# Final report

## Investigating privacy and confidentiality risks in ISC data

Project code: V.ISC.2143

Prepared by: Jonathan Chan, Thilina Ranbaduge, Paul Tyler and Thierry Rakotoarivelo
Commonwealth Scientific and Industrial Research Organisation

Date published: 12 January 2023

PUBLISHED BY
Meat & Livestock Australia Limited
PO Box 1961
NORTH SYDNEY NSW 2059

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

# Executive summary

## Background

Integrity Systems Company (ISC) holds a significant amount of data related to the livestock movements for the Australian red meat industry. The broad concern for ISC is that it is equipped to manage the balance of privacy and confidentiality against the potential use of the data for the benefit of industry. Currently ISC receives data access requests from stakeholders or third-party organisations for access to livestock movement information. ISC responds to data access requests on a case-by-case basis, releasing data that follows strict processes, such as data treatment, permissions, and consultation with stakeholders. Invariably this results in releasing data with low risk. As the processes themselves are used to manage risk, current processes require select members within ISC, with relevant skills and experience, to assess the risk in the data request and authorise the release of the data for its purpose detailed in the data request. The level of processing of the data is designed to reduce risk to suit users and environments where risks in the data may be poorly controlled, whether those risks are likely or not.

ISC acknowledges the data it holds could provide value to industry and that current practices probably restrict the utility and wide use of data released. As such, ISC wanted to better understand the privacy and confidentiality risk in the data so future data releases could be of more benefit to industry. This project aimed to understand the risk in the high value data held by ISC and ways to build capability within ISC to assess the privacy and confidentiality risk more broadly.

## Objectives

The objectives of the project were:

- Identify a high value dataset held by ISC where ISC would benefit from better understanding privacy and confidentiality in that data.
- Assess attributes and combinations of attributes from the high value dataset, that could act as identifiers or quasi-identifiers.
- Assess current ISC data de-identification processes and assess risk level.
- Create a toolbox of processes for ISC to use to assess risk and manage de-identification of their datasets.
- Build internal ISC capability around assessing risk in data and environments so they can employ those techniques in the processes of assessing data sharing risks and de-identification procedures.
- Build internal ISC capability in the processes required to safely share data to minimise risk of identification of individuals and/or businesses to a suitable level for ISC.

## Methodology

The project was structured in 5 stages, which is reflected in the structure of the final report. The stages were as follows:

- Identify a target dataset of high value with potential privacy and confidentiality risk, especially when considering re-identification risk.
- Analyse re-identification risk from attributes and combinations of attributes.

- Examine key environments in which the data does or may (hypothetically in the future) reside and how risk is controlled (or not) by environment controls.
- Consider controlling and mitigating risk through various data treatments.
- Understanding risk appetite and assessing combined risk from data and environment as it might emerge for ISC.
- Produce a toolbox of processes and worksheets to assist the assessment of risk for proposed data release.

As the project team worked through this process, ISC staff were introduced to key concepts and stepped through processes in a series of 6 workshops of 2 hours each, and the concepts of the toolbox were introduced. The final workshop worked through the toolbox with a fictious scenario.

## Results/key findings

The data attributes and table formats that were agreed for consideration by CSIRO were found to have definite paths for re-identification. Most of the data is not directly about people. However, people are closely associated with the activities recorded in the data, posing some potential issues of privacy, without involving data specifically about people. As the data is specifically about animal movement, it is intrinsically related to the activities of businesses and so pose confidentiality issues. In summary regarding the data:

- The data contains some personal data.
- People activities might be inferred.
- Confidential business information is likely contained in the data.
- A business is identifiable, not just by identifiers, but by records or combinations of records.

Release data is reported to be highly aggregated, to a point where most of this risk is mitigated. ISC processes are generally thought to mitigate risk.

Data environments were considered and generally found appropriate. High risk data was reported to be in systems that provide strong controls. Generalised and aggregated data is more likely to be used within ISC and further treated data released to less controlled environments as part of data requests.

As part of this work, ISC staff were trained in additional capabilities around assessing data and environment risk that would enable ISC to have more flexibility in the kind of data they may provide, especially where the destination environment is secure and well controlled.

Several other key findings include:

- Internally, risk is controlled by different access. This appears (on the surface) to be appropriately set (though a full examination has not occurred).
- Data release beyond ISC involves a complex process, partly driven by stakeholder groups and ISC policy.
- There is an underlying conservative approach to data release, driven by ISC obligations as custodians of industry data and stakeholder (producer) expectations associated with the sensitivities of animal movement and related business activities.

- There appear to be significant amounts of information available outside ISC (i.e., on publicly available website) that could be used as background information (The scale and risk posed by this may need further exploration to better understand how this impacts risk).

## Benefits to industry

The work has identified that ISC's processes for releasing data are conservative. This reflects an uncertainty over privacy and confidentiality risk in the broader ecosystem. It also impacts what is considered safe to release in particular situations. This likely acts as a barrier to data sharing of higher utility data. If the perceived issues of privacy and confidentiality are true, then the conservative approach is likely appropriate. However, this comes at a cost of preventing further insights into the broader industry that may provide significant benefits. To enable potential greater benefits to industry from richer data sharing, it would be valuable to develop a better understanding of why the current barriers exist and determine if there are solid justifications for those barriers. This would provide for more informed decision making. This work seems to suggest that there is room to move in providing richer data to more people.

## Future research and recommendations

This report suggests several areas ISC/MLA could continue to explore. These include:

- Further investigation to understand how the corporate structures and obligations of ISC/MLA might be driving current data practices.
- Research stakeholder attitudes to data sharing and the sources of any concerns.
- Keeping a current catalogue of publicly available data in areas that overlap or are adjacent to data held by ISC/MLA.
- Consider developing a set of well designed (appropriate treatment with known risk) “typical” releases that might serve multiple purposes for common data requests that could be pre-approved for release to particular kinds of environments (e.g., universities).