SCARLATTI
Evaluation | Analytics | Insights

Prepared for:
Our Land and Water

Multilateral Data Sharing

Lessons from a case study with organisations managing beef genetics data

December 2023

Scarlatti.co.nz

Contents

Key contact

If you have any questions about this document or require any further information, **please contact Adam Barker (Director, Scarlatti)** by phone (021 220 9026) or email (adam.barker@scarlatti.co.nz).

Executive summary

The Multilateral Data Sharing project

Multilateral data sharing (MLDS) is a long-held goal of many organisations in New Zealand's food and fibre industries. It is also an efficient way to maximise the utility of information held across a network of organisations. In 2022 Scarlatti worked with a range of organisations to develop ideas to assist organisations working in New Zealand's food and fibre industries to facilitate MLDS.

In Phase 1 of the resulting Multilateral Data Sharing project (Phase I) proposed two tools to formalise systems and processes for groups interested in data sharing:

1. A data sharing analysis framework which formally analyses actors' motivations and hesitations, and enablers that could shift each actor towards data sharing.
2. A negotiation mechanism which offers a structured process to support the formation of MLDS agreements.

In Phase II of the Multilateral Data Sharing project, Scarlatti extended the work in Phase I by undertaking a proof of concept with Beef and Lamb New Zealand (B+LNZ) Genetics, associated breed societies and other related parties (a 'data sharing collective' in the language of the framework developed in the first part of this work). This enabled us to evaluate the framework and mechanism, thereby refining the original tools and delivering enhanced versions for future applications.

B+LNZ Genetics has several objectives that align with the Our Land and Water (OLW) National Science Challenge, including reductions in land erosion (breeding smaller cows) and nitrogen excretions (through breeding indices). However, these objectives depend on data sharing between numerous stakeholders.

The beef genetics data case study

The actors in the data sharing collective (referred to from here on as 'the collective') were:

- B+LNZ Genetics
- Beef breed societies
  - Angus New Zealand (Dave Warburton, Breed Development Advisor)
  - New Zealand Herefords (Robert Peacock, Vice President)
  - New Zealand Shorthorn Beef Association (Russell Proffit, President)
- Performance Beef Breeders New Zealand (PBB)

Other parties were not part of the collective but were regularly mentioned in discussions and could potentially be involved in a wider data sharing agreement in time:

- Meat processors
- Genetics analysis service providers
- Commercial farmers

Research approach

We tested the data sharing analysis framework by:

- Convening group meetings
- Conducting confidential one-to-one discussions
- Carrying out an analysis using the data sharing collective framework
- Documenting observations in a reflexive journal
- Reflecting back findings and observations
- Drawing out general lessons from specific anecdotes

We had initially anticipated using the findings from the steps described above to inform and test the negotiation mechanism proposed in Phase I. However, during this work it became evident that this was unlikely to be appropriate for the collective because:

- Parallel negotiations were already progressing. Without a formal mandate to act as a facilitator, it was inappropriate for Scarlatti to run a separate negotiation process in tandem.
- During discussions, we started to question how effective the highly structured approach that we developed in Phase I would be.

Conclusions

The case study validated several ideas proposed in Phase I of the MLDS project and highlighted some opportunities to refine both the data sharing analysis framework and the negotiation mechanism. The case study has:

- Validated the idea that 'soft' human and organisational factors, and incentives, are as important, or more important, than data sharing mechanisms.
- Reinforced our belief that a formal, structured framework for analysing actors' motivations, hesitations and enablers is useful. At the same time, it has highlighted gaps in our original data sharing collective framework.
- Suggested that we should do more to frame 'soft' human and organisational issues in the data sharing collective framework. In particular, the case study has highlighted the importance of understanding the level of trust between actors as a part of the process.
- Reinforced some of our ideas about the negotiation process – such as the value of a facilitator – but challenged our thinking in other areas – such the value of a fixed process to run a negotiation.

Refinements to the tools

We used the findings from the case study to refine the two tools proposed in Phase I of the project by:

- Replacing the proposed negotiation mechanism with a list of six ideas that parties working toward MLDS could incorporate into their discussions:
    1. Acknowledge the complexity

2. Use an independent facilitator
3. Undertake a formal analysis
4. Recognise and address human/soft organisational factors upfront
5. Break negotiations into stages
6. Consider a ‘yes or no’ approach

- Expanding the data sharing collective framework, structuring it as a list of questions that the group should work through, and reorienting it to focus on human/soft organisational factors, incentives, tech c.f. motivations, hesitations, enablers.

Introduction

Multilateral Data Sharing

Multilateral data sharing (MLDS) is a long-held goal of many organisations working within New Zealand's food and fibre industries. In this document MLDS refers to the exchange of *datasets* between three or more *actors* – farms, businesses, industry organisations, research providers and/or government agencies that work throughout the food and fibre industries.

It is widely accepted that MLDS would provide benefits to farmers, and the organisations that work with them. Benefits include reducing the time and cost involved in providing or acquiring data, as well as opening access to more timely and precise information. Although there is only limited evidence directly supporting this notion, the logic is that MLDS would facilitate the use of tools that support farm management, decision-making, and benchmarking – further enabling New Zealand's primary sector to compete within international markets.

