Document 1

	Checklist for dealing with Administrative Data		
	Dataset Details	**Response: Processor to Grower consignment note.**	**Response: Industry Aggraded Data**
1	Why was it collected?	Processor to grower. Data collected would be used for substitution of correctly collected surveys forms or used for confrontation. Data provided by processors is in form of a consignment note supplied to the growers	Macadamia Industry collected data. Each processor has agreed to provide a 3rd party with aggregated data on prices paid and quantity nut processed. This is used and published for industry to use and evaluate who the industry is going. The 3rd Part aggregated data covers 90% of the Australian crop. AMHA add the addition 8% via phone conversations and estimates the final 2%. When published, this data represents 100% of the macadamia nut production on an calendar year.
2	How was it collected?	Data collected via feedback from processors to growers on quantity, cost etc. of the macadamia nut processed. This would also include ABN, growers address etc.	Processors send aggregated data to 3rd party, 90%. AMHA add reminder. The data provided in an aggregated total of prices paid to growers and total macadamia nut processed year ending Dec.
3	How is it used and why?	The processors use this information as a invoice for growers macadamia nut processed	The Macadamia Nut industry, processors and growers use this data to estimate how the industry is going. Academics use this data to create benchmarking for the industry. Qld Ag. See [redacted] for Benchmarking paper.
4	Look at questionnaire or form	Processors provide a consignment note to growers, mainly on a monthly basis. At this stage it's a form. A copy of the consignment note is available [redacted]	Data is available from the AMS web site. Visit http://australian-macadamias.org/industry/about-aussie-macadamias/stats
	Relevance		
5	What is the scope and coverage?	This is an invoice to the grower that covers what has been processed. Accounts are settled around the 23 of each month. Final yearly accounts are steeled by the 23 Dec each year.	the scope of the final data set represents 100% macadamia nut production. Included GVP.
6	How is geography recorded and to what does it relate?	To the grower level. Should contain business name, location, ABN etc.	At this stange it only at Australia level.
7	What is the collection unit? Is there more than one type?	Business	Australian Industry level only
8	Was there changes in the collection unit over time?	Processors should have , as a matter TAX regulation a copy of invoicing for a number of years	Time series good for about 5 years worth, data set over that period less accruate.
	Timeliness		
9	What is the time between when the data is collected and when it's available for use?	Data available monthly, about 23/24 of each month	Data is published data November each year.
10	Is there a delay? Why?	no, not at this time	no that can be seen.
	Coherence		
11	Administrative changes? What impact?	none	non at the moment.
12	Reference period used?	collected monthly	Calendar year
13	Data item definitions, classifications used?	Total production of processed nuts at a 10% water content. Also provides grower and processor names. Grower information available, see consignment note.	total production of macadamia nuts, average price paid at 10% moisture. Total Australian.
	Accuracy		
14	How is administrative data updated and how often?	As this is a consignment note, it is issues once the processor has processed the nuts and payment to the grower.	total production of macadamia nut collected by each processor as a final figure produced for that season.
15	Missing data? What proportion?	non at this stage	90% is collected direct from processors, 8% phone collected by AMHA and 2% is estimated. Coverage is 100% is Australian production of macadamia nuts.
16	Has data been adjusted to account for missing values?	none at this stage	2% is eastimated.
17	Does the accuracy of the records vary for different information/units collected?	it may as different processors use different systems to produce his information. More to examine	No, data published is simialr to data published by ABS
	Accessibility		
18	Is the dataset made publicly available?	No, would have to collect under the Stats Act.	have to check is currect
19	Which format is it available in?	various, not specified atm.	available from AMS web site
	Interpretability		
20	Is there information available to help users understand the administrative data?	should be able to provide a concordance for data items supplied.	non yet
	Institutional Environment		
21	Is there an agreement, legislation or authority that the data was collected under?	Letter of exchange is currently being produced. Will have to talk to processors and then send letter once an agreement has been reached.	data availabel from web site. Have to check if currect year available, ie 2016 data available Nov 2016.
22	Are there factors that may impact the quality of the data provided by the agency?	The way the data is formatted, send to ABS, how it is compiled form the processor.	weather, system issues, 3rd part probles aggregating and compiling data. Indutry is relient of data and issues shodl be addressed well before publication. Data can also be confronted against levies data.
23	Do they have risk management processes in place?	should not be an issue as it's a legal requirement to keep records of business transactions. Data showing growers personal information maybe an issues, one that will have to be considered for each processor.	Not know yet
24	How are errors/risks dealt with?	unknown at this stage	Not know yet

1 -24 Sourced from Methodology in corporate manuals and ABS Data Quality Framework

Document 2

Australian Bureau of Statistics

Canberra Office
ABS House
45 Benjamin Way
Belconnen ACT 2617
Phone 1300 135 070

Locked Bag 10
Belconnen ACT 2616

www.abs.gov.au
ABN 26 331 428 522

Mr S47F

Thank you for meeting with S47E, S47F as part of the Horticulture Innovation Australia (HIA) Macadamia Nut 'Gross Value of Production' project (HIGVP) for the Australian Macadamia Nut Industry and the Australian Bureau of Statistics (ABS) alternative data source project.

As an outcome of this meeting, I forward this letter of exchange between the ABS and Macadamia Nut Processing Co (MPC) that represents an agreement for MPC to supply data contained in the *'Annual Grower Summary Report' to the ABS*. The data required is at the grower (unit record), these are outlined in *Table 1.*

These data will be used by the ABS as part of the HIGVP and the ABS alternative data source project. The data will be utilised for confrontation and analysis for the GVP and for internal analytical processes. The ABS will securely store these data provided by MPC on a restricted access basis to ABS officers directly engaged in using these data for the production of ABS statistics.

As agreed with S47F MPC will supply 2 years of individual annual grower report data for a minimum of 50 growers. ABS will provide MPC with a secure process to transfer grower (unit record) data into the ABS.

The ABS operates under the *Census and Statistics Act 1905* (Cth) (the Act) which guarantees data held by the ABS will be kept confidential. The Act legally binds all ABS staff to never release personal information to any individual or organisation outside the ABS, including other government agencies. Section 19 of the Act makes it an offence to divulge any information collected under the Act. The ABS must also comply with the *Privacy Act 1988* (Cth), and the Australian Privacy Principles contained within.

The ABS will securely store these Data provided by MPC on a restricted access basis to ABS officers directly engaged in using the Data for the production of ABS statistics.

If the ABS becomes aware that these Data may have been lost, stolen, accessed or used in an inappropriate manner, the ABS must:

a) immediately notify MPC in writing, giving details of the suspected or possible breach;

b) do everything necessary to remedy the unauthorised access to, use or disclosure of the Data or to prevent the suspected or possible breach of this clause;

c) give MPC all assistance required in connection with proceedings which MPC may institute for breach of confidence or otherwise.

I have forwarded two copies to be signed, one for MPC records and one to be returned in the replied paid envelope to the ABS.

Australian Bureau of Statistics

Canberra Office
ABS House
45 Benjamin Way
Belconnen ACT 2617
Phone 1300 135 070

Locked Bag 10
Belconnen ACT 2616

www.abs.gov.au
ABN 26 331 428 522

Signature: .....

S47F

Date: 4/5/2017

16 May 2017

Lisa Wardlaw-Kelly
Head of Office (Hobart) and Program Manager
Environment and Agriculture Statistics Branch
Industry Statistics Division
Australian Bureau of Statistics
(P) (03) 6222 5802 (F) (03) 6222 5824
(E) **lisa.wardlaw-kelly@abs.gov.au**
(W) **www.abs.gov.au**

Mr S47F
Macadamia Processing Co (MPC)
Chair of the Australian Macadamia Handlers Association (AMHA)
Office: S47F Mobile S47F
S47F

Table 1.

(MPC) Labels from Annual Grower Summary Report

Grower	Heavy Germination
ABN	Heavy Immaturity
Delivery Date	Heavy Discolouration
Consignment Number	Mould
Variety	Insect
Delivery Weight NIS (Wet)	Internal Discolouration
Delivery Weight 10% Moisture	Notional Premium Value $/kg
Moisture %	Notional Commercial Value $/kg
Premium KR %	Notional Reject Value $/kg
Commercial KR %	Freight Subsidy $/kg
Reject KR %	Additional NIS payment (above notional)
Shell %	Total Final $/kg
Light Germination	Consignment Value
Light Immaturity	Industry Levy
Light Discolouration	

ABS visit confirmation . Macadamia

S47E, S47F to: S47F

27/02/2017 02:59 PM

Cc: S47E, S47F

Bcc: S47E, S47F

This message is digitally signed.

Afternoor S47F

Thanks for taking me through a few things yesterday over the phone this morning.

We would like to vist you on the Wednesday 15th and Thursday morning of the 16 March 2017 in Lismore to visit 2 Macadamia Processors and 4 Macadamia Growers. It we are able to do 4 visits on Wednesday 15 March 2017, of about an hour each and 2 visits on Thursday morning 16th March 2017, again, for a hour each.

S22

Could you supply me with the company names and locations of the growers and processors that we will be visiting

Thanks again for organising this. Below are general questions we would like to ask, but depending on the answers, may prompt up to enquire with others.

General guide for **Growers**

- weight of product to processor vs weight of production at the farm, (((if grower was given an ABS agricultural census form, with a 30 June end date.))).
- if the weight of product leaving farm gate is known
- actual weight via feedback from processor (((vers the estimates for ABS surveys, likely variants .))
- losses between farm gate and processor, ie on farm sale of macadamia nuts, other losses
- on farm growing and processing vs pure processors
- when levies due to be paid, impact of timing of harvesting and how collected. grower pays or processor pays for grower?
- If any other surveys/government organisations or research development corporations have asked about pesticide & fertiliser application, land management practices, water use and so on.

General guide for **Processors**

- weight of product to processor vs weight of production at the farm, (((if processor was given an ABS agricultural form, with a 30 June end date.))).
- if the weight of commodity leaving farm gate is known, and the same as the weight the processor ? what are the main differences
- when levies due to be paid, impact of timing of harvesting and how collected. grower pays or processor pays for grower?
- how many growers below to this processor.
- are there growers that change processors
- payments to farmers splits across financial year, harvested on fin year, payments in next fin year? levies payments
- what level of data is kept by producers, weights/prices/location/grower/water content/levy payment/
- husk residue? what happens to it
- If any other surveys/government organisations or research development corporations have asked

about pesticide & fertiliser application, land management practices, water use and so on.

Australian Bureau of Statistics contacts:

S47E, S47F (M S22) S22 @abs.gov.au

S47E, S47F (M S22 S22 @abs.gov.au

Macadamia contact details:

Mr S47F

Macadamia Processing Co (MPC)

Chair of the Australian Macadamia Handlers Association (AMHA)

Office: S47F

S47F

cheers

S47E, S47F

Agriculture Admin Data Accounts

Accounts Development | Agriculture and Environment Branch | **Australian Bureau of Statistics**

(P) S22 **(F)** S22

(E) S22 **@abs.gov.au** **(W)** **www.abs.gov.au**

Australian Bureau of Statistics

Agricultural Administrative Data Initiative Timeline

November 2016- February 2017

Sprint 1:
What were our findings?

- Researched and gained industry knowledge to identify potential data source
- Entered project information and correspondence into JIRA
- Analysed aggregated data from industry, negotiated with industry some sourced from industry websites comparing with ABARES and ABS data
- Investigated and identified commodities to progress
- Spoke with Agriculture BSC to find out past engagement contacts and engagement history
- Organised and initiated contact with industry
- Developed an Engagement spreadsheet to collate and compile the contacts and outcomes from each industry body meetings

February 2017- April 2017

Sprint 2:
Stakeholder Engagement Strategies

- Initiated contact with Industries to arrange meetings via phone and email to obtain data, metadata, industry aggregate data etc.
- Analysed aggregated data from industry, negotiated with industry some sourced from industry websites comparing with ABARES and ABS data
- Created documentation for contact. Introductory and follow up emails
- Provided industry bodies with engagement letter, reminder letter and agenda for meetings
- Meeting with ADAU to discuss docking
- Meeting with AIS/C re ADMIN data
- Identified industries who could provide data to the ABS
- Complied outcomes and industry concerns from meetings to the Engagement spreadsheet
- Travelled and met with industry representatives (Macadamias, Citrus, and Cane Growers etc.)
- Developed and tested the Data Assessment Tool to investigate the quality of the administrative dataset

April 2017- May 2017

Sprint 3:
Final Phase

- Prioritising potential industry data to be integrated within the ABS. Prioritised commodities to move to the next stage i.e.: better, best, awesome,
- Provide base industry knowledge for the ABS to target next stage of admin data collection, where to from July 2017?
- Negotiated and analysed aggregated data received from industry. Some sourced directly from industry websites comparing with ABARES and ABS data
- Prioritising potential industry data to be integrated and/or used for comparisons within the ABS
- Created Letter of Exchange DAPL and ADAU input
- Docked and received Macadamia data ADAU
- Using the data assessment tool to analyse whether the data was fit for purpose
- Findings of each commodity outcomes to be included in final report
- Report writing

May 2017- June 2017

Evaluation

- Final report write up
- Commodity selection
- Team reviews
- Ag BSC discussion
- Final roundup

MACADAMIAS GVP PROJECT

S22

ADI levies pilot	• The Pilot explored whether data from nut processors may be a viable alternative data source for improving the quality of estimates of the Gross Value of Macadamia Production (GVP). • The ABS utilised data from nut processors with the intent of: ▪ reducing survey burden on growers ("collect once, use many times") ▪ maximising the utility of data already collected by processors ▪ improving the re-use of Australian Government and private administrative datasets ▪ considering methods for broader use of administrative data sources within the agricultural statistical system, including as a component of a dynamic dataset.
Findings	• The Pilot concluded that the collection of solely commodity data from 5-10 commodity groups has a minimal effect on the number of growers allocated in sample survey designs. • The study suggested that subsequent investigations: ▪ trial the use of levy reports from all macadamia processors to generate data through a full processor collection of production and value, with business/location links (ABN/PIC) ▪ extend an opportunity to other horticultural commodities to participate ▪ include tree numbers from benchmarking reports ▪ incorporate data generated by local government rates calculation processes ▪ specifically address data governance and ethical issues • A key recommendation is that manual levy receipts reporting and Levies Online submission be standardised to collect a minimum set of metadata, including a location identifier
Phase 2 pilot outline	• There are two objectives: 1. Assess levy consignment notes from all macadamia processors, and other horticultural commodities, to demonstrate the wider applicability of levies databases to the statistical system. Almonds and avocadoes are motivated to participate. 2. Validate the observations from the first pilot.

Future work streams	• **Stakeholder engagement**: advise and consult on findings and next steps; partner with horticulture industry bodies • **Resourcing**: identify/develop resources and project plans • **Extend Phase 1**: continue research on linking grower location to processor based data; influence developments in the levy system • **Phase 2 research**: identify, capture and assess the potential of select alternative data sources through industry/commodity business as usual processes (ie levies); evaluate and report.
Draft timeline	• **November 2017:** Approach organisations to identify data needs and potential sources of information; agree to a joint project plan (utilising existing Hort Industry contacts for Macadamias, and new contacts for other commodities such as Avocadoes and Almonds). • **February 2018:** Assessment of potential data sources for extending Macadamia's pilot (e.g. additional grower's data) and other interested industries. • **March 2018:** Agreement from stakeholders on priorities. • **April 2018:** Data sources secured and ready for evaluation. • **June 2018:** Report outlining learnings and recommendations for future work (Note that this project may extend over a two years subject to funding or ASL provisions.)
Partners	• **Commonwealth government:** ABS, DAWR. • **Industry representative bodies:** Hort Innovation, GRDC, Avocadoes Australia, Almonds Australia, Australian Macadamia Society

S22

ABS – ABARES BILATERAL
ABARES, 44 Mort Street, Braddon, ACT
Wednesday 01 November 2017: 11am – 4:30pm

AGRICULTURAL STATISTICS PRIORITIES WORKSHOP

Wednesday 01 November 2017, 11:00am – 4:30pm

Australian Bureau of Agricultural Resource Economics and Sciences (ABARES)
44 Mort Street, Braddon, ACT

Attendees: S47F *(Chair),* S47F, S47F, S47F, S47F, S47F *[ABARES]*
Jacky Hodges, Lisa Wardlaw-Kelly, S47E, S47F, S47E, S47F *and* S47E, S47F *[ABS]*

PURPOSE of MEETING

This meeting intended to:

S22

2. discuss ABARES and ABS projects to improve agricultural statistics,
3. agree on priorities for joint work plan, and
4. discuss a joint course of action.

S22

S22

S22

AGENDA ITEM 4 (was 5): Presentation – ABS priorities to improve agricultural statistics

- Lisa walked through a presentation covering the context for the Agricultural Administrative Data Initiative (AADI), project work undertaken and planned, and lessons learned to date. Key observations included:

4. AADI team researched 23 commodities, worked with 10 for further analysis, and is planning to work with 4 in more detail, including macadamias and dairy.

- Lisa described the **Macadamias pilot**. ABS Gross Value of Production data consistently differs by about 19% from industry estimates. The industry supported ABS acquiring data from macadamia processors (levy recepts/levy consignment notes) under the *Census and Statistics Act* to analyse estimates in more detail.

- Processors collect information through the levies system that has the potential to meet ABS needs – eg volume, weight, quality assessment, water use. If designed well, macadamia levies receipt data could replace collected data without providing additional data points.

- The Pilot proved the concept by successfully partnering with industry, with ABS acquiring data using its new infrastructure and processes. Assessing macadamia levies data indicated that ABS may not need to rely on direct collection from growers, and could continually improve their growers frame using levies data, however a number of issues would need to be resolved before this could be tested at scale.

