the key questions for consultation are set out in Appendix A. Submissions will be published on the Energy Ministers' website, following a review for claims of confidentiality.

Further enquires on this consultation paper can be sent to the project team at info@esb.org.au.

Submission close date	13 February 2023
Lodgement details	Email to: info@esb.org.au
Naming of submission document	[Company name] Response Data Services Delivery Model Consultation Paper December 2023
Form of submission	Clearly indicate any confidentiality claims by noting "Confidential" in document name and in the body of the email.
Publication	Submissions will be published on the Energy Ministers website, following a review for claims of confidentiality.

The ESB intends to hold a webinar on the material covered in this paper on **7 February 2023**. Interested parties are invited to register their interest by email to info@esb.org.au

Appendix A: Primary consultation questions

This consultation paper reflects initial consultations with energy market bodies, government data users and other energy data stakeholders.

The Energy Security Board invites stakeholders to share their views on the approach for developing energy data services by responding to the questions in this consultation paper.

Consultation Questions - Summary
Q 1: Are there any priority data services missing from the analysis? **Q 2: Are there other barriers that inhibit data services not identified here? Can you provide examples or case studies of these barriers or are there examples of where data services are addressing any key barriers?** **Q 3: The ESB welcomes feedback on the features proposed for data services delivery models. Are there other considerations that should also be taken into account?** **Q 4: What are stakeholder views on the appropriate scope for data services in the short to medium term?** **Q 5: Are there other data service delivery models that could be considered?** **Q 6: Are there better governance models for the AEMO dedicated unit proposal, outlined in the example?** **Q 7: Are there other benefits, challenges and implementation issues that should also be considered? Are there any cost considerations that haven't been explored in this paper?** **Q 8: The ESB welcomes stakeholder views on whether the proposed criteria are appropriate to assess alternative data service delivery models?** **Q 9: The ESB welcomes stakeholder views on the initial assessment of the strengths and weaknesses of each model presented here?** **Q 10: The ESB welcomes stakeholder feedback on its proposed phased approach to delivery of data services.** **Q 11: The ESB welcomes feedback in particular on how well models deployed for this first phase of delivery (e.g. AEMO dedicated unit model) might be able to evolve or transition to other future models.** **Q 12: The ESB welcomes views on what might be priority features, services or data sets as part of this first phase.** **Q 13: The ESB welcomes stakeholder insights and examples which may be relevant in testing these illustrative examples or providing additional examples.**

Appendix B: Illustrative examples - How will data services impact data users?

This appendix provides a range of illustrative examples of the expected impacts of data services on different data users and needs.

These include:

A. State policy body (Class A) seeking to test the effectiveness of a program.
B. Australian University (Class B) looking to undertake research on vulnerable consumers.
C. Local government groups looking to test impact of climate programs.

> **Consultation questions:**
>
> **Q 13: The ESB welcomes stakeholder insights and examples which may be relevant in testing these illustrative examples or providing additional examples.**

Illustrative example A

User	State policy body (Class A).
Use case	Seeking to test the effectiveness of a new program providing CER and energy efficiency support to households.

Data service	Without data services and initial reforms	With data services and initial reforms
Data catalogue	There is limited information about what data is available. Data user is unable to test data availability to meet their needs. They may miss the opportunity to gain insights on their program or improve consumer outcomes due to assumptions about limits on data.	Data users can seek information on what data is available, who holds it and any relevant constraints. Data users have the opportunity to seek data to solve problems.
Data requirements gathering	When a data user seeks advice to gain insights on their program, there is no clear AEMO entry point or experts tasked with engaging with them to test their needs. Where a data user does make a request often the onus is on the data user to frame and scope the question and outcomes required. This is risky without expertise in the data and contexts, often leading to investment in analysis which hit barriers and have limited outcomes. AEMO experts may be diverted from their priority tasks seeking to assist, but without clear frameworks, methods and data curation may not fully identify risks.	Data user has a clear and identified expert point of reference and a framework for testing their questions and options. Data Services experts are available to scope and test the approach, identifying hurdles and risks early to ensure improved outcomes prior to investment. This provides an opportunity to also improve innovation, with a range of approaches considered and tested early in the process.
Data custodian	AEMO knows it is unable to release confidential data to the user and is cautious due to ambiguity in their current requirements – so will have to do the analysis for the user.	AEMO knows that the Class A body will be able to access the data and/or results. Data services team has dedicated expert resources available to either assist with curated data access or