Substantial effort and investment have been directed at attempts to unlock the benefits. For example, current and recent initiatives that relate to MLDS include:

- Agritech Data Reference Group
- Aotearoa NZ Agri Data Exchange initiative
- Trust Alliance New Zealand
- Commercial tools such as Agrigate and FarmIQ
- Prior to the 2023 election, the National party released their 'Getting back to farming' policy. This policy includes a 'no duplication rule' – meaning farmers will only be required to supply data once, transferring data sharing responsibility to the officials who received the data.

Despite initiatives like these, and the past attempts to solve the problem, widespread MLDS has not yet emerged. Apparently, the barriers to MLDS are even more formidable than the designers of previous initiatives have appreciated.

The Multilateral Data Sharing project (Phase I) – developing the tools

Background

In 2022 Scarlatti worked with a range of industry experts and relevant stakeholders to develop ideas to assist organisations working in New Zealand's food and fibre industries to facilitate MLDS.

This work was prompted by an observation that past and present attempts to facilitate MLDS have, or are, mainly focussed on data interoperability and/or systems to facilitate data exchange. In a sense, these are the more concrete components of MLDS. The incentives of actors to share data, and the economic, organisational, and social costs of data sharing, had not received the same level of attention.

It was observed that the current situation with respect to MLDS in New Zealand's food and fibre industries loosely resembles the well-known prisoner's dilemma. In game theory, the prisoner's dilemma is a game that presents a situation where two parties, separated and unable to communicate, must each choose between co-operating with the other or not. The highest reward for each party occurs when both parties choose to co-operate, but it is difficult to co-ordinate this outcome. A

hypothesis was formed that framing the current situation as a game theory problem might allow the development of solutions by introducing new incentives to break the deadlock.

The hypothesis was initially explored on the assumption that solutions to data interoperability would be addressed by other work, and that data sharing technologies already exist or could easily be developed. That is, the work largely focused on the incentives, and the economic and social barriers. However, it was observed that attempting to decompose complex data sharing problems by assuming that data interoperability/data sharing technologies and incentives/barriers would be treated independently did not work well. Even in the case of simple models of data sharing ecosystems, it became difficult to separate these.

The project evolved to take a more holistic approach to understanding this problem. Using a game theory lens still proved valuable by providing a more systematic approach to unpacking the incentives and barriers, ultimately leading to a proposed negotiation mechanism. Combining this with stakeholder engagement, qualitative logic, and thought experiments, the findings from this report are arguably not only more practical, but more representative of real-life.

Concepts from Phase I

Data sharing collective

The work in Phase I framed a MLDS scenario in which a group of organisations (*actors*) come together to negotiate a data sharing agreement between themselves. The group was termed a *data sharing collective*.

It is assumed that actors in a data sharing collective are at least partly motivated by bringing about collective benefits for an industry, region or similar. This means the beneficiaries of the desired data sharing may include actors that are not present at the table – in particular individual farmers. It could also mean that actors are more willing to share information about their *motivations* and *hesitations* to share data than they would be if they were negotiating purely for commercial advantage.

The data sharing collective framing is probably a good model for initiatives like the Aotearoa NZ Agri Data Exchange initiative noted above, and for the DataLinker initiative discussed in the Phase I report. It is also a good model for the case study discussed in this report.

However, a data sharing collective does not describe all initiatives relating to MLDS. For example, initiatives like Trust Alliance New Zealand and the Agritech Data Reference Group are probably better described as *enablers* that facilitate data sharing collectives to operate. We acknowledge, therefore, that the simple model of a data sharing collective that we use in this work is an idealisation and that the ideas presented in here will need to be adapted for different situations and contexts.

Importance of human and soft organisational factors

The work in Phase I highlighted the importance of 'soft' factors as being as important, or possibly more important, than data and technical challenges. These include differing organisational priorities and hidden conflicts of interest. We build on this idea in the current project.

Tools to assist in forming MLDS agreements

In Phase I we proposed two tools to assist in forming MLDS agreements:

- A *data sharing analysis framework* which aspires to provide a robust way to characterise and analyse a *data sharing collective*, inclusive of the *actors*, their *datasets* and *applications*. An important part of the framework is a formal analysis of:
    - *Motivations* – *Actor*-specific benefits to *data sharing* in a determined *collective.*
    - H*esitations* – *Actor*-specific barriers to *data sharing* in a *collective.*
    - *Enablers* – Things which make *data sharing* possible – overcoming *hesitations* and unlocking *motivations*.
- A *negotiation mechanism* which provides a structured process to support the formation of a *data sharing agreement.*

We tested and updated these tools in Phase II of the project, and present the findings in this report. We also used the findings to update the tools developed in Phase I.

The Multilateral Data Sharing project (Phase II) – beef genetics case study

The current project extends the work in Phase I by undertaking a proof of concept in partnership with Beef and Lamb New Zealand (B+LNZ) Genetics, associated breed societies and other related parties. At the time this project was undertaken, this case study group was working toward a data sharing agreement. Therefore, the goal was to use the case study group to evaluate the analysis framework and negotiation mechanism proposed in Phase I, thereby refining these tools and delivering enhanced versions for future applications.