- Next steps are to extend the pilot to all processors to generate statistics for production and value, linking ABN and farm location. ABS is keen to influence inclusion of location information – eg Property Identification Code (PIC) or similar – that can link levies data to farm location. (Incidentally, ABS met with GRDC in late October to discuss this.)

S22

S47E

S47E

S22

AGENDA ITEM 6: Finalise joint work plan

- Four alternative data source evaluation projects on the Joint Work Plan had been agreed by the meeting – S22 macadamias/avocadoes pilot S22

S22

AGENDA ITEM 7: Discussion between ABARES and ABS project teams on implementing the joint course of action

S22

- The four pilot administrative data acquisition/assessment projects were endorsed as presented. S22

- The general approach agreed was fivefold:
    1. collaborate on papers being developed for AgSOC **(ACTION 7)**
    2. write to AgSOC in the next two weeks to advise on the supported pilots, request other data in these and other domains for assessment **(ACTION 8)**
    3. ABS and ABARES to collaborate to identify core data needs and opportunities to access and assess administrative data **(ACTION 9)**
    4. ABS and ABARES to collaborate to develop metadata standards for core data **(ACTION 10)**
    5. set up meetings with pilot collaborators in identified motivated industries, starting with S22 S22 Hort Innovation, the Australian Macadamia Society, S22 **(ACTION 11)**.

S22

ACTIONS

7. ABARES and ABS: Collaborate on papers being developed for AgSOC
8. ABARES: Write to AgSOC in the next two weeks to advise on the supported Joint Work Program pilots, and request identification of other data and needs in these and other domains for prioritisation and assessment.
9. ABS and ABARES: Collaborate to assess core data needs and opportunities to access and assess administrative data identified through action 8
10. ABS and ABARES: Collaborate to develop metadata standards for core data

11. ABARES: Set up meetings with pilot collaborators in identified motivated industries, starting with S22 Hort Innovation, the Australian Macadamia Society, and S22

S22

S22

ABS – ABARES BILATERAL
ABARES, 44 Mort Street, Braddon, ACT
Wednesday 01 November 2017: 11am – 4:30pm

AGRICULTURAL STATISTICS PRIORITIES WORKSHOP

Wednesday 01 November 2017, 11:00am – 4:30pm

Australian Bureau of Agricultural Resource Economics and Sciences (ABARES)
44 Mort Street, Braddon, ACT

Attendees: S47F *(Chair),* S47F, S47F, S47F, S47F S47F *[ABARES]*
Jacky Hodges, Lisa Wardlaw-Kelly, S47E, S47F, S47E, S47F *and* S47E, S47F *[ABS]*

PURPOSE of MEETING

This meeting intended to:

S22

2. discuss ABARES and ABS projects to improve agricultural statistics,
3. agree on priorities for joint work plan, and
4. discuss a joint course of action.

ACTIONS

S22

7. ABARES and ABS: Collaborate on papers being developed for AgSoc (ACTION 8)
8. ABARES: Write to AgSoc in the next two weeks to advise on the supported Joint Work Program pilots, and request identification of other data and needs in these and other domains for prioritisation and assessment.
9. ABS and ABARES: Collaborate to assess core data needs and opportunities to access and assess administrative data identified through action 8 (ACTION 9)
10. ABS and ABARES: Collaborate to develop metadata standards for core data

11. ABARES: Set up meetings with pilot collaborators in identified motivated industries, starting with S22, Hort Innovation, the Australian Macadamia Society, and S22

S22

S22

S22

S22

S22

S22

S22

Macadamias

Results from Pilot Study of Macadamias

- Both ABS and Macadamia industry used a very similar 'average price per kgs' for calculation of GVP.
- Main difference in the way the estimate for volume is collected and derived.

Confidence in ABS

	2013-14	2014-15	2015-16
REACS/ Ag Census Production	31,840,639.00	39,783,457.00	38,362,801.50
VACP Nuts Received	17,403,000	36,045,830	40,150,100
VACP Processor Count	4	13	14
AMHA (Calendar Year)	35,200,000.00	43,600,000.00	48,300,000.00
Published Count of Providers (with production)	776	838	549
Diffence in Numbers of Growers	9.4%	7.9%	-34.5%

Nut in Husk (NIH) Nut in Shell (NIS) Kernel

Macadamias key issues

- Processor data seemed more accurate than self reporting on the Census
- Portability of produce was a measurement issue
- Levies data relatively poor quality compared to ABS and AMHA
- Legislation fairly straightforward to incorporate data into ABS
- Use of ABN as a integrator produced some odd results
- Removing commodity information from ABS survey forms had no impact on reducing burden (in aggregate).

JWP 1.1 proposal – levy and related data

- Expand macadamia pilot to rest of industry for the purposes of including results for the calculation of GVP

S22

- Test the readiness of Horticulture industry to move to a 'HortBase' style of data collection.

Administrative Data Initiative results

- Analysed 23 commodities
- Secured 10 for further analysis
- Developed partnerships with 4 for detailed review
  - Macadamias (unit record, levy based)

S22

S22

Agriculture Round Table

ABS House, Belconnen, Canberra

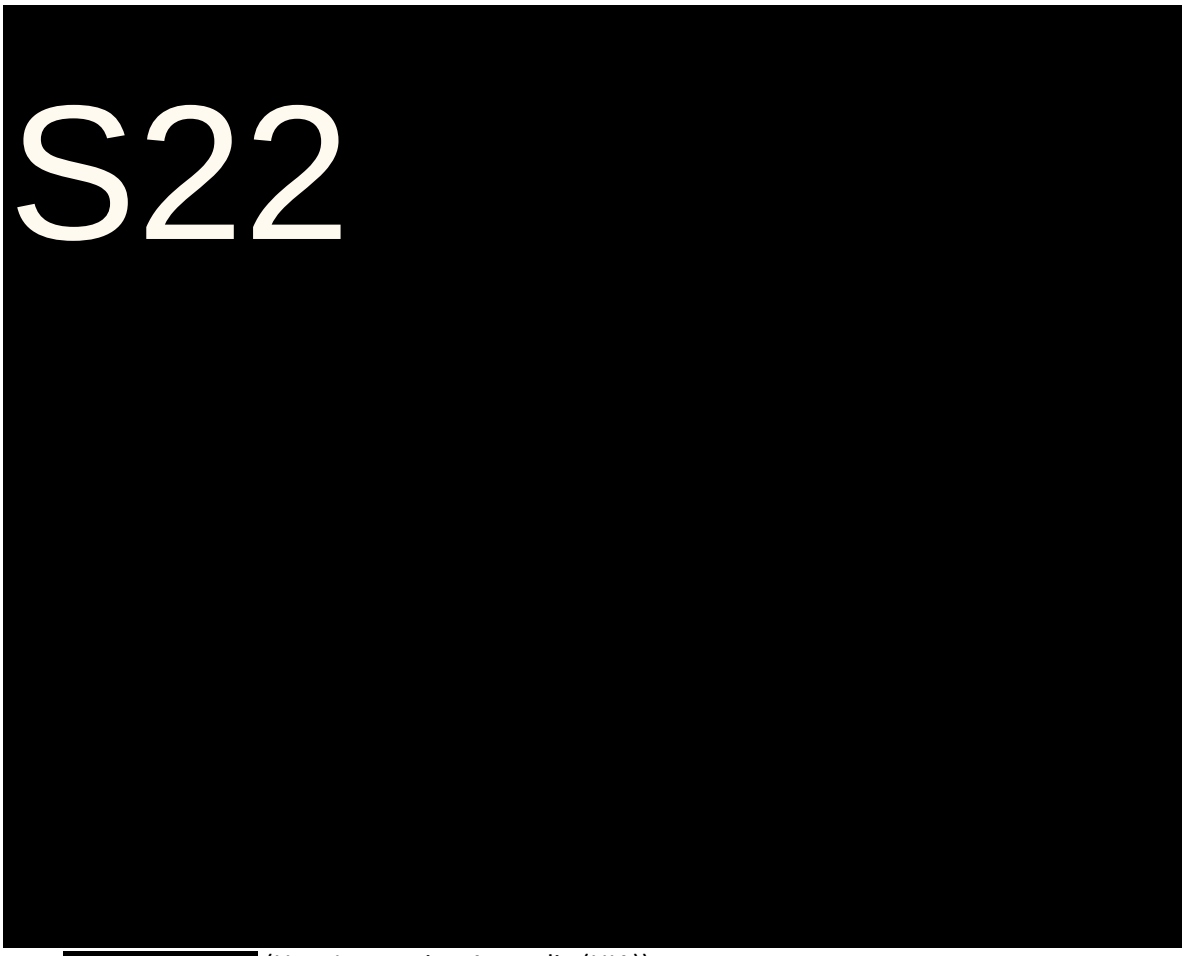
22 February, 2018

9.00am – 12.30pm Morning Session

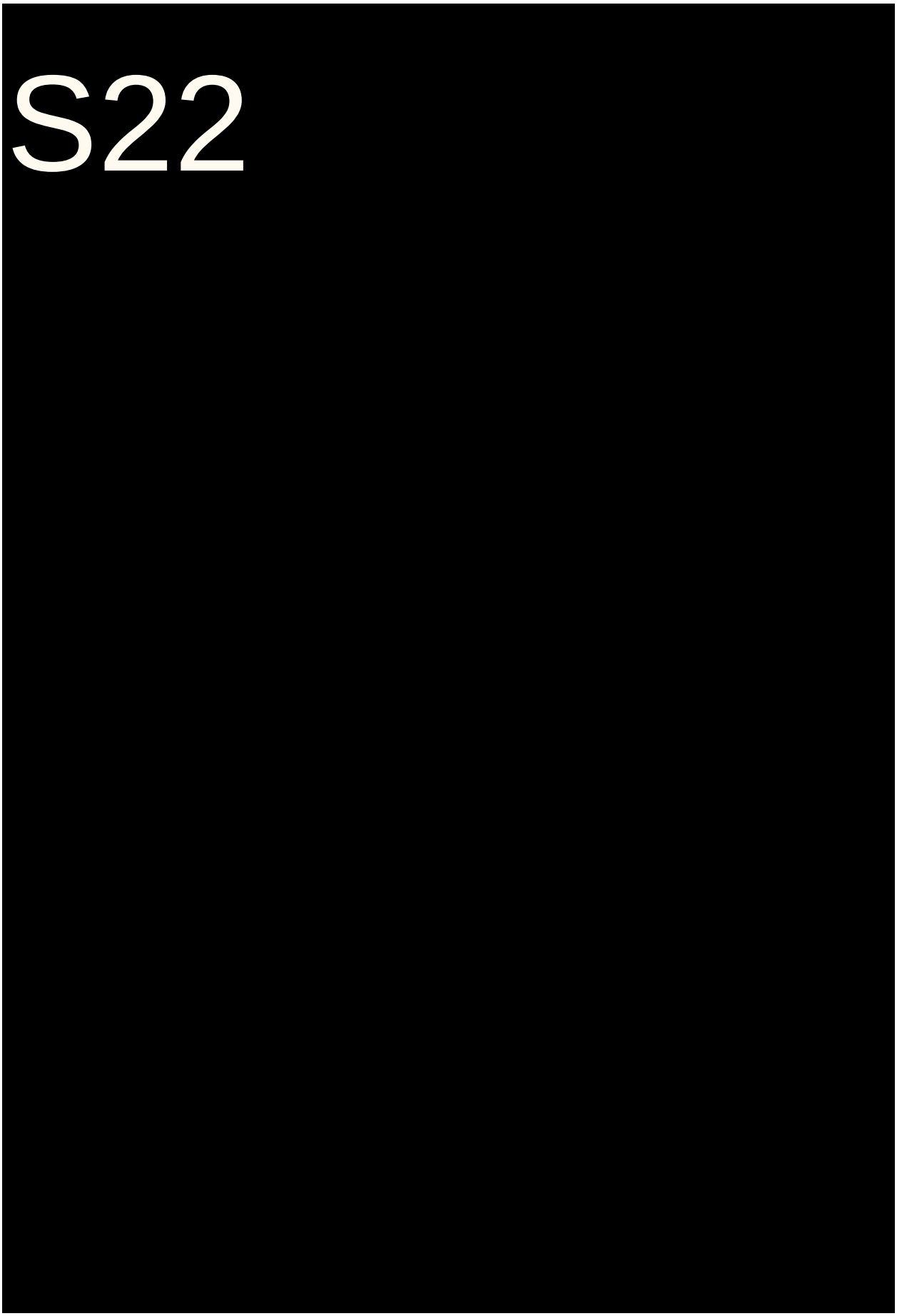
Purpose: Around 35 key stakeholders from across government and industry came together to work on opportunities to advance the agricultural statistics transformation agenda. Recent developments – particularly collaborative work by ABS and ABARES with industry, and the Precision to Decision (P2D) projects were discussed, and lessons shared. Participants reflected on these. They also discussed emerging data needs, and new initiatives and opportunities relevant to their industries.

9.00am – 9.10am Introductions

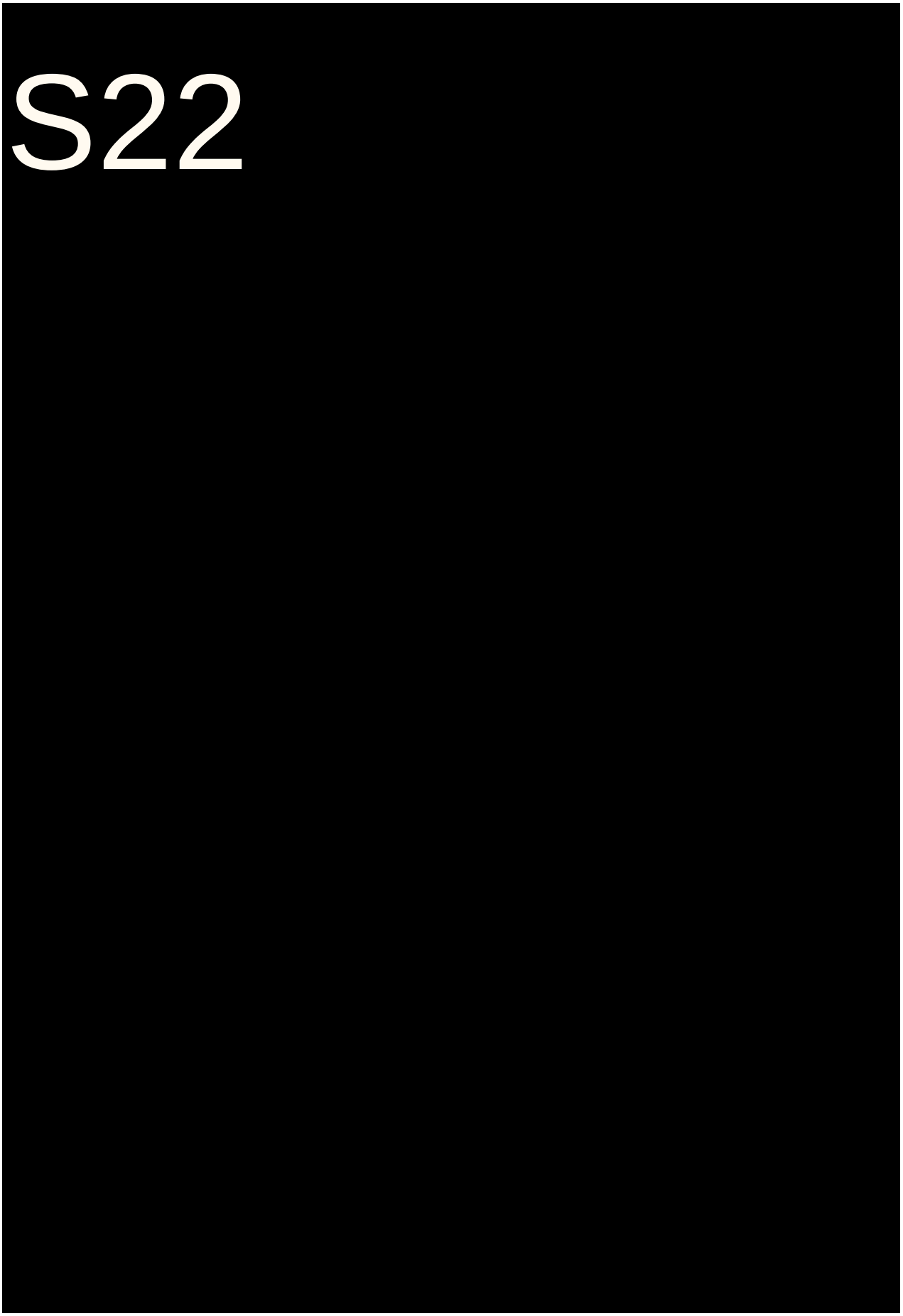
Attendees



- S47F (Hort Innovation Australia (HIA))