	AEMO does not have dedicated resources to meet these needs. AEMO will have to consider the priority of the work against internal needs to identify if/when resources available can be made available. This has often required months of delay. Often the appropriate resources are in AEMO's planning and forecasting areas and are subject to a range of cyclic reporting needs, causing months of delay to await an available window in the reporting cycle. Without a clear role in analytics, AEMO data sets are curated and managed for operational purposes. Its structure and quality are frequently not pre-optimised to support analytics. This can mean that for every analytics task undertaken data quality and standards issues need to be identified and re-addressed, costing time and causing duplication.	undertake a full range of services, on a fee-for-service basis. They have some capacity to scale and manage priorities in a timely fashion. Recognising ongoing data services needs, data is effectively curated to support improved outcomes and efficiency. This includes identification and active management of data quality issues, identification of standard repeatable methods and linking, improvements to the data architecture (such as pre-categorisation of data and ensuring standardised linking keys such as address) and a range of pre-existing published aggregates and de-identified datasets to support self-service where possible and scoping of analytics.
Data Contracting	AEMO and the state body negotiate a bespoke agreement. Lawyers on both sides recognise their requirements are ambiguous and are cautious. It takes over 6 months. A range of scoping elements in the problem remain unclear and result in risks emerging at a later point.	Data services have standard published terms, including a template MOU specifically for Class A state policy bodies. After testing the scope of services in the data requirements phase to identify risks and limitation and develop clear agreed outcomes, the Class A body signs an MOU for the specific services in 2 weeks.
Data Accessing	With limited ability to release data AEMO provides only aggregate outcomes and only when clearly protected. This means the data user may not be able to interrogate results or understand unexpected outcomes. Where AEMO does seek to provide large datasets, there may be significant bilateral work with the agency seeking appropriate safe ways to transfer data, with not standard systems in place.	Data services have a data lab environment allowing the data user to view and access the data safely. They use this environment over the course of the work to show the user results at different phases. As a Class A data user, the agency is allowed to see full details and interrogate data or undertake their own analysis.
De-identification	De-identification is limited as standard methods are not agreed. Limited research into better methods and re-identification risks mean that the approach is cautious. Only broad aggregates are supported, which are often inappropriate for the analysis needed.	Data services de-identify the meter usage data, using a standard perturbation method, reducing the sensitivity of the data without limiting its value and insights.
Compilation and analysis	Linking can be challenging and time consuming without standard keys. Privacy issues can limit linking if individual consumer consent is not provided in the program data, allowing the state body to release data to AEMO. There are no standard methods for commonly undertaken analyses – requiring time consuming development in each case and a range of barriers and challenges.	Data services uses a standard linking method with third party address keys to link the state bodies program data to NMIs. The Class A state body does not have extensive internal expertise and resources in data analysis and so seeks support from Data Services for the insights and analytics. Because testing the impact of a program or change on a group of households is a common need, data services also use a number of standard approaches to short-cut analysis. This includes robust approaches to control groups, pre-existing

		classifications to "cluster" similar types of household patterns and consider driving characteristics, and methods to test the statistical relevance of impacts and variations.
Visualisation/in sights	With limited de-identification and concerns over social licence, the data user is constrained from secondary disclosure, limiting ability to share results as evidence.	Data services supports expertise in how to frame, communicate and protect results, allowing for greater impact and release of insights. As a Class A data user, the data sharing agreement does not impose restrictions on secondary disclosure. Data services provides advice on which datasets may be sensitive to share and which are fine to release. The Class A data users is informed to manage protect data and manage their own data sharing risks.

Illustrative example B

User	University research (Class B).	
Use case	Research on vulnerable consumers.	
Data service	*Without data services and initial reforms*	*With data services and initial reforms*
Data catalogue	There is limited information about what data is available. Data user is unable to test data availability to meet their needs. They may miss the opportunity to gain insights on their program or improve consumer outcomes due to assumptions about limits on data.	Data users can seek information on what data is available, who holds it and any relevant constraints. Data users have the opportunity to seek data to solve problems.
Data requirements gathering	When a data user seeks advice to gain insights on their program, there is no clear AEMO entry point or experts tasked with engaging with them to test their needs. Where a data user does make a request often the onus is on the data user to frame and scope the question and outcomes required. This is risky without expertise in the data and contexts, often leading to investment in analysis which hit barriers and have limited outcomes. AEMO experts may be diverted from their priority tasks seeking to assist, but without clear frameworks, methods and data curation may not fully identify risks.	Data user has a clear and identified expert point of reference and a framework for testing their questions and options. Data Services experts are available to scope and test the approach, identifying hurdles and risks early to ensure improved outcomes prior to investment. This provides an opportunity to also improve innovation, with a range of approaches considered and tested early in the process.
Data custodian	AEMO knows it is unable to release confidential data to the user - so will have to do the analysis for the user and release only aggregate results. Resourcing constraints may cause major delays, particularly if the work is not supporting with wider AEMO priorities.	As a Class B body, AEMO knows it will likely be able to release the data.
Data Contracting	AEMO and the university lawyers negotiate a bespoke agreement – including discussion of aggregate limitations. Lawyers on both sides recognise their requirements are ambiguous and are cautious. It takes over 6 months with a range of issues around IP and secondary disclosure to allow research to be published.	Data services have standard published terms, including a template agreement specifically for Class B universities. This includes a range of accreditation requirements. The University is already accredited with the Commonwealth DATA scheme – which AEMO confirms and recognised. The Class B body signs an agreement for the specific services in 4 weeks.
Data Accessing	With limited ability to release data AEMO provides only aggregate outcomes and only when clearly protected. This means the data user may not be able to interrogate results or understand unexpected outcomes.	Data services have a data lab environment allowing the data user to view and access the data safely. They use this environment over the course of the work to show the user results at different phases. As a Class B data user, the university is allowed to see full details and interrogate data and undertake their own analysis.