B+LNZ Genetics has several objectives which align directly with the Our Land and Water (OLW) National Science Challenge, including reductions in land erosion (breeding smaller cows) and nitrogen excretions (through breeding indices). These objectives depend, in part, on data sharing between numerous stakeholders. The case study also serves to refine the two tools for other data sharing purposes that also contribute to OLW's goals.

During this process Scarlatti was, for the most part, an observer to the discussions underway between the actors rather than being an active participant or a facilitator. That said, the process of talking to the actors' representatives and sharing our findings back was acknowledged as contributing to the negotiation process by offering different perspectives and insights to members of the data sharing collective. These helped to clarify thinking and shape data sharing goals.

Purpose of this report

This report serves two main purposes. Firstly, it documents findings from the beef genetics data sharing collective case study. These are drawn from group and individual discussions, along with a reflexive journal maintained throughout the case study. We have also incorporated the case study group's collective feedback about benefits, limitations and possible refinements from their perspective.

Secondly, it updates the concepts and tools developed in Phase I of the project, and more explicitly frames these two tools for use in future MLDS negotiations as such. This includes incorporating new ideas that arose from the case study and simplifying previous recommendations.

The beef genetics data case study

The actors

The actors in the beef genetics data sharing collective (referred to from here on as 'the collective') were:

- **B+LNZ Genetics**, represented by Dan Brier, General Manager. B+LNZ Genetics has several objectives that are dependent on MLDS between themselves and other organisations including PBB and breed societies. For example, they are investigating how to use genetics to reduce nitrogen excretions, methane emissions and land erosion (through breeding smaller cows) in New Zealand's beef industry. These genetic traits would be incorporated into breeding indices, allowing breeders and farmers to demonstrate/identify animals with these desirable traits.
- **Beef breed societies**. All beef breeders in New Zealand are members of a breed society, however not all members of breed societies are beef breeders. Three breed societies participated in the case study data sharing collective, represented by:
  - Angus New Zealand (Dave Warburton, Breed Development Advisor)
  - New Zealand Herefords (Robert Peacock, Vice President)
  - New Zealand Shorthorn Beef Association (Russell Proffit, President)
- **Performance Beef Breeders New Zealand (PBB)**, represented by Harry Faas, General Manager. PBB is a limited liability company whose shareholders are the breed societies. PBB acts as a service provider to stud farmers, with a variety of services including managing the flow of data between breed societies and evaluation companies.

Other parties were not part of the collective but were regularly raised in discussions and could possibly be involved in a wider data sharing agreement in time:

- **Meat processors**. While not part of the discussions, meat processors are an important part of the system to generate value from beef genetics data. Their potential involvement was discussed extensively during this work.
- **Genetics analysis service providers**. Several firms currently provide analytical services to organisations in the collective (or could potentially do so). As with meat processors, these service providers were mentioned regularly in this work.
- **Commercial farmers**. Commercial farmers are the ultimate beneficiaries of the work being undertaken by the collective. While they were not directly involved, their interests were front of mind throughout the discussions.

Research approach

Aligning to existing discussions

In an ideal pilot of the tools developed in Phase I of this project, the collective would have been formed as a part of the case study, and Scarlatti would have taken a facilitator role. Arguably, this would have provided the most opportunity to test the tools in a systematic way.

In this case, however, the collective was already formed and was already working towards a data sharing agreement when the case study started. The research approach evolved to fit around the discussions already in progress. In effect, the research became a parallel process to the main negotiation. However, this is not to say that the research was entirely passive – the process had an influence by reflecting observations and insights back to the actors in a two-way sharing process.

Testing the data sharing collective framework

Specific steps in this work were:

- Group meetings
  - These were held at the establishment phase of the project, initially with B+LNZ Genetics and PBB. These two actors were responsible for identifying which other actors should be invited to join the collective.
  - A group meeting was then held with all five actors to reaffirm context of collective, mandate, and how process would unfold.
  - A group meeting was also held after the data sharing analysis framework had been applied, to assess its value for the collective and seek feedback on refinements. It is important to note that we did not share the actors' viewpoints with one another at this meeting, as this would have breached the confidentiality and trust we sought to maintain throughout.
- One-to-one discussions
  - Each actor was invited to a confidential one-on-one discussion with Scarlatti.
  - The purpose of these discussions was to understand which datasets and applications were held by each actor, and draw out their motivations, hesitations and potential enablers to forming a MLDS agreement.
- An analysis using the data sharing collective framework
  - A formal analysis of the information shared through the one-on-one discussions was carried out to identify areas of commonality/discord.
- Documenting observations in a reflexive journal
  - The process of applying the framework was documented as part of undertaking the proof of concept.
  - Over the course of the project, the research team's thoughts and observations were noted in diary form. This brought to light potential opportunities for improvement and the framework was refined accordingly.
- Reflecting back findings and observations
  - The key themes from the analysis were socialised with the collective and iterated as required.
- Drawing out lessons from specific anecdotes that could inform data sharing more generally.