S22

S22

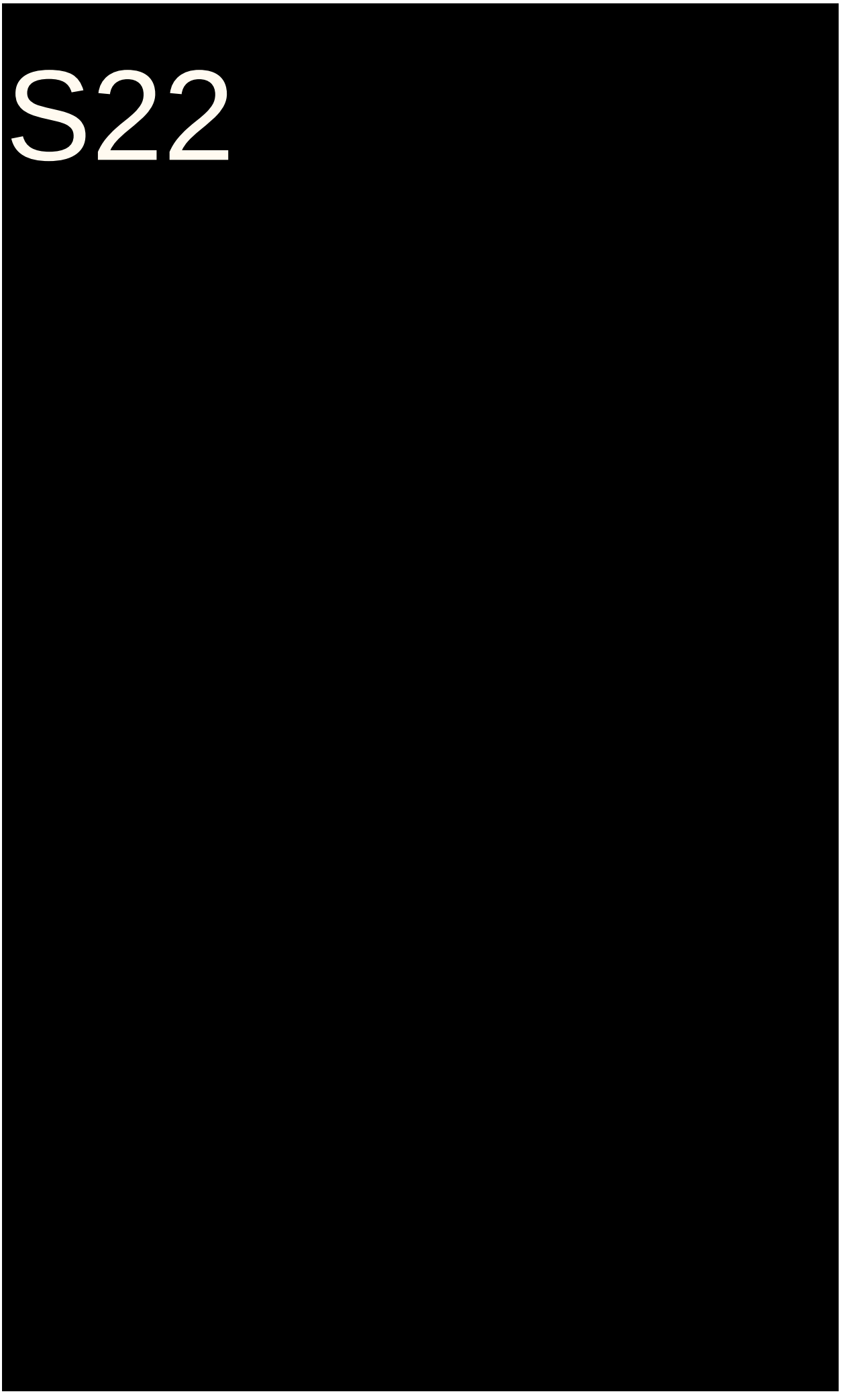

S22

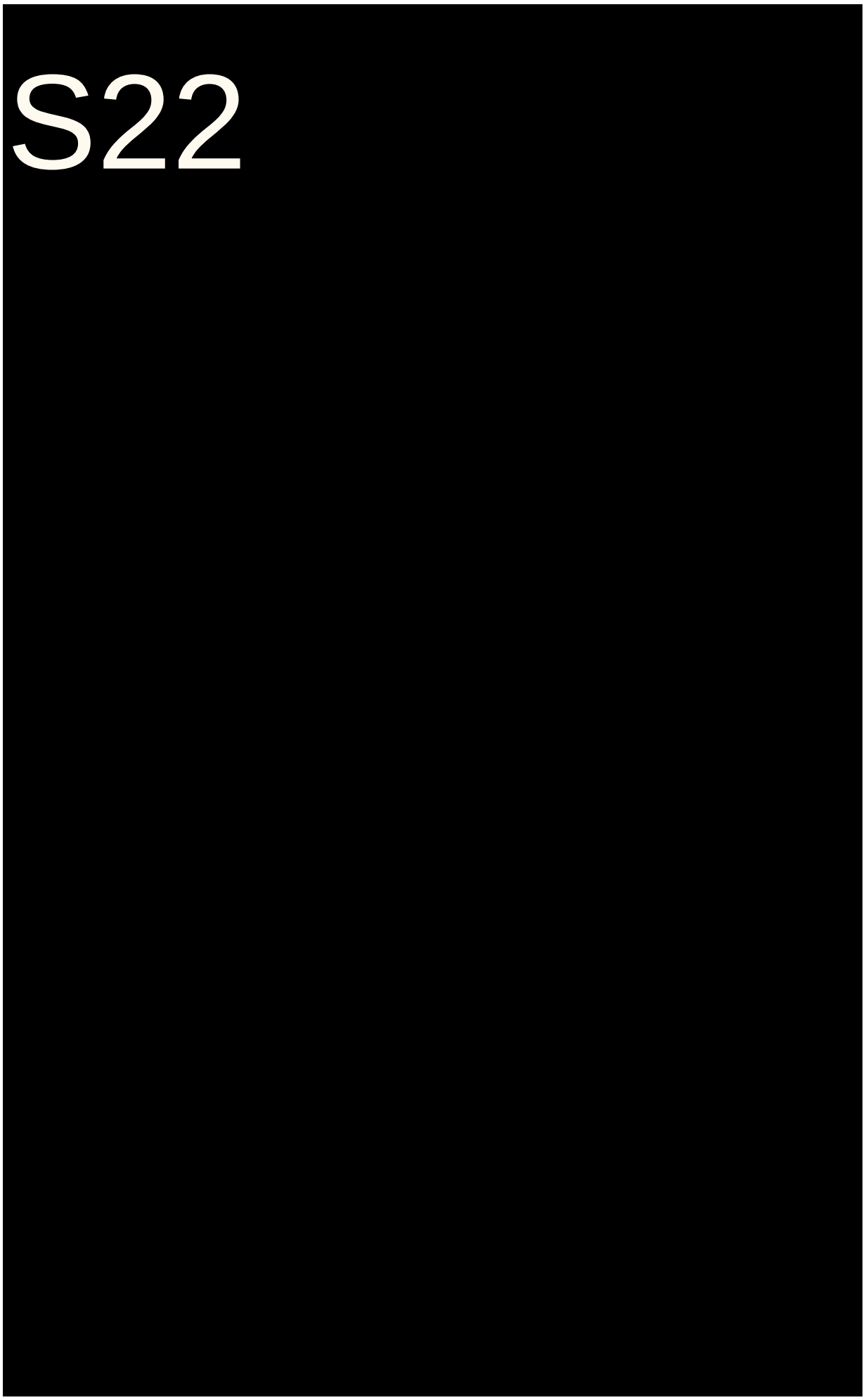
10.15am – 12.15am Information and Updates

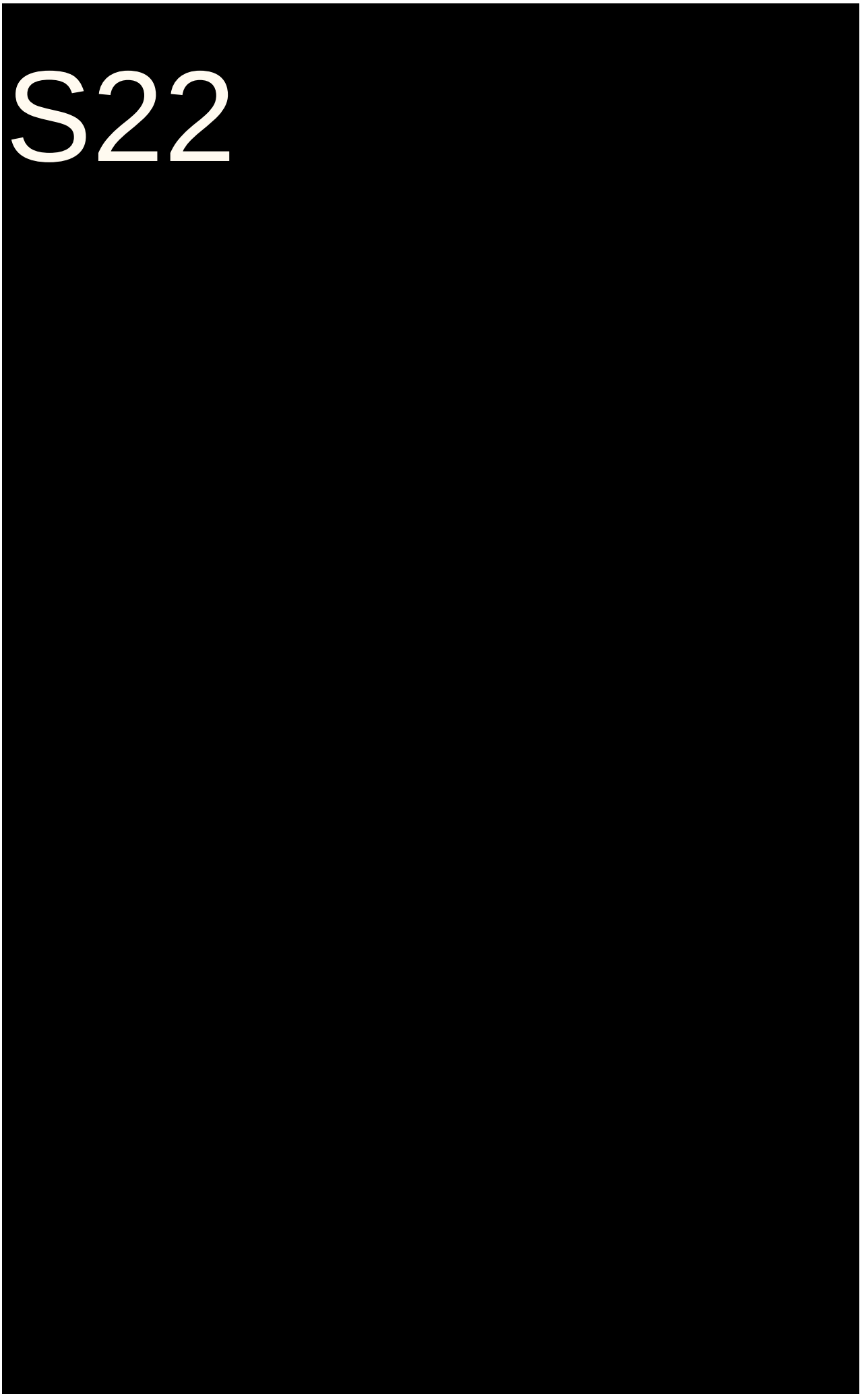
Vignettes of work progressed to date focusing on outcomes, lessons and links to the Transformation Roadmap

- Presentation – S47F (HIA):
  - Macadamias pilot project with ABS
  - Largely spoke to points in the PowerPoint
  - Cautioned that there is a very concentrated supply chain in the macadamia industry, while other industries have a far more complex structure
  - Few processors collect data on behalf of the industry
  - To assess the suitability of processor data for use in national statistics (to calculate GVP), having a close collaborative relationship between ABS and HIA and other industry bodies was key
  - The ability to officially collect and manage data under the Act was also critical
  - The project has closed as it is considered to have delivered on its objective, S22
  - S22