De-identification	De-identification is limited as standard methods are not agreed. Limited research into better methods and re-identification risks mean that the approach is cautious. Only broad aggregates are supported, which are often inappropriate for the analysis needed.	To allow the Class B body to publish research, data services agreed a de-identification methods to allow release of results.
Compilation and analysis	Linking can be challenging and time consuming without full visibility of protected data. There are no standard methods for commonly undertaken analyses – requiring time consuming development in each case and a range of barriers and challenges.	The Class B university seeks to undertake most of its own research. However, data services support this by linking AEMO data to the universities protected data safely and providing it back to the University through the protected data lab. It also advises on standard methods, related learnings from past analysis and data quality issues.
Visualisation/insights	With limited de-identification and concerns over social licence, the data user is constrained from secondary disclosure, limiting ability to share results as evidence.	Data services supports expertise in how to publish research results safely and meet the requirements on the Class B body.

Illustrative example C

User	Local government or community groups.	
Use case	Seeking to test impact of local climate programs.	
Data service	*Without data services and initial reforms*	*With data services and initial reforms*
Data catalogue	There is limited information about what data is available. Data user is unable to test data availability to meet their needs. They may miss the opportunity to gain insights on their program or improve consumer outcomes due to assumptions about limits on data.	The data user can see what data is available and issues around its availability to identify opportunities.
Data requirements gathering	When a data user seeks advice to gain insights on their program, there is no clear AEMO entry point or experts tasked with engaging with them to test their needs. Where a data user does make a request often the onus is on the data user to frame and scope the question and outcomes required. This is risky without expertise in the data and contexts, often leading to investment in analysis which hit barriers and have limited outcomes. AEMO experts may be diverted from their priority tasks seeking to assist, but without clear frameworks, methods and data curation may not fully identify risks.	Testing whether outputs/insights can likely be provided which meet the need, while being appropriately de-identified, is even more important where AEMO cannot release protected data (as the party is not Class A/B) (frequently this may be probabilistic and can be impacted by sample sizes)
Data custodian	AEMO knows it is unable to release confidential data to the user (as neither Class A/B) - so will have to do the analysis for the user and release only aggregate results. Resourcing constraints may cause major delays, particularly if the work is not supporting wider AEMO priorities.	As data cannot be released, Data Services provides resources to the analysis (likely fee-for-service) without delay. Data Services may also curate a range of public or self-service datasets, particularly in case like this one where a range of users may seek similar information (for example regional aggregates of electricity-driven emissions over time allowing all local governments to track aggregate impacts). Priorities for these kinds of aggregates may be negotiated through public consultation with data users.
Data Contracting	AEMO and data users lawyers negotiate a bespoke agreement – including discussion of aggregate limitations. Lawyers on both sides recognise their requirements are ambiguous and de-identification methods limited. So the agreement is cautious, setting very broad definitions of aggregation requirements and limiting detail in outcomes.	Data services have standard published terms, including a template agreement specifically for non-Class A/B proponents. Standard de-identification definitions are included and well tested, with a good understanding of the technical limitations, allowing the maximum level of detail possible without unreasonable risks.

Data Accessing	AEMO provides only aggregate outcomes, with the user having limited ability to interrogate or test outcomes.	Data services have a data lab environment allowing the data user to view and access data safely while still protected. They use this environment over the course of the work to show the user results at different phases and discuss what can be released.
De-identification	De-identification is limited as standard methods are not agreed. Limited research into better methods and re-identification risks mean that the approach is cautious. Only broad aggregates are supported, which are often inappropriate for the analysis needed.	Standard and advanced de-identification methods, with tested re-identification risks, allow for more useful aggregates to be provided with clearer security.
Compilation and analysis	Linking can be challenging and time consuming without full visibility of protected data. Where both set of data to be linked must be protected, caution in both parties currently limits linking, particularly where different protection methods have not been tested to provide confidence.	The local government has program data that could be linked but must also be protected. Data Services supports tested safe third-party linking methods to allow the data to be linked without breaching the local government data protections requirements.
Visualisation/insights	AEMO does not have a mandate to support ongoing portals or products for a wide range of users.	Data Services recognises that some outputs from this study could be usefully standardised for wider users with similar needs (such as regional electricity-based emissions over time). Data Services consults user groups and supports a new ongoing, self-service set of products on their portal.