Testing the negotiation mechanism

We had initially anticipated using the findings from the steps described above to inform and test the negotiation mechanism proposed in Phase I. This would have included:

- Drafting an initial agreement using the outcome of the data sharing collective framework.
- Delivering the initial agreement to each actor individually and commencing negotiations.
- Negotiating the agreement through an iterative process, refining and re-presenting the agreement to each actor until an outcome is reached.
- Documenting our application of the negotiation mechanism to identify potential opportunities for refinement, as we did for the application of the data sharing collective framework.

However, during this work it became evident that this was unlikely to be appropriate for the collective. There were two main reasons for this:

- Parallel negotiations were already progressing. Without a formal mandate to act as a facilitator, it was inappropriate for Scarlatti to run a separate negotiation process in tandem.
- During discussions, we started to doubt how effective the highly structured approach that we developed in Phase I would be.

Instead of running a formal negotiation process, we used group discussions to test more specific ideas to incorporate into data sharing negotiations. These included:

- The use of an independent facilitator.
- Role-playing the part of other participants in the negotiation.
- A voting mechanism, proposed in Phase I, in which *actors* vote yes or no to a data sharing proposal.

Current state of the beef genetics data sharing discussion

As of November 2023, the beef genetics case study group had made good progress in their parallel discussions towards data sharing. They felt that there was no need for Scarlatti to develop and present the negotiation mechanism, as this had perhaps happened albeit in a different and more informal manner. They believed this was due to the open-minded nature of the actors, the reasonably small size of the group, and their consensus-driven approach to discussions.

However, they did indicate that there may be merit in adopting such an approach for larger, more complex groups wanting to develop MLDS agreements.

Observations from the case study

Data sharing collective framework: Ideas reinforced

Many of the ideas identified in Phase I of this work were reinforced by the Phase II case study. Some examples are discussed here.

Actors and the data sharing collective

Even a 'simple' data sharing collective contains lots of complexity

Sharing beef genetics data sharing is (arguably) simpler than, say, sharing the many data sets that could contribute to a farm environment plan. In particular, the number of actors involved is comparatively small. Nevertheless, a negotiation between the small number of actors involved in the beef genetics data sharing collective proved capable of generating sufficient issues for the negotiation to outlast this six-month research project.

One of the actors said that until their involvement with this project, and the application of the framework, they were unaware of the depth of complexity within the collective – this was a valuable learning for them.

Adding actors to the collective increases the value of data sharing but adds to complexity.

One idea raised in discussions was bringing meat processors into the data sharing collective to contribute carcass phenotype data to the data sharing collective. The extra data could prove valuable to the members of the collective as well as to commercial farmers (although this wasn't universally agreed). However, involving processors would have created additional complexity and introduced additional questions around their motivations, hesitations and enablers.

Datasets and data flows

Quality and quantity of data held by different actors may vary

In the beef genetics case study, we heard that some breeders/farmers collect all phenotype data at all time points and some DNA test each calf. Others only report pedigree and some weight data. This variability is both within and across breed societies.

Similarly, there is no particular or uniform data standard for describing beef carcasses (although we understand that Australia has some standards for describing carcasses that could be adapted for us in New Zealand).

This variability meant that the value created and received by different actors was uneven, impacting on each actor's motivations and hesitations for data sharing.

Not all actors have the same level of connectivity to data sharing platforms

Some meat processors, for example, have links to FarmIQ – a farm management application and data sharing platform – but this is not universal. Others have their own farmer portals.

Motivations, hesitations and enablers

Hesitations and enablers go beyond issues of data sets and data flow

The main motivation to share data among the beef genetics group was to benefit from better genetic analysis made possible by larger data sets. This appears an almost archetypal case where the benefit of a larger dataset is the reason for the data sharing collective to form.

While barriers, or hesitations, did include those relating to data interoperability and the like, the hesitations that featured most prominently in our discussions typically related to human and organisational issues. For example, for breed societies, the control and management of their breed's genetic data is their raison d'etre. Consequently, if widespread sharing is good collectively, it may represent a threat to their purpose. Experience of similar changes within the sheep breeding industry reinforces this hesitation.

Actors' individual motivations and hesitations may not be in balance.

The potential involvement of meat processors provides an example of where the costs and benefits for individual actors are not well-aligned. Meat processors could potentially contribute considerable value to a beef genetics data sharing collective by making carcase data available. However, this would impose short-term costs such as connecting their systems to a data sharing platform and/or conforming to a standardised way to characterise carcases. In contrast, there are few short-term benefits for meat processors to be had from data sharing (or even long-term benefits in an industry where profitability relies more on throughput than carcase quality.)

Hesitations, motivations and enablers may not be immediately apparent

Some hesitations and enablers relevant to the case study arose for the first time during discussions for our research project. That is, they were not evident enough to have been openly discussed without our intervention, albeit they would likely have come up in time. Two examples were:

- The role of historic organisational behaviours in reducing trust between actors – a hesitation.
- The idea of involving meat processors as an enabler (see above and below).

Data sharing collective framework: Refinements

Applying the framework with the case study provided us with several ideas on how it could be refined.