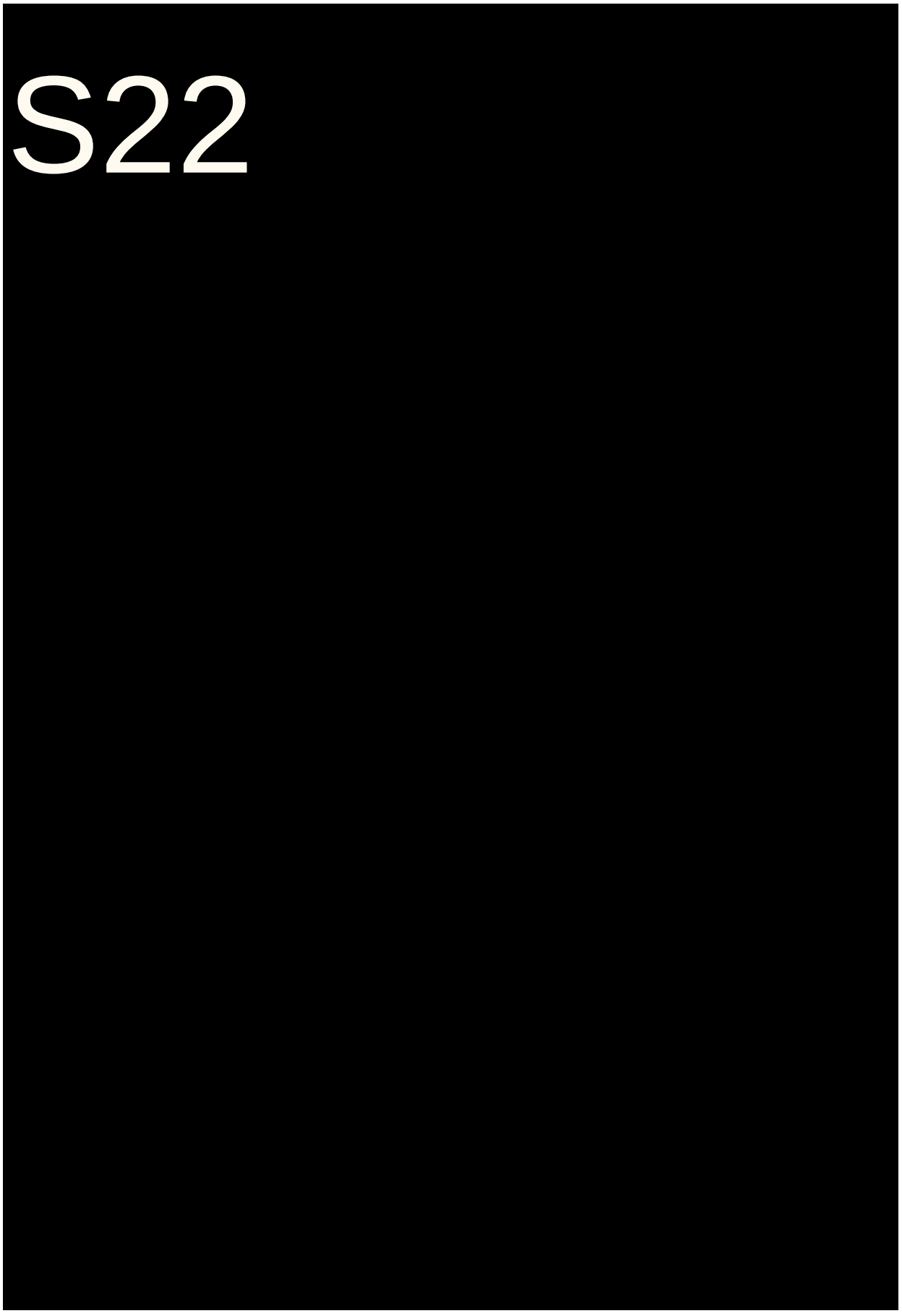
S22

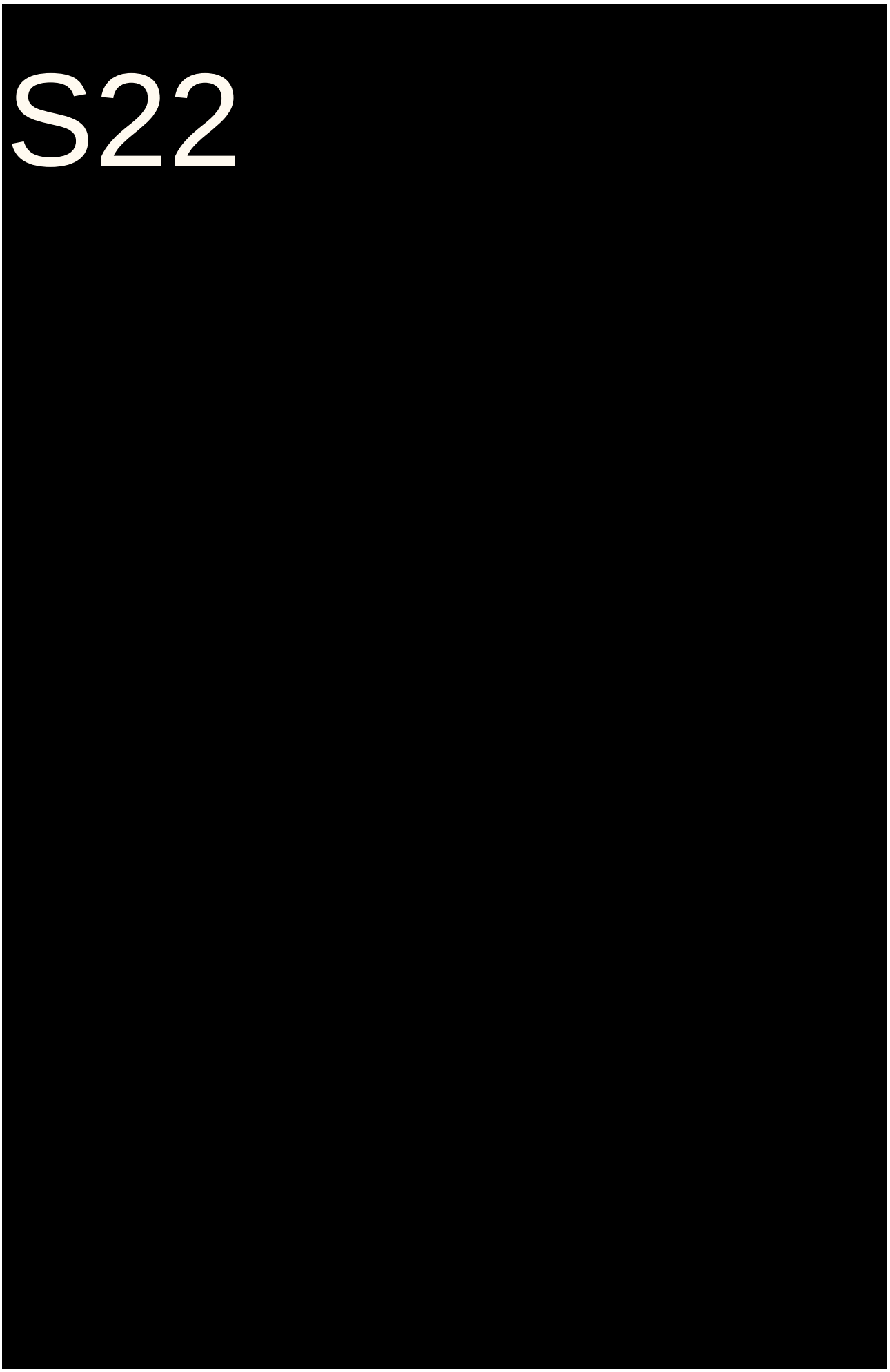

S22

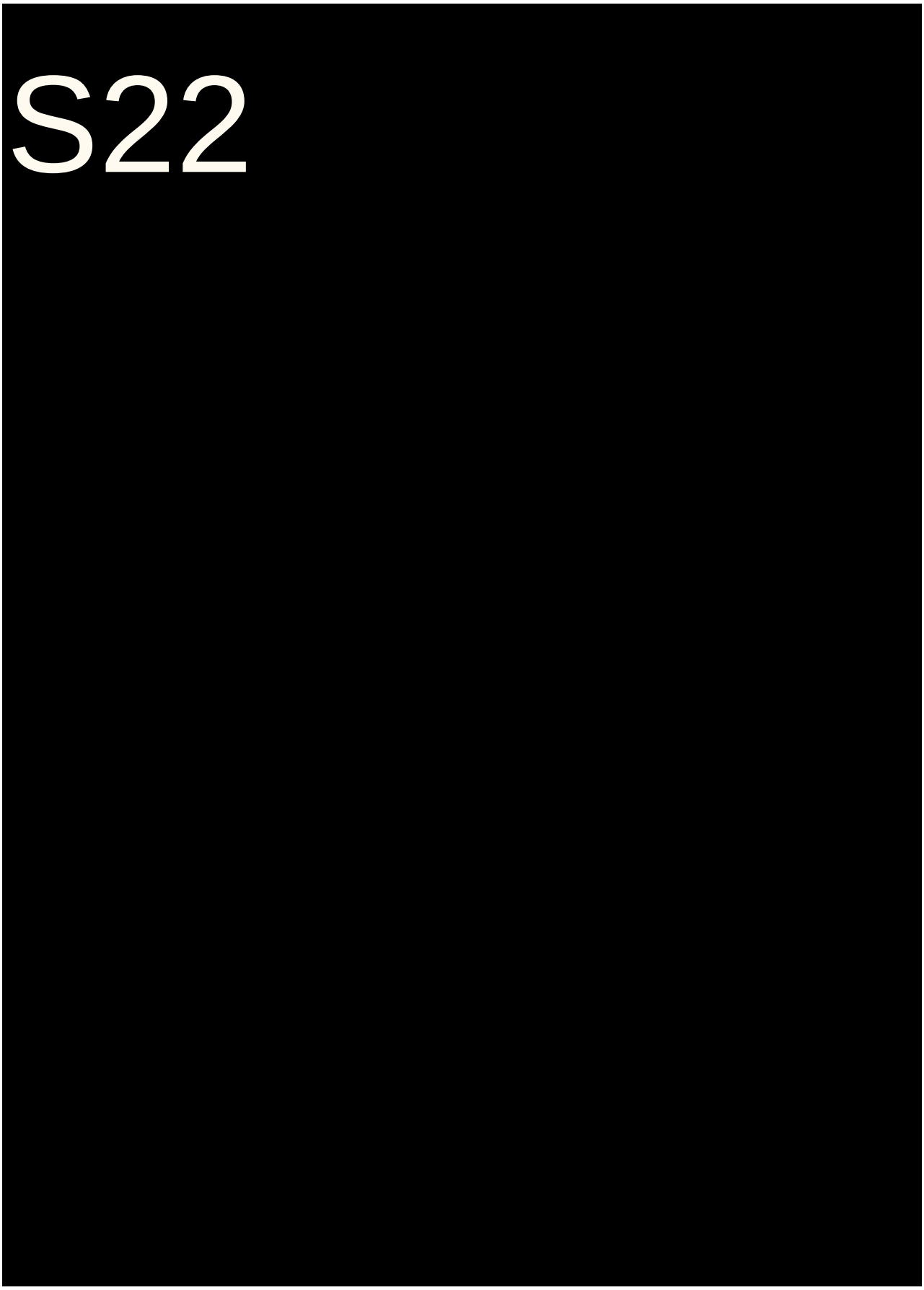

S22


S22

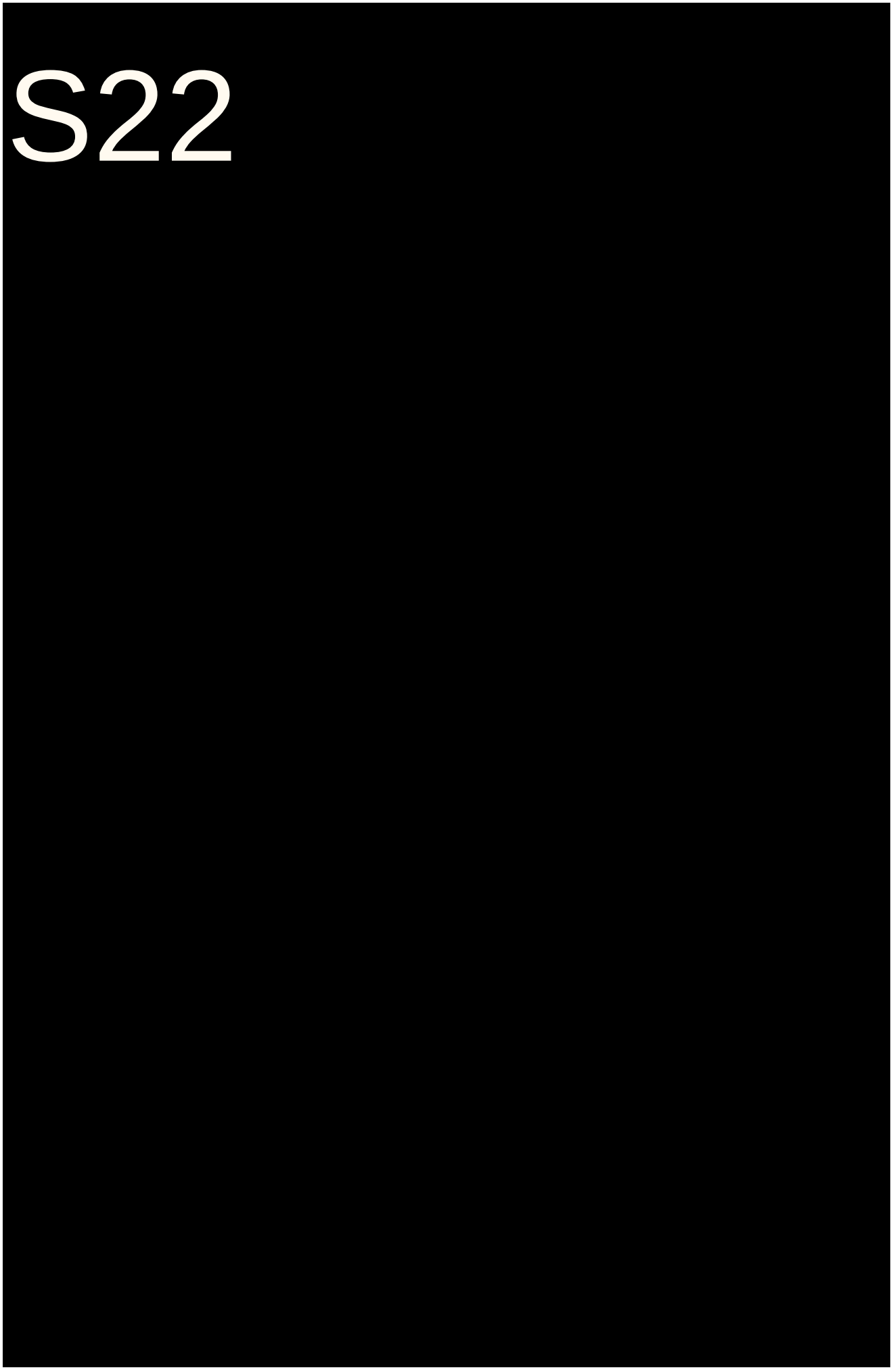
S22


S22


S22


S22

S22


S22

S22

S22


S22

Agricultural Administrative Data Initiative

S47E, S47F

Australian Bureau of Statistics

Outline

- Background
- Aims and outcomes
- Project work

Systemic issues and stakeholder concerns	Proposed solutions
A lack of strong governance and coordination for coordinated approach to statistical production	The establishment of mechanisms to improve coordination and governance across the statistical system
Agricultural statistical system lacking a central plan to guide investment, leading to potential duplication and inefficiency	The establishment of a Foundation Dataset that applies existing frameworks
Over-reliance on surveys leading high respondent burden, potential data quality issues and higher costs	Exploration of alternative data sources, including through establishing an Administrative Data Initiative
Under-utilised opportunities to harness innovative new technologies and statistical methods	Adopting new and emerging technologies in statistical use and production
The need for a culture of open data to make the best use of existing sources	Establishment of a one-stop portal for agricultural statistics to maximise the value of existing data sources and encourage a culture of open data

NASR call for action

Coordination and governance

Six key actions are needed for ABS and ABARES to fulfil their statistical leadership role

1. Supporting industry collaboration to develop and implement common data standards and data sharing protocols
2. Explore potential of farm management systems, satellite technology and precision agriculture
3. Engaging in pilot projects with industry to utilise administrative data
4. Utilising environmental-economic accounting as a framework for information relating to sustainable agriculture;
5. A collaborative project with ABARES to extract statistical value from the agricultural levy process
6. Exploring potential to consolidate surveys across industry and government

Most recent achievement is the in-principal agreement between ABS and ABARES to support a joint forward work plan

What is: Agricultural Administrative Data Initiative?

- An ABS 'Innovation Project' associated with the Carbon Farming Initiative and National Agricultural Statistics Review.
- Utilising a cross-divisional, multi-disciplinary team to design and deliver the statistical solutions to support policy makers and reduce the burden on agricultural businesses.
- Inclusive agenda looking to maximise public-private partnerships.

Aims and outcomes

- Consolidate
  - Surveys and third party data are minimised (5-10 year view)
  - Aware government, open data environment
- Involve
  - Maximise public/private partnerships
  - Enhance academic & research initiative (Big data)
- Improve
  - Drive up quality where possible
  - Ultimately, decrease regulatory burden

The 'Ag Census' of the future?

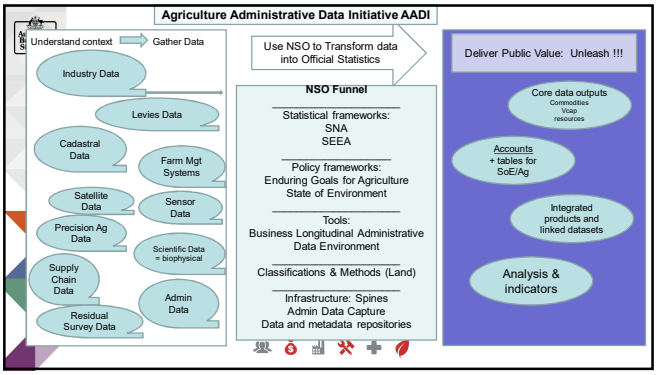
Current outputs meet *most* user needs but at an unsustainable price

What if users had small area data *every year*, with *less burden*, *less cost*, *better quality* and more *integration*?

Develop new sources over next five years
Agricultural levy payer data
Industry data sources (processors, gins, storage)
Farm management systems
Big data: Satellite, sensors, precision agriculture
Trade data
Annual integrated collection/ Pop census etc
Water, energy and biosecurity regulators
NRM information sources (eCond)
Valuers general/ Surveyors general
Modelled data
Other admin data: taxation/rebate/scheme/SBR
Direct collect essential items only (land frame?)

Farmer's views?

Agriculture Administrative Data Initiative AADI



Guiding principles for the AADI

- The ABS is the right choice for this type of initiative.
- The ABS has some principles for new data:
  - Administrative or third party data preferred (collect once, use many)
  - Build trust and confidence with respondents and providers
  - Use of strong statistical foundations including classifications of land, business and activities
  - Opening data for use in policy, planning and informing the community, whatever the level of detail.

Our Challenges

- Third party data is collected for reasons beyond statistical purposes
- Continuity and a consistent approach are hard work
- Connecting diverse data is intensive work
- Need to maintain connection to land, produce, environment and management practices
- We need to keep or create strong connections between datasets to maximise their utility

Three main programs of activity

- Strategic
- Data driven/pathfinders
- Infrastructure foundations

Strategic

- Proposed outputs:
  - Paper on recommendations for reducing regulatory burden on Agriculture
  - Information paper on results of the data substitution/integration projects

Data driven/pathfinders

- Proposed projects:
  - Industry partnerships (Macadamia, pork, egg)
  - Research partnerships (Food agility CRC)
  - Government partnerships (Levies)
- Targets:
  - Data substitution for the System of Environment-Economic Accounts – Agriculture, Fishing and Forestry
  - Start with willing partners for test case and then key commodities

Related projects in parallel

- Leveraging from related projects:
  - System of Environment Economic Accounts – Agriculture, Fishing and Forestry.
  - Ecosystems accounts for the Great Barrier Reef Region.
  - Use of Satellite Imagery for Land Management
  - Collaborations with State Departments of Primary Industries and Departments of Environment (eg, Access to project catalyst – Qld)
  - Drought resilience analysis

Infrastructure/Foundations

- Proposed projects:
  - Alignment with internal business processes
  - Develop a proposal to harvest farm management information directly (5-10 year view – similar to Standard Business Reporting)

Timeline

- Strategic
  - Paper on Ag Futures: March 2017
  - Results of Data substitution work June 2017
- Data
  - Expected between March 2017 to June 2018
- Infrastructure – Late 2017

Presentation
ABS Environment and Agriculture Statistics Branch: New Directions

S47E, S47F

Burning platform

- We can't maintain the old ways of doing things
- Community and government expect us to use and share existing data holdings
- Society is facing new problems that require new data and new ways of measuring progress
- Expectations and costs are increasing while govt revenue is decreasing

Environment and Agriculture Statistics Goals

Goal 1: Increase the use of ***alternative data sources*** for environment and agriculture statistics	Goal 3: Improve public understanding and use of ***environmental-economic accounting***
Goal 2: ***Reduce survey burden*** on Australian farm businesses	Goal 4: ***Unleash*** environment and agriculture statistics ***to reach audiences*** across government, research and the community

Our Goals and Initiatives

1. Increase

The use of **alternative** data sources for environment and agriculture statistics

By 2021 60% of our program data will be compiled from alternative sources

Major initiative:

Administrative Data Initiative

- Progressively replace agriculture survey content with alternative data sources
  - Agriculture levy data
  - Precision agriculture systems
  - Satellite data
  - Industry data
- Harness transformation to create required infrastructure/capability

The 'Ag Census' of the future?

The Administrative Data Initiative

Is it possible to provide small area data *every year*, with *less burden, less cost, better quality* and more *integration*?

Develop new sources over next five years
Agricultural levy payer data
Industry data sources (processors, gins, storage)
Farm management systems
Big data: Satellite, sensors, precision agriculture
Trade data
Annual integrated collection/ Pop census etc
Water, energy and biosecurity regulators
NRM information sources (eCond)
Valuers general/ Surveyors general
Modelled data
Other admin data: taxation/rebate/scheme/SBR
Direct collect essential items only - through friendly means

ADI: three main programs of activity

- Strategic – exploring ways to reduce burden through new technology and big data, improve communication
- Pathfinder projects/ pilots
  - Industry partnerships
  - Research partnerships
  - Government partnerships
- New foundations - for data collection, storage and access, using ABS transformed capability and leveraging industry capability

Our Goals and Initiatives

2. Reduce

The survey **burden** on Australian farm businesses

By 2021, the cumulative survey burden on farm businesses by Government has reduced by 20% and farmer perceptions of burden have improved

Major initiatives

- Through collaboration with ABARES rationalise the number of agricultural surveys by Australian Government
- Partner with industry and research sectors to reduce survey duplication and burden
- Engage with transformation to explore better ways to collect information from farmers

Our Goals and Initiatives

3. Mainstream

The understanding and use of **environmental-economic accounting** (EEA)

By 2021 the use of environmental economic accounting in mainstream reporting about the environment and economy has increased by100%

Major initiatives:

Work with governments to agree national approach to EEA and common indicators

Progressively integrate aspects of EEA's into the National Accounts and agriculture statistics

Increase application of EEAs for State of the Environment and sustainability reporting at state, national and international levels

Why do we need environmental accounting?

'A country could exhaust its mineral resources, cut down its forests, erode its soil, pollute its aquifers, and hunt its wildlife to extinction, but measured income would not be affected as these assets disappeared.'

Repetto et al

'Choices between promoting GDP and protecting the environment may be false choices once environmental degradation is appropriately included in economic performance.'