Actors and the data sharing collective

Parallel discussions should be identified

Rather than one collective discussion, the actors in a data sharing collective are likely to be having side conversations with one other and / or having discussions with organisations outside of the collective on related data sharing issues.

Alongside our case study, parallel discussions were being held between the same actors and including Simmental New Zealand – another breed society – who are not part of the collective.

Data sharing fatigue should be acknowledged

Where discussions about data sharing have been ongoing for some time, there may be a sense of fatigue resulting in a general reluctance to engage.

In the case study, there has been some previous discussion with Silver Fern Farms but that has not progressed – there is a sense of being 'over it' and 'hounded'. This also meant we were unable to engage with them as part of the project.

Decision-making processes should be considered

Individual participants in a data-sharing discussion will rarely have the complete authority to commit their organisation to the terms a data sharing agreement. In the case of the beef genetics collective, each breed society is governed by a board containing people with varying appetite for change. In general, it will be useful to understand:

- To what extent do the people directly involved in the negotiations have influence over their organisations' participation?
- How concentrated or diffuse is the decision-making about data-sharing within each actor?
- How aligned are the different decision makers and influencers within each organisation?
- What will be needed to help each organisation to reach a decision?

Datasets and data flows

Traceability of data along the supply chain should be clarified

In the case study there were questions/uncertainty around the ability for carcass data to be linked to genetics data. A key issue here is the length of the chain of purchase/sale between breeder and processor – animals can be moved from farm to farm several times without pedigree information following.

Interesting, NAIT data which does follow animal movements, is by default not allowed to be used to share information along the supply chain for breeding purposes (or for any purpose other than biosecurity). This is legislated to address privacy and security concerns. However, there is now a tick-box to indicate consent to share the data.

Motivations, hesitations and enablers

Building personal and organisational trust is perhaps the most important enabler to consider

The topic of trust came up several times in the work of our work with the beef genetics collective.

In one of these cases, it came up in a conflict between participants - one participant expressed to another that they didn't trust their intentions. In another example, a participant reflected on past behaviours of one of the actors and noted that this created a level of institutional mistrust albeit changes in people and behaviours had mitigated this.

We observed that some actors were uncertain about others' motivations and what they wanted to do with the data if it were shared. We heard that some of the members of individual breed societies were also hesitant about sharing data because of this perception.

Participants also pointed out the impact that changes of personnel within organisations may have, with fresh relationship-building needing to occur at each point of 'disruption'.

We take from these observations the general lesson that organisational priorities, personalities and historic/present issues of [mis]trust should be recognised and addressed early on in the process, as they may form the basis of many hesitations.

The 'capturability' of perceived benefits should be considered

The case study assumed that there would be demand for Estimated Breeding Values (EBVs) to cross breed, because commercial farmers want to compare across breeds. However:

- The underlying assumption of B+LNZ Genetics that EBVs are demanded by farmers is untested.
- It is unclear that the payment mechanisms exist to fully reward farmers for genetic gain as farmers are rewarded for the quantity but not the quality of their production.
- There is a level of farmer education needed on value of genotype over phenotype – the capability/aspirations element needs clarification (c.f. the assumption of demand – to what extent do farmers want it).

A broader lesson is that a range of factors will constrain the ability for farmers and other actors to capture benefits from data sharing. These constraints should be considered when assessing the value created by data sharing.

Negotiation process: Lessons from the case study

The case study offered provided various lessons to apply to other MLDS discussions.

Talking about data sharing mechanics may slow progress by 'hijacking' the actors' focus

Discussions among members of the collective about the mechanics of data sharing and data interoperability, may have actually *slowed* progress on data sharing. These technical discussions arguably meant that less time was available to address the human and organisational hesitations, that arose in the discussions.

More than one round of individual discussions will be necessary – allow time

While we were not formally playing a facilitator role in this case study, the way that we interacted with the collective did partly resemble this. This provided us some lessons that could be applied to facilitation process more widely.

After one round of individual discussions, we had gained a reasonable understanding of the context of MLDS in the beef genetics industry, and established a level of trust with the actors such that they had begun to disclose their motivations, hesitations and enablers (see the point above about trust). However, this also introduced a number of questions that would have been useful to further our understanding. This indicates that one round of individual discussions is likely to be insufficient and that the process cannot be circumvented.

We raised the idea with the actors of canvassing actors' motivations, hesitations and enablers at the beginning via a survey. They felt that this could potentially work but reinforced the point that whichever method was used to gather this information, it could not be rushed. They believed that the use of an external facilitator was valuable, in that they could ask these questions in such a way that it would be perceived as constructive rather than "picking a fight" – which could be the case if another member of the collective asked them.

Defining a common data sharing language may help

Having a shared understanding and language around data sharing may be useful to shape solutions and avoid misunderstanding. One particular area is in defining and differentiating between the concepts of data ownership, management, control, use, privacy...

There is value in having a formal structure to uncover these points

The case study reinforced our view that using a structured approach to analysing a data sharing collective will be valuable. It allows the facilitator to:

- Focus on the principles for MLDS rather than the mechanics.
- Question individual actors' motivations, hesitations and enablers.

The members of the collectivebelieved that the formal structure of this framework has given them

> *"a chance to think about this from a helicopter point of view rather than from the trenches. [It's been] good to have an outsider's objective perspective."*

Individual, anonymous discussions are beneficial

Most actors in the case study were willing to engage in honest, open conversations knowing that what they said would not be directly conveyed back to the collective.

This degree of anonymity allowed them to say things that might have remained unsaid had the discussions been undertaken in a group situation with the whole collective.

The discussions also helped some actors get clarity on what their desired future looked like, and raised valuable points they may not have considered otherwise.

The facilitator should be an external party

It is valuable for the facilitator not to have 'skin in the game', as this neutrality helps actors to feel comfortable with the process. In this case study we observed participants talking to us about points that they were not willing to discuss directly with other actors.

An external facilitator will likely be a non-subject-matter expert. While that may seem a disadvantage, it potentially helped in this case study in one way. Within the case study there were some conversations that seemed to be contradictory. An example was around the flow of animal data and the ability to link genetic data to carcass phenotype data. Sending back the notes from the one-on-one discussions to each individual allowed them to reflect, fact-check, edit, and add further details. This helped clarify issues and knowledge gaps for the wider collective.

Broad conclusions and possible next steps

We have drawn several broad conclusions drawn from this MLDS Phase II project. The case study has:

- Validated the idea that 'soft' human and organisational factors, and incentives, are as important, or more important, than data sharing mechanisms.
- Reinforced our belief that a formal, structured framework for analysing actors' motivations, hesitations and enablers is useful. At the same time, it has highlighted gaps in our original data sharing collective framework.

- Suggested that we should do more to frame 'soft' human and organisational issues in the data sharing collective framework. In particular, the case study has highlighted the importance of understanding the level of trust between actors as a part of the process[1].
- Reinforced some of our ideas about the negotiation process – such as the value of a facilitator – but challenged our thinking in other areas – such the value of a fixed process to run a negotiation.

We have evolved the outputs from Phase I of this work to reflect these findings in the next section.

[1] During the time that this work was being undertaken, Scarlatti attended a conference in which one of the key speakers was Ivar Ravn, Director of SEGES Innovation in Denmark. Denmark appears to have largely solved the problem of multilateral data sharing in their food and fibre sector. Ivar opened his talk by arguing that the reason why this has been possible is that farmers have a high level of trust in the institutions that serve them. Presumably also, those institutions also have a high level of trust in one another.

Six ideas and 40 questions – Two tools for parties working towards MLDS

This section applies lessons from the case study to refine the two tools proposed in Phase I of the project. We do this by:

- Replacing the proposed negotiation mechanism with six ideas that parties working toward MLDS can draw on to shape the negotiation process that they use to work towards a data sharing agreement.
- Expanding the data sharing analysis framework and structuring it as a hierarchical list of 60 questions.

Tool 1 – Six ideas for parties negotiating MLDS agreements

The ideas presented below are intended as a 'pick-and-mix' selection for parties negotiating data sharing agreements to draw from as fits their particular context.

1. Acknowledge the complexity

The process to reach a multilateral data sharing agreement is an involved negotiation. it requires that multiple actors, with different goals, constraints and resources, work through a process to ensure that technical problems are solved, incentives are aligned, and that human and organisational pitfalls are navigated. In Phase I of this work, we made an analogy to the complexity of multilateral trade negotiations, and we think that analogy still holds true.

By acknowledging the complexity at the start of the process, it is more likely that the parties working towards a data sharing agreement will:

- Build into their negotiation plan the level of structure and formality of process needed to maximise the chance of success. We hope that this will make them more likely to consider using some of the approaches below.
- Budget the time and cost needed.
- Commit the management focus required to ensure that each actor fully engages with the process.

2. Use an independent facilitator

A facilitator could play several roles to help progress a data sharing negotiation. These include:

- Undertake the formal analysis discussed below.
- Develop terminology so that collective have a shared understanding of potentially ambiguous concepts like 'data ownership'.
- Talk with actors individually and collectively to surface issues, and test solutions that could be incorporated into proposals.
- Provide a trusted go-between that parties can talk to about topics that they are not willing or able to bring to the collective as a whole.

- Develop proposals to put to the collective.

Ideally the facilitator would not have a direct interest in the data sharing being discussed. Being skilled in mediation / problem solving / conflict resolution may be more useful than having skills in data / technology or subject matter expertise.

3. Undertake a formal analysis

By formal analysis we mean using the data sharing analysis framework presented below to understand and document:

- The data sharing collective as a whole, including the actors, the purpose, datasets and data flows, and applications.
- The actors individually including the organisational factors that influence their decision-making, and their motivations and hesitations to share data.

Importantly, we propose that the analysis is shared across the data sharing collective. This will help to ensure that all actors have a shared understanding of the issues and opportunities. An assumption here is that the MLDS negotiation process is different to one, say, between businesses negotiating a commercial agreement in that the purpose of the data sharing collective is likely to be for industry and / or national good rather than commercial gain. That said, it is acknowledged that this step may be constrained by the need to avoid sharing anything that actors have disclosed in confidence.