Stiglitz Commission 'Beyond GDP'

Our Goals and Initiatives

4. Unleash

Environment and agriculture **statistics** to reach audiences across government, research and the community

By 2021, citations of ABS environment and agriculture statistics have increased by 100%

Major initiatives

- Increase appropriate microdata access through in-postings
- Improve confidentialisation and dissemination systems
- Use new platforms (BLADE , MADIP) for data integration
- Deliver useful information to industry and policy audiences (eg NFF 'Australian Farmers' website)

Our 'value chain'

Strategy	Development	Delivery
Engage with the external environment to anticipate needs, assess opportunities and identify priorities	Undertake development, implement pilots and partnership projects and produce experimental statistics	Produce reliable ongoing statistics that reach users in the right ways

ABARES/ABS Collaboration

- Builds on National Agricultural Statistics Review
- Aims to improve the quality and efficiency of the agricultural statistics system and to reduce the survey burden on farming businesses.
- Through two mechanisms:
  - Increasing the use of alternative data sources
  - Improved coordination and communication
- Complements the strategic objectives of both agencies
- Joined up engagement and coordinated investment

Consultation so far

- Huge goodwill & offers of help
- Some industry RDCs feel that they have better data sources
- Examples of best practice
- Emerging technologies
- Ideas to attract farmers to provide good data and deliver value to them
- Respect for ABS role as custodian

Progress!

- Focus on four-five key projects
- Look at how we can integrate these data
- Data Quality Assessment Tools
- Key is to understand how the ABS can adopt industry data

- Wool
- Macadamia
- Avocados
- Dairy
- Citrus

Next Steps

- ADI phase 1 finishes 30 June
- ADI phase 2 under discussion
- More work on
  - Levies
  - Foundation data set

- ADI phase 2
- Option: industry led and funded tools for growers
  - Integrated with Whole of Government data needs
  - Includes State Government
  - Includes RDCs
- Option: low touch approach
  - Steady roll out over 5 years.

• weight of product to processor vs weight of production at the farm, (((if grower was given an ABS agricultural census form, with a 30 June end date.))).	
• if the weight of product leaving farm gate is known	
• actual weight via feedback from processor (((vers the estimates for ABS surveys, likely variants .))	
• do all your macadamia nut got to the closest processor or do you pick ? is so on what basis	
• losses between farm gate and processor, ie on farm sale of macadamia nuts, other losses	
• on farm growing and processing vs pure processors	
• what sort of record keeping for you undertake, electronic, paper, etc and in what format, ie tonnage by week or month or specific processor?	
• when levies due to be paid, impact of timing of harvesting and how collected. grower pays or processor pays for grower?	
• If any other surveys/government organisations or research development corporations have asked about pesticide & fertiliser application, land management practices, water use and so on.	
• any surveys that you are have received, who from ??	
• do you use either the ABS or AMS published data to help your business	
what other types of information could be useful to help	

Processors

There is a difference between what AMC estimate produce for tonnage of macadamia nut produced and what ABS produced. Are you aware and what do you thing the difference could be contributed to ? (AMS and S47F S47F only)	
***DO NOT quote: There is around a 12Mill Ton diff, 2015/16.* (AMS and S47F only)**	
where does the data from AMS come from and is what form (AMS and S47F only)	
the weight of product to processor vs weight of production at the farm, (((if processor was given an ABS agricultural form, with a 30 June end date.))).	
if the weight of commodity leaving farm gate is known, and the same as the weight the processor ? what are the main differences	
when levies due to be paid, impact of timing of harvesting and how collected. grower pays or processor pays for grower?	
what sort of record keeping for you undertake, electronic, paper, etc and in what format, ie tonnage by week or month or specific farmer?	
how many growers below to this processor.	
are all farmers macadamia nut processed at the closest Processor	
are there growers that change processors	
payments to farmers splits across financial year, harvested on fin year, payments in next fin year? levies payments	
what level of data is kept by producers, weights/prices/location/grower/water content/levy payment/ etc	
If any other surveys/government organisations or research development corporations have asked about pesticide & fertiliser application, land management practices, water use and so on.	

S22

S47E, S47F Broad summary of Macadamia Pilot Project (R... 18/12/**** 12:23:22 PM

Broad summary of Macadamia Pilot Project (Response to: Header - Macadamia Nut project)

Environment & Agriculture Statistics Branch WDB S47E, S47F 21/09/2018 01:50 PM

S22

Macadamia Pathfinder Pilot

This pilot was jointly funded by the ABS and Horticulture Innovation Australia. Iy specifically assessed the potential for data from macadamia processors to be used in the production of gross value of agricultural production (GVAP) estimates. The pilot included site visits to growers and processors to understand how data are collected and managed through the supply chain. A data quality assessment was undertaken to measure the readiness of the processor data for integration into ABS statistical processes.

The macadamia pilot revealed that the levy consignment reports issued by the nut processors are a rich source of data that are provided to individual growers and to DAWR. These reports have potential to perform a critical role in the agricultural statistics system, not only by replacing key survey content, but by enabling linkage of a range of key variables.

One of the other potential difficulties in using the Australian Business Number (ABN) as a link between the Agricultural frame and the unit produced through the macadamia study was the frequent change of ABN or unit ownership. ABNs also have problems around the location of the farm business and the location of the commodity. The attribution of production information to a particular geographical location is critical to producing statistics on yields, productivity and area of holding for example.

The AADI pilot found that this could be achieved by improving the address location field or allocating codes to each land plot or cadastre used for agricultural production. State governments, for example, assign a property identification code (PIC) based on the cadastre of properties where livestock are grown. In another example, Local Government also assigns property classification codes for valuation purposes. If an existing code of this kind were able to be included on the levy consignment report and associated with an ABN, this would represent a breakthrough in the transformation of the agricultural statistics system.

A key finding of the macadamia pilot was that levy based reports may hold the key to systematising the collection of most agricultural statistics without imposing additional burden on growers. These findings can help to guide the redevelopment of the levy collection system currently being undertaken by the DAWR, for example by recommending that levy consignment reports for manual and Levies Online submission are standardised with a minimum set of meta data fields including the current requirement for ABN and production, with the inclusion of a PIC or other locational element.

It was recommended that phase 2 of AADI support the use of levy reports from all macadamia processors to generate statistics through a full processor collection in the Value Add of Commodity Production through AADI processes would be a first step in leading alterative data collections. The AADI phase 2 approach would also continue research on linking location to processor based data.

S22

S47E, S47F S47E, S47F Agriculture Admin Data Acc... 24/02/2017 07:50:32 AM

Basics

S47E, S47F /Staff/ABS 24/02/2017 07:50 AM

Send	To Environment & Agriculture Statistics Branch WDB@ABS cc
Subject	Fw: Macadamia GVP project MC16013 contacts

S22

S22

S47E, S47F

Agriculture Admin Data Accounts

Accounts Development | Agriculture and Environment Branch | **Australian Bureau of Statistics**

(P) S22 **(F)** S22

(E) S22 **@abs.gov.au** **(W)** **www.abs.gov.au**

----- Forwarded by S47E, S47F Staff/ABS on 24/02/2017 07:50 AM -----

From: S47F <S47F @horticulture.com.au>
To: S47E, S47F @abs.gov.au>,
Cc: S47E, S47F <S47E, S47F @abs.gov.au>
Date: 22/02/2017 04:54 PM
Subject: Macadamia GVP project MC16013 contacts

HI S47E, S47F,

Good to chat earlier.

S47F from Macadamia Processing Co (MPC) is the Chair of the Australian Macadamia Handlers Association (AMHA) and will be the best person to speak to regarding processor data. He will be able to answer many of your questions and provide guidance on how to move forward. I have spoken to him and let him know that you will be in touch. He is very supportive of the project. Details below:

S47F
Office: S47F

S47F said that there are actually 10-12 processors, however 4-5 make up 80% of the crop.

S47F is the Project Lead on the Macadamia Crop forecasting project. He will be able to provide you with information on this project. It is quite advanced.

S47F
Principal Scientist (Biometry)
Agri-Science Queensland
Department of Agriculture and Fisheries
Telephone S47F; **Email** S47F @qld.gov.au
Postal Address: 2A East, Ecosciences Precinct, **GPO Box 267**, Brisbane Qld 4001
Street Address: 2A East, Ecosciences Precinct, 41 Boggo Road, Dutton Park Qld 4102

S47F is the project lead on the Macadamia Benchmarking Project. S47F would be another useful contact.

S47F
Senior Development Horticulturist
Department of Agriculture and Fisheries

T S47F E S47F@daf.qld.gov.au W www.daf.qld.gov.au
Maroochy Research Facility, 47 Mayers Road, Nambour QLD 4560
PO Box 5083 SCMC Nambour QLD 4560

Still waiting to get in contact with S47F so I can give him background on the project and that you will be in touch. Also waiting for some grower suggestions.

Thanks,

S47F

S47F | **Data Analyst** | Horticulture Innovation Australia Limited | Level 8, 1 Chifley Square, Sydney NSW 2000 | Ph: S47F | Fax: 02 8295 2399 | Email: S47F@horticulture.com.au

From: S47F
Sent: Wednesday, 22 February 2017 12:41 PM
To: S47E, S47F@abs.gov.au>
Cc: S47E, S47F@abs.gov.au>
Subject: RE: This mornings meeting

Hi S47E, S47F,

Thanks for sending this through. In the first instance I will get you the detail of the processors contact details. I will also draft a reply to the other questions that you have noted.

When you mention 'forms design' and identifiers, will this depend on the information that we are able to provide you, or based off the initial contact with the processors?

Regarding the requirements from the ABS it would be useful to know the specific team members who will be conducting the consultation with the processors. Also are the questions that you have mentioned under *data requirement* relate to what you are seeking from Hort Innovation in the first instance, or will these also be asked of the processors. I don't think we have access to the processors data, and this will need to be obtained from them directly, I am unsure if it is aggregated.

Thanks,

S47F

[redacted] | **Data Analyst** | Horticulture Innovation Australia Limited | Level 8, 1 Chifley Square, Sydney NSW 2000 | Ph: [redacted] | Fax: 02 8295 2399 | Email: [redacted]@horticulture.com.au
From: [redacted] [mailto:[redacted]@abs.gov.au]
Sent: Wednesday, 22 February 2017 11:31 AM
To: [redacted] <[redacted]@horticulture.com.au>
Cc: [redacted] [redacted]@abs.gov.au>
Subject: RE: This mornings meeting

Hi [redacted]

Spoke to [redacted] and here are the answers to your questions.

Key stakeholders *– we would imagine this would include the AMS and the four processors at a minimum, but could also include the project leads from the crop forecasting and benchmarking projects. We are happy to make the required introductions*

- At a minimum we need to talk with the people who have the volume and price data. It would be useful to go through how the macadamia goes from farm to market and the measurement points. (visit at least one, possibly two?)
- Would be useful to talk with other users of this data (crop forecasters etc) to get a feeling for their view on the quality of the data and why ABS isn't meeting their needs.
- Any academic interest in this space would be useful, eg if anyone is thinking about precision agriculture/big data style information harvesting from
- Others?

Data requirement

- What data is collected yourself and\or commission others to collect for you?(eg. weight, quality, processed, farm gate, aggregated)
- How much data do you have (eg. how far back, original, vetted, etc).
- How often do you collect these data? (eg. calendar year, financial year, etc)
- What descriptors/information about your data do you have? Is there a specific set of 'metadata' (explain) or information is embedded in a spreadsheet?
- Are you willing to provide a sample set?
- Are there any barriers to providing these data on a regular basis? (eg. Ease of extraction)
- Are the data collected in one central location or do we need to poll the four processors?

Timing's for consultation and when you are planning on commencing

- We have started collecting relevant data/metadata/scope and compiling on what basis the ABS collects Macadamia Nut data. We have commenced work on quality assessment, processes for bringing data in and storage.
- There will be a delay over the forms design, so getting the identifiers documented and put onto a form would seem to be vital and urgent.

Is there also anything that you require from the ABS ?

cheers
[redacted]

Agriculture Admin Data Accounts

Accounts Development | Agriculture and Environment Branch | **Australian Bureau of Statistics**

(P) S22 **(F) (02) 6252 7933**

(E) S47E, S47F @abs.gov.au **(W)** www.abs.gov.au

S47F ---20/02/2017 10:42:16 AM---Hi S47E, S47F – I had just sent this to S47E, S47F – but will repeat again in light of your note. If you are ab

From: S47F <S47F @horticulture.com.au>
To: S47E, S47F S47E, S47F @abs.gov.au>,
Cc: S47E, S47F S47E, S47F @abs.gov.au>
Date: 20/02/2017 10:42 AM
Subject: RE: This mornings meeting

Hi S47E, S47F – I had just sent this to S47E, S47F – but will repeat again in light of your note. If you are able to provide anything on this then that would be good, but understand that it may be best to wait until S47E, S47F returns…

Hi S47E, S47F

In lieu of this morning's phone hookup, could you please confirm the following to allow us to facilitate the commencement of this project:

- *key stakeholders – we would imagine this would include the AMS and the four processors at a minimum, but could also include the project leads from the crop forecasting and benchmarking projects. We are happy to make the required introductions*
- *data requirements*
- *timings for consultation and when you are planning on commencing*

Thanks very much,

S47F

S47F | **Data Analyst** | Horticulture Innovation Australia Limited | Level 8, 1 Chifley Square, Sydney NSW 2000 | Ph: S47F | Fax: 02 8295 2399 | Email: S47F @horticulture.com.au


S22

S22

S47E, S47F Initial meeting invite with Macadamia Nut (Res... 18/12/**** 12:23:22 PM

Initial meeting invite with Macadamia Nut (Response to: Macadamia Nut Contacts and corro (HEADER ONLY))

Environment & Agriculture Statistics Branch WDB S47E, S47F 24/02/2017 07:53 AM

S22

S22

Invitation: MC16013 Macadamia GVP analysis inception call

Mon 20/02/2017 10:00 AM - 10:30 AM

Attendance is **required** for S47E, S47F

Chair: S47F @horticulture.com.au

Sent By: S47F < S47F @horticulture.com.au>

Location: Sydney MR3 (Glasshouse)

S47F @horticulture.com.au S47F has invited you to a meeting. You have not yet responded.

S47F

Description

Hi All,a

As the contract for MC16013 has been executed it would be timely to hold an initial inception call to discuss the project in more detail, including key contacts for ABS to speak to regarding data sources to underpin investigations.

In addition to the data that will be collected by the processors, there are currently two key levy paying projects that may be of relevance to this project: MC15009 Macadamia Crop Forecasting 2015-2018 and MC15005 Benchmarking in the Macadamia Industry 2015-18. These projects operate at an advanced level with good support from industry and growers and have been running for a number of years. I have attached the 2016 crop forecast report (2017 report still to be finalised), and an analysis of the benchmarking data from 2009 to 2015 for background reading.

Participants:

S47F -Data Analyst
S47F - Macadamia Fund Manager
S47F - Macadamia Relationship Manager
S47E, S47F - ABS Director
S47E, S47F - Lead Analyst

S22

Speak to you then.

Regards,

S47F

[redacted] | Data Analyst | Horticulture Innovation Australia Limited | Level 8, 1 Chifley Square, Sydney NSW 2000 | Ph: [redacted] | Fax: 02 8295 2399 | Email: [redacted]@horticulture.com.au<mailto:[redacted]@horticulture.com.au>

[Facebook]<https://www.facebook.com/HortInnovation> [Twitter] <https://twitter.com/hort_au>

[cid:image3d263c.JPG@d4a548fd.4fbe0098]<
http://horticulture.com.au/membership-application-form/>

- imageece716.PNG - image5d8620.PNG - image3d263c.JPG - MS102 2016 forecasting report - 2016 03 31.docx - macadamia-industry-benchmark-report2009-2015.pdf

S22

S47E, S47F Ho S47E, S47F Could you dig out the below comp... 08/03/2017 02:43:03 PM

Basics

S47E, S47F /Staff/ABS 08/03/2017 02:43 PM

Send	To S47E, S47F /Staff/ABS@ABS cc S47E, S47F /Staff/ABS@ABS, S47E, S47F Staff/ABS@ABS, Environment & Agriculture Statistics Branch WDB@ABS
Subject	Macadamia industry contacts. [redacted]

S22

S22

Ho S47E, S47F

Could you dig out the below companies data from where every you can. They are all Processors, I haven't got the name of the growers as yet. I will not tell them I have their data at all, just need it to get an idea of what size etc we are looking at. I may be able to confront what they give us to what was collected by the ABS, depends really on what, if any, data they provide.

S47E

These are processors we will have a talk to over the phone the week after next"

S47E

cheers

S47E, S47F

Agriculture Admin Data Accounts

Accounts Development | Agriculture and Environment Branch | **Australian Bureau of Statistics**

(P) S22 **(F) (02) 6252 7933**

(E) S47E, S47F**@abs.gov.au** **(W) www.abs.gov.au**

S22

BCS 2018/19 Design Documentation

Table of Contents

S22

For the 2016/17 cycle the BCS BSC implemented a change in the collection method by providing both the innovation and IT usage statistics as well as unit-level data through a single survey vehicle. For the purposes of this documentation the term BCS will refer to this new survey. In the process of implementing the single survey vehicle in 2016/17, several changes were made to the sample design including changes to the stratification and the employment size-based CE thresholds. Details of the changes are in 2016/17 summary document.

S22

The 2018/19 cycle continued the single survey design, as implemented in previous two cycles. However the 2018/19 BCS sample design has implemented a couple of changes compared to previous cycle, including:

S22

2. A total of 184 units that were In-Scope for both the BCS and BLD but dropped out from the BLD panel J for cycle 2016/17 were requested to be put back to the BLD for this cycle to support more informative longitudinal analysis with sufficient BLD units.

The main purpose of this document is to highlight the aspects of the BCS design for the 2018/19.

1. Structure of 2018/19 Design

This section looks to provide a brief overview of the BCS, BLD as well as the Growth Hubs top-up designs.

1.1 Overview of the BCS, BLD, and Growth Hubs top-up designs

The BCS design is, for all intents and purposes, a normal ABS survey design. The main point of differentiation with most other surveys is the design of the BCS is predominantly based on the rates of binary (0/1) variables (the averages of which can be considered reasonable approximations of proportions). Furthermore, the BCS allocation is designed to meet the accuracy constraints which are unusual in that they are essentially standard error constraints for proportions rather than RSEs for continuous variables. The process for finding this BCS allocation can be found in Section 3.

The sample currently combines across three processes.

(i) The BLD which is a carry-over of the longitudinal survey. The 2017/18 survey included 3 panels (Panel 9, J and K), For the 2018/19 cycle, the Panel 9 was removed as scheduled, leaving only Panel J and K as the live serving panel. Given the longitudinal nature of these units to provide a steady-state, all units on Panel J and K for cycle 2017/18 were completely carried over to the 2018/19 cycle. Additionally, a total of 184 units that were In-Scope for both the BCS and BLD but were dropped out from the BLD panel J for cycle 2016/17 were put back to the BLD for this cycle.

(ii) The core BCS, which continued under the sample design regime implemented in 2017/18, except that the design constraint for the variable of "Total income received from orders taken over the internet" has been removed.

(iii) The growth hubs top-up, which also continued under the sample design implemented in 2017/18.

1.2 Applied sample design changes

Maximum sample size

The sample size achieved in 2017/18 was 10,437 units, which was under the required maximum sample size of 10,200 units. For the 2018/19 cycle, it was requested to drop the maximum sample size to around 9,500, given the fact that Panel 9 of the BLD will be rotated out, and the design constraint for the variable of "Total income received from orders taken over the internet" will also be removed.

Treatment of existing BLD panels

Given that all units selected in both the BCS and BLD get sent the BCS form (and BLD-only units received a short version of that form) then no longer running BLD would not require any changes to the BCS form.

The size of the 'Core' sample (see below for further details), which was the sample used to meet the accuracy constraints, turned out being around **5,985 units**. And the size of the Growth-Hubs top up turned out being around 1,184. Given the maximum sample size required for the BCS is 9,500, it ended up being the case that it was not only possible to continue selecting all 2,256 of the live units within the two remaining BLD panels (Panel J and K) that were still active as of the 2017/18 cycle, but also putting back to Panel J an additional 184 units that were In-Scope for both the BCS and BLD but were considered dead and dropped out for cycle 2016/17.

'Core' and 'Flexible' samples

As per the suggestions of the ABS Business Charter, the BCS sample design aimed to rotate out approximately 1/3rd of the sampled sector every year. This resulted with the sampled sector spending at most three years being sampled by a survey. The 16/17 redesign for the BCS had seen a large number of data requests or changes in sample design and these would all likely alter the size and distribution of the allocation. This in turn would impact on the rotation rates within the (old) BCS sample.

Some of these changes or requests were considered to be permanent or otherwise long-term while others were thought to only be needed for a short duration. Following the 2016/17 cycle, the 2018/19 survey considered the ‘old’ BCS as the ‘Core’ sample of the new single survey. This ‘Core’ component of the survey is what is placed under the rotation policy.

All the short-term data requests from the 2016/17 design are considered as the 'Flexible' component of the new BCS (see table 1.2.1). Given that these short-term requirements will likely change from year-to-year, and so these units will only need to be enumerated on an ad hoc basis then it was decided that it was not necessary to have forced rotation for these units;

As noted in greater detail in the 'Selection' section of this document it was decided that, in order to not overly burden units that are selected as part of the Flexible component of the BCS, these units are selected in such a way that they are unlikely to be reselected in the BCS as part of the Core sample in the near future.

Table 1.2.1 lists how the requested changes for the 2016/17 BCS sample design were classified in terms of being long-term or short-term requirements.

Table 1.2.1 List of requirements for BCS

Long-term requirements	Short-term requirements
Reduce overall sample size (the reduction is a once-off and the intention is to keep the sample size to this level for the foreseeable future)	Continue following existing BLD panels
Changes to CE bounds	Additional sample top-up, focussed around Growth Hubs
In addition to have the old BCS design constraints also having constraints or sample top-ups for: • Collaboration with Higher Education • Innovation • Comparisons with CIS outputs in the relevant parts of the population	

The main reasoning behind separating the BLD panels from the BCS Core is that the aim for the BLD was to maintain the same units across cycles. If BLD strata were matched to BCS strata (which would be straightforward if there were no non-employing BLD units kept on) and these BLD units were selected with first preference (in order to maximise the continuation of these panels from their current state) then this would alter the level of rotation for the current BCS selections. It would also cause more clumpy rotation for the BCS into the future once these BLD panels run their course. This concern is particularly notable for the most recent BLD Panel. Panel K. This has a very different scope to the other BLD panels (as well as the BCS) because it only selected units that were micro-startups ie. businesses with 1-10 employees where these businesses came into existence at most a year before the 15/16 reference period. Introducing Panel K into the existing BCS stratification would result in an over-representation of micro-startups.

The main down-side of having these BLD units be treated as part of the 'Flexible' component is that doing so cuts into the amount of sample available to satisfy all the other short-term requirements. Further details on this can be found in the 'Allocation' section that deals with the Growth Hubs top-up.

Finally, it should be noted that even though BLD data was followed for five years the intention is that any non-BLD data selected as part of the new BCS will only be kept in sample for at most three years, as was the case under the old BCS. The BSC was comfortable with continuing this practice because (anecdotally) they found that most of the most interesting data collected as part of the longitudinal analyses was collected in the first three years a unit was in a BLD panel.

2. Sample Design constraints/specifications

2.1 BCS Core

2.1.1 Core scope

The scope for the frame used for the BCS Core (as well as the BLD and Flexible selections) is:
The subset of the ABS Common Frame consisting of live units with ITW role (employing units), and excluding:

- Units in ANZSIC classes
  - 6240-6249 (Group 624),
  - 6330-6339 (Group 633),
  - 7500-8399 (Divs O, P),
  - 9540-9559 (Groups 954, 955) and
  - 9600+ (SubDiv 96); and
- Units with SISCA starting with 3, 6.

2.1.2 Core design constraints

Table 2.1.1 contains the standard error constraints that the BCS design needed to satisfy, and Table 2.12 contains the stratum-level sample size constraints that the Core design needed to satisfy.

Table 2.1.1. Accuracy constraints used for BCS allocation.

Variable	Classification	Values	Accuracy Constraint
Proportion of businesses which placed or received orders via the INTERNET	Australia	-	[redacted]
	ANZSIC06 Division	A, B, C, D, E, F, G, H, I, J, K, L, M, N, Q, R, S	
	Stratification Size	1, 2, 3, 4, 5	
Innovation rate for innovations in goods and services, operations, organisation, or marketing	Australia	-	
	ANZSIC06 Division	A, B, C, D, E, F, G, H, I, J, K, L, M, N, Q, R, S	
	* ANZSIC06 Divisions within the scope of the Community Innovation Survey	B, C, D, F, I, J, K, M	
	Stratification Size	1, 2, 3, 4, 5	
	Reported Employment Class	1, 2, 3, 4	

Collaboration for the purposes of Innovation	Australia	-	S47E
	ANZSIC06 Division	A, B, C, D, E, F, G, H, I, J, K, L, M, N, Q, R, S	
	Stratification Size	1, 2, 3, 4, 5	
	Reported Employment Class	1, 2, 3, 4	

The Reported Employment Class of 4 is the agglomeration of the Stratification Sizes of 4 and 5. The other three Reported Employment Classes are equivalent to their Stratification Size counterparts. The definitions for the Stratification Size classes can be found in Section 2.1.3

Table 2.1.2 Minimum sample sizes for Growth Hubs top up allocation

Stratum Size	Minimum Sample Size
999 < STRAT_DES	Stratum population size
STRAT_DES <= 999	Maximum of 8 and 1/500 of the stratum population size.

Note that the minimum sample size cannot exceed the stratum population size, so where the minimum sample size calculation exceeds the stratum population size, it is set to the stratum population size.

In addition to the constraints in tables 1 and 2, the Core design must also satisfy the following constraints:

1. Voluntary rotation rate to be about 1/3 per cycle;
2. ABS Selection System rotation controls to allow for a 6 year holiday for units in sampled strata which have been in sample for 3 years in-sample;
3. The maximum total sample size for the BCS was to be about 9,500;
4. The design must meet the accuracy constraints in table 1 with an expected response rate of 90% for all sampled strata;

2.1.3 BCS stratification

Figure 2.1.1 below describes the way the stratum information for the Core is put together. Table 2.1.3 defines all the possible values.

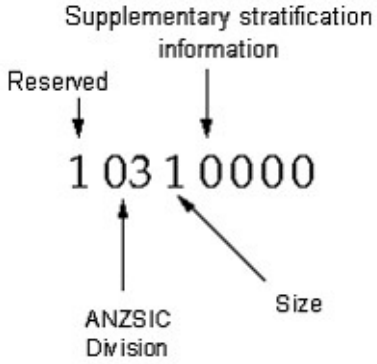


Figure 2.1.1 Diagrammatic representation of stratum definition for the Core sample.

Table 2.1.3. Core stratification values for a unit on the frame.

Digits(s)	Name	Values
1	Reserved	1 when in-scope of BCS, 9 otherwise.
2 to 3	Industry division	ANZSIC06 Divisions 01, 02, ..., 19, when in-scope of BCS, 99 otherwise.

4	Size	9 = Not in BCS scope, 1 = STRAT_DES 0 to 4 2 = STRAT_DES 5 to 19, 3 = STRAT_DES 20 to 199, 4 = STRAT_DES 200 to 999, 5 = STRAT_DES 1,000+
5 to 8	Supplementary stratification information	0000 when in-scope of BCS, 9999 otherwise.

2.2 BLD

From the 2016/2017 cycle new panels are no longer selected for the BLD. Further information on the scope, stratification and design constraints of the historical BLD panels can be found in **[4]**.

2.3 Growth Hubs top-up

As mentioned in Section 1 the Growth Hubs top-up was the only data or sample design requirement from the BSC, DIIS or BSM that was not covered by the Core allocation.

2.3.1 Growth Hubs scope

The scope for this Growth Hubs top-up are all ANZSIC classifications of a given level that contain Growth Hubs. The subtlety to be aware of here is that Growth Hubs are defined at the ANZSIC Class level, but this might not necessarily be the level at which selections for this top-up were taken. The reason why the top-up selections might be taken at a different level of ANZSIC Classification is that ANZSIC Class is a very fine level of stratification and therefore stratifying at this level might cause the minimum sample, let alone the actual sample size, for this top-up to be prohibitively large given the overall sample size constraint for the BCS.

Table 2.3.1 lists the scope in terms of ANZSIC Classes and the corresponding ANZSIC Sub-divisions for the Growth Hubs top-up.

Table 2.3.1 Growth Hub scope

Growth Hub	ANZSIC Classes	ANZSIC Sub-Div
Advanced Manufacturing	1811, 1812, 1813, 1821, 1829, 1831, 1832, 1841, 1842, 1851, 1852, 1891, 1892, 1899, 2311, 2312, 2313, 2319, 2391, 2392, 2393, 2394, 2399, 2411, 2412, 2419, 2421, 2422, 2429, 2431, 2432, 2439, 2441, 2449, 2451, 2452, 2461, 2462, 2463, 2469, 2491, 2499	18, 23, 24
Mining, Equipment, Technology and Services	1012, 1090, 2462, 2491, 3109, 6922, 6923, 6925, 6962	10, 24, 31, 69
Energy	0600, 0700, 1011, 1701, 1709, 5021	06, 07, 10, 17, 50
Food Products and Agribusiness	0121, 0122, 0123, 0131,0 132, 0133, 0134, 0135, 0136, 0137, 0139, 0141, 0142, 0143, 0144, 0145, 0146, 0149, 0151, 0159, 0160, 0171, 0172, 0180, 0192, 0193, 0199, 0201, 0202, 0203, 0411, 0412, 0413, 0414, 0419, 0529, 2461, 1521, 1831, 1111, 1112, 1113, 1120, 1131, 1132, 1133, 1140, 1150, 1161, 1162, 1171, 1172, 1173, 1174, 1181, 1182, 1191, 1192, 1199, 1211, 1212, 1213, 1214, 6620	01, 02, 04, 05, 11
Medical Technologies and Pharmaceuticals	1841, 1842, 2411, 2412, 3491, 3720	18, 24, 34, 37
Professional, Scientific and Technical and Computer Design Services	6910, 6922, 6923, 6924, 6925, 6962, 6970, 6999, 7000	69, 70
Information Media and Telecommunications	N/A	54, 55, 56, 57, 58

2.3.2 Growth Hubs design constraints

The maximum sample size for the Growth Hubs top-up was based on the sample left over after taking into consideration the:

- Maximum sample size for the BCS survey vehicle;
- The amount of sample allocated to the Core and the BLD panels;

This ended up being around 1,200 units being available for the Growth Hubs top-up.

Given that there were not many units available for this top-up there were no explicit accuracy constraints set for the Growth Hubs top-up. Instead, it was decided to distribute sample in order to have a minimum sample of 8 and/or a maximum weight of 500 in the relevant Growth hubs strata.

2.3.3 Growth Hubs stratification

Figure 2.3.1 describes the way the stratum information for the Growth Hubs is put together and Table 2.3.2 defines all the possible values for this stratum identifier.

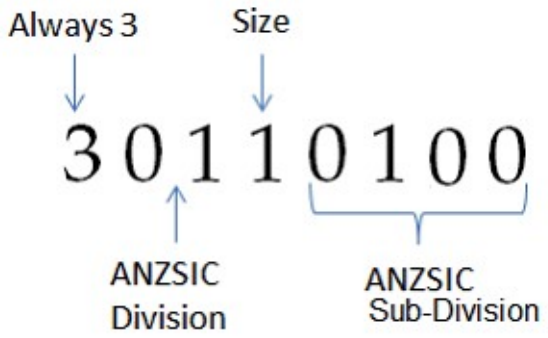


Figure 2.3.1: Diagrammatic representation of stratum definition for the growth hubs frame

Digits(s)	Name	Values
1	Reserved	Always takes the value 3
2 to 3	Industry division	ANZSIC06 Divisions 01, 03, 05, 06, 09, 12 and 13
4	Size	1 = STRAT_DES 0 to 4 2 = STRAT_DES 5 to 19, 3 = STRAT_DES 20 to 199, 4 = STRAT_DES 200 to 999 5 = STRAT_DES 1,000+ (CE'd)
5 to 8	2-digit ANZSIC Sub-division	2-digit ANZSIC Sub-divisions which contain Classes in scope of the growth hubs

Table 2.3.2: Stratification values for a unit on the Growth Hubs frame

S22

4. Calculate stratum variances of constraining variables

The aim of this process is to produce a sample allocation that satisfies the required accuracy constraints for key estimates. This allocation is based on the expected means and variances of the responses to key design variables in the 14/15 cycle. The variance constraints for the Core involve a total and three rates (that is, quantities of the form Y/X, where X and Y are estimates of total):

1. The total income received from orders taken over the internet;
2. The estimated number of businesses which placed or received orders via the INTERNET (Y), divided by the estimated number of live responding businesses (X);
3. The estimated number of businesses with an innovation in goods and services, operations, organisation, or marketing (Y), divided by the estimated number of live responding businesses (X); and
4. The estimated number of businesses with that Collaborated for the purposes of Innovation (Y), divided by the estimated number of live responding businesses (X);

These expected means and variances are estimated from data collected in the two previous BCS cycles ie. the BCS 2016/17 and BCS 2017/18 survey data. The assumption being used here is that recent historical data will be a good predictor of the data that will be collected in the 2018/19 cycle. Also, the intention behind using multiple cycles worth of data rather than a single cycle is to stabilise the values that are input into the design process. The means and variances are averaged across the two cycles and this provides some protection against the allocation changing suddenly from one year to the next, especially if a past year had more volatile data than the other year.

Another component of stabilising the input data is to minimise the impact of extreme and unrepresentative values of Internet Income (a skewed variable) on these estimated means and variances. This was done by using the winsorised values for Internet Income (which are calculated during the estimation processes of these two previous cycles) rather than the reported values for Internet Income.

For each cycle of BCS data, the stratum variances for total income received from internet orders were calculated by:

$$s_{yh}^2 = \frac{\sum_h w_{ih}(y_{ih} - \bar{y}_h)^2}{\left(\sum_h w_{ih}\right) - 1}$$

where

w_{ih} is the weight of unit i in stratum h

y_{ih} is the value of income received from internet orders for unit i in stratum h

$\bar{y}_h$ is the stratum mean of income received from internet orders for stratum h

The method used to calculate the stratum variances of the rates is the standard approximation (used in the ABS) for the population variance of a rate $R=Y/X$:

$$\frac{1}{X^2}\left(s_y^2 + R^2 s_x^2 - 2R s_{xy}\right)$$

The s_y^2 , s_x^2 and s_{xy} are calculated at the stratum level, while the X and Y values are calculated at the level for the estimate in the design constraint. For example, if the estimate is a State-level estimate, X and Y would be calculated at the State level, and if it was an Industry-level estimate, X and Y would be calculated at the Industry level. The totals are calculated with means from past data and "blown up" to the total using this year's stratum population counts.

The optimisation problem that needs to be solved, to produce a minimum allocation is:

$$\sum_{h=1}^{H} n_h$$

minimise

$$\sum_{h=1}^{H} \frac{N_h^2}{n_h}\left(1-\frac{n_h}{N_h}\right)\frac{1}{X^2}\left(s_y^2 + R^2 s_x^2 - 2R s_{xy}\right)$$

subject to $< v$

where n_h is the sample size and N_h is the population count in stratum *h*, for some variance constraint, *v*.