This analysis step may require an iterative process as actors build:

- Trust in other actors and become more willing to share information.
- Internal understanding about, for example, the costs they will experience.

4. Recognise and address human/soft organisational factors upfront

It is tempting to treat MLDS as primarily a data and technology problem. However, the work done across both Phase I and Phase II of this project highlights that the human and soft organisation issues are more likely to be stumbling blocks than data management issues.

These human and soft organisational issues may be current, or they may be based on historical tensions that have generated a level of ongoing discomfort or mistrust amongst present members of a collective. It is therefore fundamental to recognise and address these factors at the beginning of the process, if negotiations are to succeed.

The use of the data sharing analysis framework provides an opportunity for a skilled facilitator to draw out any issues and work to identify a solution. Potentially a facilitator could work with the data sharing collective to develop a set of design principles that guide the data sharing negotiations.

5. Break negotiations into stages

To facilitate progress, it may be useful to formally break a negotiation into distinct stages with milestones and outputs at each stage. For example:

1. A discovery phase in which the collective works though the data sharing analysis framework together.

2. **A trust building phase** in which the collective identifies trust deficits and puts in place solutions to address these. Other human and organisational hesitations could also be identified and addressed at this stage.
3. **An incentivisation phase** in which the costs and benefits for all the actors are discussed, and the negotiation focusses on approaches to ensure that all actors received benefits that at least cover their costs.
4. **A data management phase** in which the parties address the issues of data interoperability, data quality, data ownership and usage, …

6. Consider a 'yes or no' approach

In Part I of this work, we developed a negotiation mechanism in which a facilitator would table a proposed data sharing agreement and offer each actor a confidential vote to be in or out. The catch, however, is that all parties need to vote to be in for the proposal to be agreed.

This mechanism arose in Phase I from the exploration of game theory as a way to analyse data sharing problems. That work identified that actors can be part of negotiations without a serious intent to commit to an agreement and/or actors may not have fully developed their own thinking about the costs and benefits of participating in the agreement. This negotiating mechanism exposes both of these.

While we believe that such a yes/no mechanism may still have merit, we have two suggestions to strengthen its value:

- Include an onboarding brainstorm as part of the formation of the collective. This could identify ways to approach actors who may be reluctant to engage in data sharing but who the remainder of the collective want/need to join in order for them to benefit.
- Identify all actors who should be part of the collective and agree what their role should be. Some may be active participants in the collective as data sharers, others may simply need to be kept informed of progress.

Tool 2 – A data sharing analysis framework in 40 questions

In Phase I of this work, we proposed a data sharing collective framework. The tool is in two parts – a guide to an analysis of the data sharing collective as a whole, and structure to understand each actors' motivations, hesitations and enablers.

In this section we update the tool presented in Phase I by:

- Presenting the framework as a set of questions. These could be asked of actors in several ways, for example, discussed in one-to-one interviews (as we did in the case study), requested using a survey, or debated as a group in a facilitated workshop.
- Incorporating insights from the case study.
- Broadening the analysis of actors' motivations, hesitations, and enablers to cover other actor-specific topics that don't neatly fit under the motivations, hesitations, and enablers headings. These are grouped under the themes:

    1. 'Soft' human and organisational factors.
    2. Costs and benefits.
    3. Datasets and data flows.

    This restructuring is motivated by the suggestion made above that a MLDS negotiation start by analysing and addressing the 'soft' human and organisational factors before going on to address costs and benefits, and only after that address issues like data interoperability.

The list of questions below is unlikely to be exhaustive but can easily be added to or adapted over time. It is unlikely that every question will be relevant to every data sharing collective. As with the six ideas in Tool 1, the questions can be used as a list to 'pick-and-mix' from. We think that simply considering whether a question is relevant will contribute to data sharing discussions.

Part 1 - The data sharing collective as a whole

The questions in this first part relate to the data sharing collective as a whole.

Actors

Forming the data sharing collective

1. Does the model of a data sharing collective apply to the proposed initiative? If not, how does it vary? Is the use of this framework sill appropriate in part?
2. Who are the actors involved in the data sharing collective?
3. What additional actors could be brought into the data sharing collective in the future?
4. What other actors are relevant to the data sharing collective but not directly involved in data sharing negotiations? For example, individual farmers.
5. Who has the power to decide which actors should be included/excluded from the data sharing collective?

Shared purpose

6. How well-aligned are the actors on the purpose of the data sharing collective? What reservations do / could any actors have about the purpose?

7. Could design principles be agreed and documented? Would this clarify how the data sharing collective should work together?

Datasets and data flows

Scope

8. What datasets are relevant to the data sharing collective?

9. What sharing of datasets between actors is being proposed?

Dataset attributes

10. For each dataset relevant to the data sharing collective, what attributes does that dataset have? Relevant attributes could include:

    - ***Domains*** – The term "*domains"* is used to describe broad topics areas such as:
        - Spatial data – like farm maps
        - Financial data – like farm accounts
        - Animal data – like animal health records.

      Each *domain* could be broken into sub-*domains*, sub-sub-*domains* and so on. There is no specific hierarchy of data *domains* proposed here as groups working on *data interoperability or data standards* will be better placed to do this.

    - ***Coverage*** – The term *coverage* relates to the number of data *domains* contained within the *dataset*.

    - ***Subjects*** – This describes the farms, individuals, or organisations whose data are included within the *datasets.* This is of particular importance when Māori are identifiable within a *dataset.*

    - ***Sample size*** – This refers to the number of *subjects* within a *dataset*.

    - ***Completeness*** – A complete *dataset* will contain all the relevant data fields. An incomplete *dataset* will be missing some fields.

    - ***Reliability*** – This refers to the confidence that a user can have that the data are correct. Self-reported data will typically have low *reliability*. Data collected by a third party providing a curation role, or data collected for administrative purposes, may have higher *reliability*.

    - ***Granularity*** – Data about a quantity could be provided at varying levels of aggregation. For example, a dairy farm's production data could be broken down at a daily level or totalled over a year. Or a farm's area could be described with a single value for effective area or characterised as a set of land blocks with each described in detail.

    - ***Sensitivity*** – Information about some *domains*, such as commercial, personal, and financial information, are likely to make a *dataset* more sensitive.

Ownership and control

For each dataset relevant to the data sharing agreement:

11. Who 'owns' the data? Which actor(s) are involved in sharing that data?
12. What mechanisms are available to manage the process of data sharing including permissions, data transfer, storage etc.?
13. Under what terms could data sharing be permitted? For example, a licence to use the dataset for a single application only.

Applications

For each *actor* and for each *dataset*:

14. What application(s) does the actor want to use the dataset for?

Part 2 – Actor-specific analyses

The questions in this second part relate to each actor relevant to the data sharing collective. Importantly, this should include actors that are not directly represented in the data sharing collective in particular, individual farmers.

'Soft' human and organisational factors

Trust-building

15. What level of trust does the actor have in the other actors involved in the data sharing collective?
16. What level of trust do the other actors involved in the data sharing collective have in the actor?
17. What is the cause of any trust issues?
18. What concerns does the actor have about the use of their data? An actor may be concerned around sensitive information within the datasets being accessed by certain actors, and the potential for them to be used in applications beyond their control. An often-cited example is the perceived risk for farmers that shared data will be used, without their consent, to monitor compliance with regulatory requirements.
19. Are any issues serious enough to warrant a formal intervention such as a mediation session? Or is it a simple communication breakdown that could be overcome by an open discussion?
20. What options exist to build trust between the actor and other people and organisations (c.f. building trust in IT systems, data security)?

Prioritisation

21. How important is this data sharing agreement to the actor? Where does this sit in a list of priorities for the relevant people within the actor organisation?
22. What history of involvement in data sharing initiatives does the actor have? What is their appetite to engage in the current data sharing collective?
23. What parallel data sharing initiatives is the actor involved in? How do these relate to the current data sharing collective?

Decision-making processes

24. To what extent do the people representing the actor at a data sharing negotiation have influence over their organisations' participation in a data sharing agreement? That is, are the decision-makers at the table?

25. How concentrated or diffuse is the decision-making about data-sharing within the actor? How many people have a say? How will they reach agreement?

26. How aligned are the different decision-makers and influencers within the actor organisation?

27. What information will be needed to help each actor to reach a decision?

Costs and benefits

Motivations

28. What time and cost savings could the actor achieve by gaining access to a dataset?

29. What time and cost savings could the actor make by not having to provide a dataset?

30. What additional applications become available to the actor by gaining access to a dataset? What value does that create for the actor? What value does that create for other actors?

31. What benefits can the actor gain through access to better information and the use of tools informed by data?

32. What non-data consideration could reduce any of these benefits for the actor? For example, farmers' capability and motivation constrain the utilisation of new tools made available through data sharing. How much do these factors reduce the benefits?

Hesitations

33. What one-off costs will the actor face to participate in a data sharing agreement? For example, the costs to connect or adapt existing IT systems, or the costs to make datasets interoperable.

34. What ongoing costs will the actor face to participate in a data sharing agreement?

35. What threats does data sharing pose to the actors' current activities or purpose?

36. What potential future opportunities for the actor could the actor perceive will be threatened or lost by sharing data? For example, does the actor have ambitions of developing proprietary data sharing tools?

Enablers

37. What additional value would be created for the actor if additional actors and additional datasets were brought into the data sharing collective?

38. Other than providing data, what could actors that stands to gain benefits greater than costs from data sharing (new beneficiaries) offer to actors that face costs greater than benefits by sharing their data (net providers)? Could net beneficiaries make 'transfer payments' to offset costs or otherwise incentivise net providers?

Datasets and data flows

Hesitations

39. What limitations do the attributes of datasets that the actor could gain access to as a result of data sharing pose for that actor? For example, do issues of coverage, sample size, reliability etc. reduce the value of the dataset for the actor's intended application(s)?

Enablers

40. What could be done to 'upgrade' the attributes of a dataset to make it more valuable to other actors? For example, expanding the coverage of a dataset created by a farmer survey by asking farmers questions on additional domains.