This is fortuitously very close to the minimisation problem that %mltalloc (the usual allocation minimisation macro used in the ABS) solves. In fact, the only difference is the calculation of the variance, where

$$\frac{1}{X^2}\left(s_y^2 + R^2 s_x^2 - 2R s_{xy}\right)$$

is needed instead of just s^2 . Since %mltalloc does not calculate these variances, but instead, relies on the user to calculate the variances for %mltalloc, we can substitute the complex rate variance calculation for the simpler sample variance formula for a mean and, in this way, persuade %mltalloc to solve the optimisation problem above.

5. Perform optimal allocation

This is performed using the %mltalloc macro. The macro requires the following sets of inputs:

1. The variance constraints from table 2.1.1,
2. Stratification as per Section 2.3.1,
3. The stratum means and variances calculated in Step 4 above;
4. Stratum population counts from the current frame, and
5. Minimum stratum sample sizes as specified in table 2.1.2;

The %mltalloc macro is set up to produce one output dataset containing a minimal allocation satisfying all the variance constraints. The output dataset contains a number of variables, including the two key variables: The stratum and allocation (the latter is always called _strsamplesize_).

This macro also produces estimates for the accuracy of key estimates at the required output levels (ie. the Expected SEs and RSEs).

6. Adjust for non-response

The non-response adjustment adjusts for expected non-response, as per specifications in Section 2.1.1. In previous cycles the procedure currently undertaken is to take the stratum allocation from the %mltalloc output (variable _strsamplesize_) and multiply it by 10/9 outside of %mltalloc. This adjustment was only applied to strata in the sampled sector. However, because the CE'd sector experiences non-response in practice (even despite IFU efforts) then the actual RSEs for the Internet Income estimates for some estimates were being greatly under- estimated. Therefore, for the 2017/18 cycle this non-response adjustment was run within %mltalloc itself.

It should be noted that running the non-response adjustment within %mltalloc, ie. in conjunction with the optimisation process, ends up not leading to any real difference to the resulting total allocation.

7. Undertake the relevant top-ups

The 2018/19 cycle of the BCS had two top-ups for the Core allocation:

1. A top-up that aimed to improve the accuracy of the Innovation rate over as wide a range over the population as population;
2. A more-targeted top-up for the Innovation rate , that was focussed on the areas of the population that aligned with the scope of the Community Innovation Survey;

The first, broad, top-up was aimed at improving the accuracy of estimates during a 'long-form' Innovation year. The I&T had expressed concerns that in these long-form years, which occur biennially, the Innovation rate estimates tend to be lower than in the short-form years. The second, more targeted top-up, was requested by DIIS because they had concerns about the validity of the Innovation rate estimates that were being used when making comparisons with the results of the CIS (as the Australian rate tends to be lower than expected).

Now, while a ratio estimate is not unbiased, and will be less biased as the sample increases, BSM feels that the improvement to bias from these top-ups will be secondary to the improvements in the accuracy of the Innovation rate estimates. Also, the improvement to accuracy can be quantified while this is not the case for the bias of the estimates. Therefore, these top-ups should probably be considered as servicing the accuracy of the Innovation rate estimates (ie. improving the confidence in the estimates) rather than their biasedness (ie. moving the estimates towards the correct value).

A single top-up value was applied to each relevant Core stratum through the following process:

1. For each of the two top-ups run the %mltalloc macro with the top-up condition added on or superseding the relevant accuracy constraint;
2. Obtain the stratum-level allocation based on these top-ups;
3. Compare the stratum-level allocation between:
    1. The original allocation;
    2. The allocation with the broad improvements to the Innovation rate;
    3. The allocation with the improvements in the Innovation rate for the CIS-aligned population;
4. Take the largest of these allocations to be final allocation for that stratum. It should be noted that this protects against a stratum having to lose sample because its allocation under the either the second or third allocation was lower than its initial allocation;

The final allocation for the BCS Core was 5,985 units. This is a decrease of 529 units compared with the 2017/18 cycle.

8. Finalise allocation

Approval for the final allocation by the I&T BSC was agreed at 9,609, which was within the initial request range of around 9,500 units. The following table presents the stratum-level allocation for the final BCS Core allocation.

Table 3.1.1 Division x Size level allocation for the 2018/19 BCS

S47E

BCS allocation and RSE information:

BCS Core allocation results_1819.xlsx

3.2 BLD allocation process

For the 2018/19 cycle there were two BLD panels that would still be operational, they were Panels J and K, which were first taken in the 2014/15 and 2015/16 respectively. In other words, there will be only Panel K in the 2019/20 cycle after which there will be no BLD panels anymore. Also, a total of 184 units that were In-Scope for both the BCS and BLD but dropped out from the BLD panel J for cycle 2016/17 were requested to be put back to the BLD for this cycle to support more informative longitudinal analysis with sufficient BLD units.

These panels were selected under the BLD scope which is not the same as the BCS Core scope. It was found that only selecting the units in the BLD panels that also fell under the BCS Core allowed for the Core, BLD panels as well as a reasonably sized Growth Hubs top-up to be selected. Table 3.2.1 lists the number of units that were selected in 2018/19 and how that number compares to the number of selections (which is the number of live units under the *BLD* scope) for the 2017/18 BLD sample design.

Table 3.2.1 BLD selections

Panel	First Cycle	Selections in 2017/18	Selections in 2018/19
J	2014/15	1,239	1,423
K	2015/16	1,017	1,017
	Total	3,365	2,440

There were a total of 2,440 units selected in 2018/19. This is a drop of 925 units from the 2017/18 cycle.

3.3 Growth Hubs Top-up allocation process

The initial request of a sample size of around 9,500 allowed for provision of the full Growth Hubs sample top-up.
The following document outlines some of the allocation options that were considered for the Growth Hubs top-up. The key takeaway points here are that:

- Allocation for the Growth Hubs could only be feasibly calculated at a level higher than the ANZSIC Class level, which is the level at which Growth Hubs are defined; and
- There was not much opportunity to have 'stratum'-level allocations to be higher than the minimum sample size (to cap the size of the weights of the selected units) without having very large allocations;

S22

Based on these constraints the I&T BSC agreed to have a Growth Hubs top-up where allocation was distributed at the Sub-division level, which meant that the expected overall size of the Growth Hub was 1,184 units. Table 3.3.1 provides the distribution of these units across the relevant Divisions.

S47E

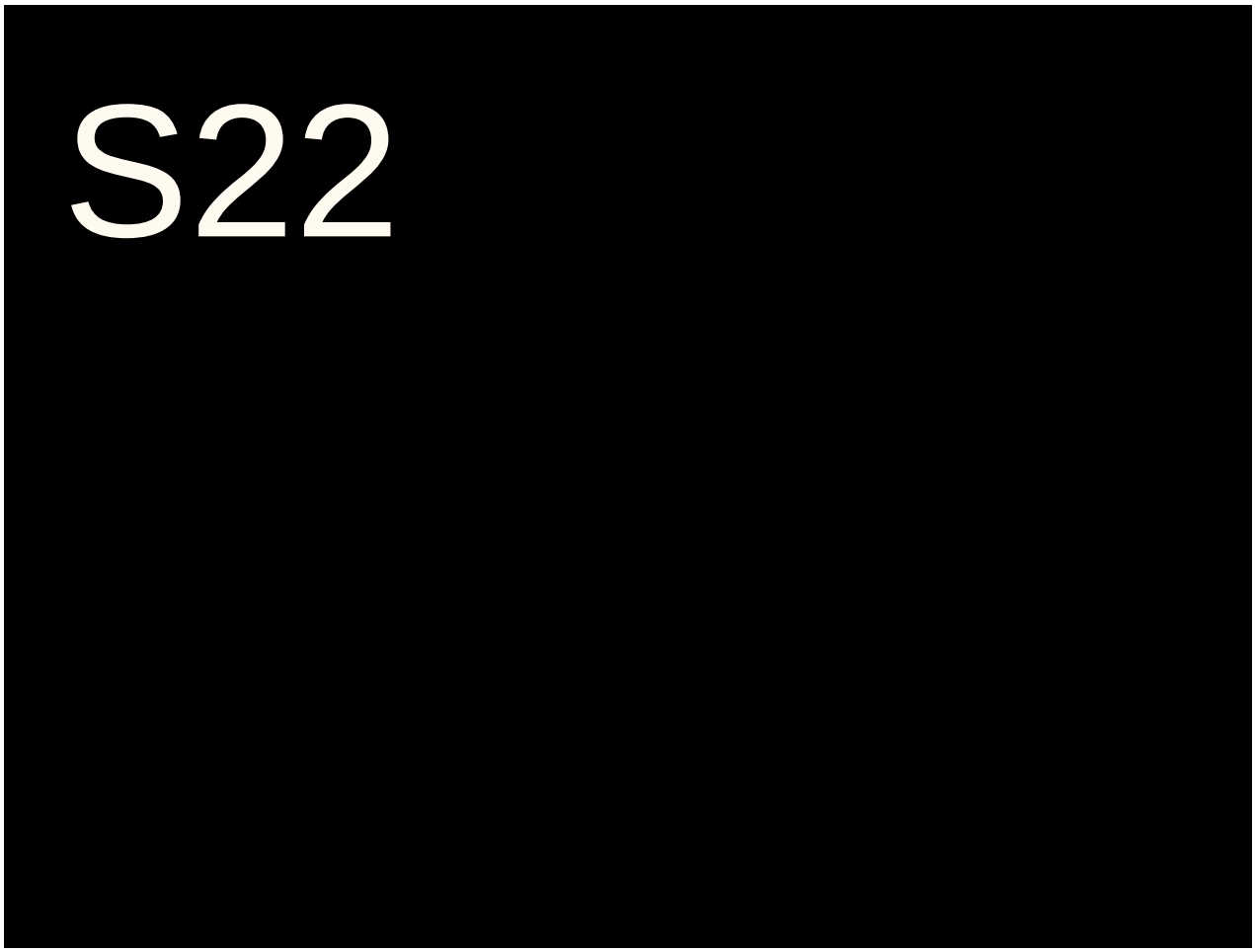
4. Selections

4.1 Procedure for selections

There are three sets of selections:

1. Select live BLD units;
2. Select Core sample by selecting units after trying to apply the forced rotation strategy to the 2018/19 BCS sample. This core sample will include the Innovation top-up. Due to how the selection processes for the Core and BLD are set up then overlap between these two allocations should be minimised by default;
3. Taking the Growth Hubs top-up by selecting units which minimise overlap with the BLD and Core samples;

While the BLD and Core units could be taken in any order the Growth Hubs top-up should be taken after both of the other sets of selections. This is because in order to maximise the amount of extra Growth Hubs units selected the units selected as part of either the BLD or BCS needed to be known first.



There are a number of assumptions underpinning this proposed process and these assumptions are listed in Section 4.2.

4.2 Assumptions for the Selection Process

Assumption 1: The Innovation top-ups are considered part of the Core design
Consequence: Units selected as part of the Innovation top-up will undergo the forced rotation that is applied where possible to units in the Core design
Rationale: This top-up is based on an aspect of the Core design (ie. the Innovation variable) and the allocation for this top-up was based on a similar process (ie. satisfying an accuracy constraint) to one

used in the Core design.

The Core design will continue aim to follow the same rotation strategy as the BCS did in previous cycles, which was following units for 3 years and then aiming to forcibly rotate those units out. This rotation strategy aimed to have a common sample of 66% every year and also to have a relatively stable level of rotation each year.

Also, it will make the selection process more straightforward in that it will make it easier to take Growth Hub selections from units that are not already selected as part of the Core design or the BLD (see below).

Assumption 2: Growth Hub selections will look to minimise overlap with Core and BLD selections
Consequence: The total sample size will be larger than if Growth Hub selections overlapped with Core and BLD selections
Rationale: The purpose behind the Growth Hubs top-up is different to the purpose behind both the Core selections and the purpose behind the BLD selections. Therefore, these selections should be based on different sets of units and it seems like it is possible to minimise overlap between these sets of selections and still have the total allocation fall under the collective maximum sample size.

Also, because the Growth Hubs stratification was not an explicit part of the Core design and because Growth Hubs only make up a small proportion of the Core's scope then it is likely that not many units in Growth Hubs would have been selected as part of the Core allocation. This means that there would not have been much of a saving in sample size from trying to maximise overlap between the Growth Hubs and Core selections.

Another benefit of minimising overlap between the Growth Hubs and the Core selections is that it minimises the chance of units selected in the Growth Hubs being selected in upcoming cycles of the Core allocation.

It should be noted that this means that any units within Growth Hubs that are selected as part of the Core or BLD allocations are not considered to be units selected as part of the Growth Hubs top up. These units will therefore not receive a Growth Hubs top-up weight and consequently not contribute to Growth Hubs (top-up) estimates.

Assumption 3: For Divisions which contain Growth Hubs the Core allocation is calculated at the Division level but the Core selections will be taken at the Sub-division level
Consequence: For the Core design the number of selections within Growth Hub Divisions are fixed at a lower level of industry classification than for non-Growth Hub Divisions
Rationale: Given the sample size requirements for the overall BCS design that Growth Hubs top-up selections will likely be taken at the Sub-division level. This is a lower level of industry classification than the level at which the Core allocation was calculated (ie. the Division level).

When calculating the integrated weights (eg. the weight when multiple sets of samples contribute to the same estimate) for units these weights needs to be based on the population and sample counts at the finest level of stratification of the collections being integrated. In other words, the integrated weights will need to be set at the Sub-division (x Size) level.

If selections for the Core are taken without taking the finer stratification into account then the resulting integrated weights will be biased. However, if the number of selections within each Sub-Division are fixed (eg. by proportionally allocating selections across the constituent Sub-divisions) then by taking Core selections randomly within the Sub-divisions the integrated weights will be unbiased. The alternative way to calculate unbiased integrated weights would be to incorporate the Growth Hubs stratification into the Core sample design but this is undesirable.

Therefore, for the purposes of taking selections and then obtaining integrated weights within Growth Hub Divisions the selections will actually be taken at the Sub-division level. It should be noted that the weight pertaining to just the Core selections would be unaffected. Also, the Sub-Divisions will share the same overlap and selection ranges as the Divisions where possible.

By way of example, consider a stratum within a Growth Hubs Division has a population of 10,000 and this Division contains three Sub-divisions with Sub-Division 1 containing 5,000 units, Sub-Division 2

containing 3,000 units and Sub-Division 3 having 2,000 units. If this stratum was allocated 120 units then under random selections it would be expected that Sub-Division 1 would have 60 units selected, Sub-Division 2 would have 36 units selected and Sub-Division 3 would have 24 units selected.

S47E

Assumption 4: The Growth Hubs top-up will include units that are not within Growth Hubs
Consequence: The Growth Hubs top-up is less efficient
Rationale: In order to properly control for overlap between the Core Selections and the Growth Hubs selections the overlap control algorithm needed to select Growth Hubs based off the same frame used to select the Core selections. This meant that units within Sub-Divisions that contained Growth Hubs but that were not actually in the appropriate Growth Hubs ANZSIC Classes were eligible for selection as part of the Growth Hubs top-up. Setting an allocation, and taking selections, at this broader level brings inefficiencies into the allocation because there is now less control over being able to selecting Growth Hubs from the relevant population (and still taking selections randomly). In other words, there was a difference between the desired scope for the Growth Hubs top-up and the actual coverage for the Growth Hubs top-up.

However, it ended up being the case that there were 758 Growth Hubs selected as part of the overall top-up of 1,184 units which corresponds to 64% of the top-up being Growth Hubs.

4.3 Outputs and Summary of selections

The following link leads to a document that contains the selections for the 2018/19 BCS. For this cycle this means that the selections include the:

- Core allocation: 5,985 units;
- BLD units: 2,440 units across two panels (the panels for the cycles 2014/15 - 2015/16 inclusive);
- Growth Hubs top-up: 1,184 units;

for a total of 9,362 units (after accounting for overlapping units between the various sets of selections).

Other points of interest around these selections are that there are:

- 3,517 units continuing on in the Core allocation. This translates to an overlap of 59%;
- 49 units that fall within both 2018/19 Core allocation and the BLD panels. All of these common units are from the continuing BCS units;

Table 4.3.1 provides the breakdown for the overall BCS sample:

Table 4.3.1 breakdown for the overall BCS sample

Summary 2018/19	Sample	Overlap with 17/18 (%)
bcs	5,985	
BCS 17/18 and 16/17 - in common	3,517	59
bld1516	1,017	
bld1415	1,423	
bld1314	0	
bld bcs - in common	49	
growth_hub	1,184	64
growth_hub and bcs - in common	190	
Combined total	9,362	

S22

S22


S22

S22

S22

Summary Document for REACS 19/20

1. Introduction

The aim of this document is to act as a reference for all major pieces of documentation, inputs and outputs associated with the REACS 19/20 sample design. It also summarises the main differences between the 19/20 sample design and the previous cycle of REACS, which was designed in the 18/19 financial year. This document contains links to:

1. Sample design specifications;
2. Key outputs, both during and at the end of the REACS 19/20 sample design process;
3. The SAS programs used for the sample design, and documentation for those SAS programs;

Special Note: Due to the potential impact of COVID-19 on the response rate of REACS 19/20, two fallback options were developed in addition to the business-as-usual (BAU) design. These fallback options were developed to address potential reduced response rates and limited resources available for intensive follow-up (IFU) as a result of COVID-19. A two tier approach was implemented. The aim of the tier approach was to ensure National level estimates are kept intact, should response rates not reach the BAU expectation. This two tier approach provides a subset of the full selection that can be deselected from IFU and possibly excluded from estimation to reduce non-response bias. The tier 2 option ensured National level estimates by applying constraints at the National level only. The tier 1 option also ensured National level estimates but provided a further reduction in sample by applying constraints at the National level for a reduced set of priority commodities as identified by the EASB. The list of the 14 priority design items are in Table 1.

Table 1: List of Priority Design Items

2. Sample design work and specifications

Project Plan for REACS 19/20: [redacted] (Subject: Plan for REACS 2019/20; Database: [redacted] S22 [redacted])
This document breaks down the work that was intended to be done as part of REACS 19/20 sample, in it notes the initiatives intended to be implemented for this cycle, in particular the relaxation of the national level RSE constraints from [redacted]% to S47E [redacted]%. Additionally, a second initiative was to include a set of units which were identified through a BSM frame maintenance model into the sample as a stand alone CE stratum. The frame maintenance model results were not implemented for this year, though, due to COVID-19 impacts.

Maximum sample size: For REACS 19/20 PEASB requested a maximum sample size of 26,000 units. This is lower than what was used in earlier REACS cycles and is mainly due to the fact that we have relaxed the national level RSE constraints The achieved sample allocation was 25,642 for BAU; 23,375 for Tier 2; and 18,823 for Tier 1.

Accuracy constraints:
The Desired RSEs are the same as those for REACS 18/19, except for the national level constraints which were relaxed from and from S47E [redacted]% to S47E [redacted]%, following the SSC recommendations. Similarly the sizevar variable was relaxed from S47E [redacted] to S47E [redacted] at the State and NRM (National Resource Management) level and area of holding (AOH) was relaxed from S47E [redacted] to S47E [redacted] at the State level. Details of the constraints are provided in this document. [redacted] (Subject: REACS 19/20 Sample Allocation; Database: [redacted] S22 [redacted]) . The full suite of Desired RSEs, differentiated by Commodity or Priority Group, is provided in the table below.

S47E

^ These constraints aren't set as part of the initial allocation process and are usually only set after consultation with the BSC. Also, the constraints are typically only applied to particular commodities rather than all commodities within that Group.

The process for allocation begins by excluding commodity restraints at the State and NRM levels. After the assessment of the initial allocation from this process, a set of important state by commodity constraints were implemented in consultation with PEASB.

Design commodities:
This spreadsheet contains the list of possible design commodities at the National, State and NRM levels for the REACS 19/20 sample design.

Summary of Sample Design 19_20_Final.xlsx

Further constraints were added to state by commodity items which were unreasonably high after national level constraints were applied are listed below.

- NSW: Dairy Cattle, Total ('M20'), Other Nut ('M33'), Oats for grain ('M35'), Peaches ('M42'), and Sorghum for grain ('M51')
- Vic: Potatoes ('M47'), and Strawberries (area of bearing age) ('M52')
- Qld: Cotton ('M17')
- SA: Canola ('M11'), Mandarins ('M27')
- WA: Nurseries ('M34')
- Tas: Apples ('M3'), Carrots ('M13'), Onions ('M13')

For further details on the State- and NRM-level constraints that were set the reader is directed to the Sample Allocation document in Section 2.

3. Outputs from REACS 19/20 sample design

Variance Outliers: [redacted] (Subject: Variance outliering REACS 19/20; Database: [redacted] S22 [redacted])
Units that reported values that were considered to be unrepresentative both within a particular year of design data and across all four years of design data compared to those reported by other units in the population were set as outliers. The design data for REACS 19/20 was REACS data for the 16/17, 17/18 and 18/19 cycles in addition to the 15/16 Ag Census data.

The tracking document provides a list of all units set as surprise outliers during the sample design process. There is also documentation of what commodity values caused these units to be set as outliers. This outliering process was done over several stages, which differed by whether the intention was to simply lower the overall sample size or to get the Expected RSEs of certain estimates below the Desired RSEs. For further details on the different stages of outliering the reader is directed to the sample allocation document provided below.

Sample allocation & final accuracy constraints for REACS 19/20:
[redacted] (Subject: REACS 19/20 Sample Allocation; Database: [redacted] S22 [redacted])
This document contains details regarding the sample allocation for REACS 19/20 that ended up being approved for use by PEASB. The set of RSE constraints that were used to generate this allocation can also be found within this document.

A spreadsheet containing the stratum and state level allocations as well as the Expected RSEs for all potential design commodities at the National, State and NRM levels can be found at the bottom of this document.

Overlap control strategy:

[redacted] (Subject: REACS 2019/20 Sample Selections; Database: [redacted] S22 [redacted])
Historically, REACS has only looked to maximise overlap with the previous cycle and this was in service of having a good quality of movement estimates. This was changed for the 16/17 REACS to allow some rotation of sampled sector units over time. For the 19/20 REACS we have continued to aim for a balance in terms of having about a two-thirds overlap in sample with the 18/19 REACS.

For the REACS 19/20 there were sets of selections for which overlap was controlled:
Maximise overlap with units selected in the last two-thirds of the REACS 18/19 (to help with overlap with previous selections);
Minimise overlap with units selected in both the REACS 18/19 and 17/18 ;
Minimise overlap with units selected in any of the REACS 16/17, REACS 15/16 or REACS 14/15.

Summary of Tier 1 and Tier 2 allocation and overlap strategy :
A similar overlap strategy was implement for the Tier 1 and Tier 2 subsamples. The Tier 2 allocation was assigned by setting the BAU allocation for each stratum as the maximum sample for the Tier 2 stratum. Similarly, the Tier 1 allocation was assigned by setting the Tier 2 allocation for each stratum as the maximum sample for the Tier 1. This resulted a reduced availability of sample for Tier 1, so to assist in the allocation process the minimum sample size for each stratum was reduced from 15 down to 10.
The program for this process is in this SAS code:. [redacted] S22 [redacted]
The overlap strategy for selection for Tier 2 units maximised the overlap between Tier 2 and BAU selections; and the selections for Tier 1 maximised overlap between Tier 1 and Tier 2 selections. There was no need for the explicit overlap between 19/20 and previous cycles because this was already achieved in the BAU overlap control strategy described above.

Approval of sample allocation by PEASB: [redacted] (Subject: Re: Notification: Allocation summary 19/20; Database: [redacted] S22 [redacted])

Sample Selections: [redacted] (Subject: REACS 2019/20 Sample Selections; [redacted] S22 [redacted])
This document contains:

- A copy of the frame that flags units that have been selected for REACS 19/20 and
- Some characteristics of the selected sample (sample counts, overlap and despatch rate);

Gold Provider strategy:
During the 2018/19 cycle a live trial of a Gold Provider (GP) strategy was conducted. Details of the assessment of the live trial are summarised here: [redacted] (Subject: Assessment on the Live Trial of the GP Strategy; [redacted] S22 [redacted]).
Results from the live trial assessment showed the GP strategy was effective in saving IFU resources spent on the GP units by delaying the commencement of IFU contacts towards them for 2 months, without causing a reduction in response rate for these units. Based on these results the subject matter area were keen to implement the GP strategy in the 2019/20 REACS. Details of the implementation of the GP strategy are here: [redacted] (Subject: GP Model for REACS 19/20; [redacted] S22 [redacted])

4. SAS programs and data used for sample design

[redacted] S22

Frame: The frame that was used in the sample design for REACS 19/20 was based on the frame S22 within the parent directory (S22). This frame includes all units considered in-scope with an economic size threshold of $40,000 and is not stratified. The stratified frame, which also distinguishes between whether or not a unit is in scope, is called S22 and can be found within the output directory.

S22

S22

S22

S22

REACS 19/20 Sample Allocation

Executive Summary

Business as usual (BAU) allocation:

The REACS 19/20 sample design has returned an allocation of 25,642 total units. This allocation is expected to result in National RSEs which are less than or equal to the RSEs desired by PEASB for all 50 commodities designed for at the National level. It is also to meet the desired accuracy for the majority of State-level estimates and NRM-level estimates while only explicitly designing for a small number of these lower level constraints. The sub-national estimates with RSEs that are expected to be larger than the RSEs desired by PEASB are all either small contributors to the national total or where the expected accuracy is not all that much larger than the desired accuracy. There were 1275 strata, with 436 of these deliberately CEd and an additional 204 fully sampled through the allocation.

An additional component to REACS 19/20 was to provide fallback options to address potential impacts of COVID-19, in particular reduced response rates and limited resources available for IFU. A two tier approach was implemented in addition to the business-as-usual (BAU) allocation described above. The aim of the tier approach was to focus the allocation to ensure National level estimates are kept intact, should response rates not reach the BAU expectation. This two tier approach provides a subset of the full selection that can be deselected from IFU and possibly excluded from estimation to reduce non-response bias.

Tier 2 allocation:

The tier 2 option applied constraints at the National level only. This sample design returned an allocation of 23,375 total units. This allocation is expected to result in National RSEs which are less than or equal to the RSEs desired by PEASB for all 50 commodities designed for at the National level.

Tier 1 allocation:

The tier 1 option applied constraints at the National level only for a reduced set of priority commodities as identified by the EASB, The priority commodity items are listed in Table 1. This sample design returned an allocation of 18,824 total units. This allocation is expected to result in National RSEs which are less than or equal to the RSEs desired by PEASB for the 14 priority design items at the National level.

Table 1: List of Priority Design Items

Design Process

The design process for REACS 19/20 has followed a similar process to recent cycles. It involved the following:

- An initial allocation was produced by using:
- The march production frame;
- Means and variances from the previous four cycles of REACS (which in this case included REACS 15/16, 17/18, 18/19 and the 16/17 Ag Census);
- The Desired RSEs from the previous design; and
- National-level constraints relaxed to S47E % for all commodities;

These means and variances represent the expected means and variances that would be

achieved by REACS 19/20 in practice, under the assumption that recent historical data would be a good predictor of future responses. Four years of historical data was used to smooth out the impact of any unusual data. Using only national constraints prevents the design from becoming inefficient through being overly constrained;

- This initial allocation had a large sample size (greater than 30,000) so a process of outliering units was undertaken, where unrepresentative units that were having a large impact on sample size were either removed, or where the CE cut-offs for commodities were adjusted. There were a total of 21 units set as outliers.

Design Constraints

The final set of Desired RSEs, or maximum allowable RSEs, for various commodities at various output levels are given in Table 2 for the BAU design; Table 3 for Tier 2; and in Table 4 for Tier 1.

Table 2. Desired RSEs for REACS 19/20 - **BAU**

S47E

Table 3. Desired RSEs for REACS 19/20 - **Tier 2**

Table 4. Desired RSEs for REACS 19/20 - **Tier 1**

Design Performance

A summary of REACS 19/20 sample design, as it currently stands, is presented in the following tables.

Table 2. Summary of the initial REACS 19/20 sample design - National level

S47E

Table 3. Summary of the initial REACS 19/20 sample design - Sub-national levels

S47E

S47E

The main results for the BAU design from Tables 2 and 3 are that:

- All Expected RSEs for National-level commodity estimates are lower than the corresponding National-level Desired RSEs, except one commodity, i.e., pineapple whose expected RSE is S47E%;
- Approximately 59% of the State-level estimates and 83% of the NRM-level estimates have expected RSEs lower than the Desired RSEs. The state-level result is lower than what was achieved in the REACS 18/19 sample design (70%), while the NRM result is similar to the 18/19's (83%);

The lower-on-average achieved RSEs for State and NRM levels are not surprising given that we don't generally aim to design for these levels. Generally, when the achieved RSEs exceed the Desired RSE, they do not exceed by much. There are a small number of State level RSEs where this is not the case. Advice was sought from PEASB and a set of 14 commodities were added in as explicit State constraints. (Refer to column "Priority_for supp" in sheet "Expected RSEs - State" in the spreadsheet below - items constrained are identified by "1").

The expected RSEs for all the National-, State- and NRM-level estimates of the required commodities are listed in the following spreadsheet. This spreadsheet also includes a stratum-, state- and sector-level breakdowns of the allocation.

Summary of Sample Design 19_20_Final.xlsx

S22

[redacted]

Summary Document for REACS 18/19

1. Introduction

The aim of this document is to act as a reference for all major pieces of documentation, inputs and outputs associated with the REACS 18/19 sample design. It also summarises the main differences between the 18/19 sample design and the previous cycle of REACS, which was designed in the 17/18 financial year. This document contains links to:

1. Sample design specifications;
2. Key outputs, both during and at the end of the REACS 18/19 sample design process;
3. The SAS programs used for the sample design, and documentation for those SAS programs;

2. Sample design work and specifications

Project Plan for REACS 18/19: [redacted] (Subject: Plan for REACS 2018/19 sample design; Database: [redacted])
This document breaks down the work that was intended to be done as part of REACS 18/19 sample, in particular it notes the two major initiatives intended to be implemented for this cycle. The first initiative is to relax the national level RSE constraints from [redacted]% to [redacted]%, and the second initiative is to include a set of units which were identified through a BSM frame maintenance model into the sample as a stand alone CE stratum.

Maximum sample size: For REACS 18/19 PEASB requested a maximum sample size of 28,000 units. This is lower than what was used in earlier REACS cycles and is mainly due to the fact that we have relaxed the national level RSE constraints. The achieved sample allocation was 27,159.

Accuracy constraints:
The Desired RSEs are the same as those for REACS 16/17, except for the Area of Holding (AOH) constraints which were tightened from [redacted]% down to [redacted]% at the national level and from [redacted]% to [redacted]% at the state level. Details of the constraints are provided in this document [redacted] (Subject: Allocation summary 18/19; Database: [redacted]). The full suite of Desired RSEs, differentiated by Commodity or Priority Group, is provided in the table below.

[redacted]

^ These constraints aren't set as part of the initial allocation process and are usually only set after consultation with the BSC. Also, the constraints are typically only applied to particular commodities rather than all commodities within that Group.

There was a change in the design process for allocation in 18/19, where commodity restraints were removed during the allocation process. Important state by commodity constraints were implemented in consultation with EASB.

Design commodities:
This spreadsheet contains the list of possible design commodities at the National, State and NRM levels for the REACS 18/19 sample design.

REACS 1819 Sample Design specs - Final.xls

Further constraints were added to state by commodity items which were unreasonably high after national level constraints were applied are listed below.

- NSW: Cherries ('M15'), Macadamias ('M25'), Melons (all types) ('M30'), Nurseries ('M33'), Pigs,Total ('M41'), Sorghum for grain ('M47'), Total area orchard fruit and/or nuts ('M56')
- Vic: Cherries ('M15'), Cut flowers ('M18'), Nurseries ('M33'), Potatoes ('M43'), Strawberries (area of bearing age) ('M48')
- Qld: Lettuce (all types) - Outdoor and Undercover ('M24'), Nurseries ('M33'), Strawberries (area of bearing age) ('M48')
- SA: Mandarins ('M26'), Oranges ('M37')
- Tas: Cherries ('M15'), Onions ('M36')

For further details on the State- and NRM-level constraints that were set the reader is directed to the Sample Allocation document in Section 2.

3. Outputs from REACS 18/19 sample design

Frame checks: Frame check was conducted before the start of the sample design process. There were a total of 138804 units on the frame, a slight decrease of 1960 from 140764 units last cycle. For more details of the frame check results, see the linked document: [redacted] (Subject: Quality Check on the REACS 2018/19 Frame; [redacted] S22 [redacted])

Frame quality improvements: Additional improvements specific to 2018/19 sampling and future quality of the frame were:
A total of 217 units that were identified using the BSM frame prediction model were also added to the final sample as a stand-alone CE stratum. For more details of the frame prediction model, see the linked document:
[redacted] (Subject: Using the frame prediction model to improve REACS frame coverage/scoping; Database: [redacted] S22 [redacted]

Variance Outliers: [redacted] (Subject: Tracking REACS Sample Design 2018/19 Progress_Allocation and Outliering; Database: M[redacted] S22 [redacted])
Units that reported values that were considered to be unrepresentative both within a particular year of design data and across all four years of design data compared to those reported by other units in the population were set as outliers. The design data for REACS 18/19 was REACS data for the 14/15, 16/17 and 17/18 cycles in addition to the 15/16 Ag Census data.

The tracking document provides a list of all units set as surprise outliers during the sample design process. There is also documentation of what commodity values caused these units to be set as outliers. This outliering process was done over several stages, which differed by whether the intention was to simply lower the overall sample size or to get the Expected RSEs of certain estimates below the Desired RSEs. For further details on the different stages of outliering the reader is directed to the sample allocation document provided below.

Sample allocation & final accuracy constraints for REACS 18/19:
[redacted] (Subject: Allocation summary 18/19; Database: [redacted] S22 [redacted])
This document contains details regarding the sample allocation for REACS 18/19 that ended up being approved for use by PEASB. The set of RSE constraints that were used to generate this allocation can also be found within this document.

A spreadsheet containing the stratum and state level allocations as well as the Expected RSEs for all potential design commodities at the National, State and NRM levels can be found at the bottom of this document.

Overlap control strategy:
[redacted] (Subject: REACS 2018/19 Sample Selections; Database: [redacted]
S22 [redacted])
Historically, REACS has only looked TO maximise overlap with the previous cycle and this was in service of having a good quality of movement estimates. The 16/17 cycle of REACS looked to minimise overlap with previous REACS cycles in an attempt to lower overall respondent burden. For the 18/19 REACS we have continued to aim for a balance in terms of having about a two-thirds overlap in sample with the 17/18 REACS.

For the REACS 18/19 there were sets of selections for which overlap was controlled:
Maximise overlap with units selected in the last two-thirds of the REACS 17/18 (to help with overlap with previous selections);
Minimise overlap with units selected in both the REACS 17/18 and 16/17 ;
Minimise overlap with units selected in any of the REACS 1617, REACS 14/15 or REACS 13/14

Approval of sample allocation by PEASB: [redacted] (Subject: Allocation summary 18/19; Database: [redacted]
S22 [redacted])

Sample Selections: [redacted] (Subject: REACS 2018/19 Sample Selections; Database: [redacted]
S22 [redacted])
This document contains:

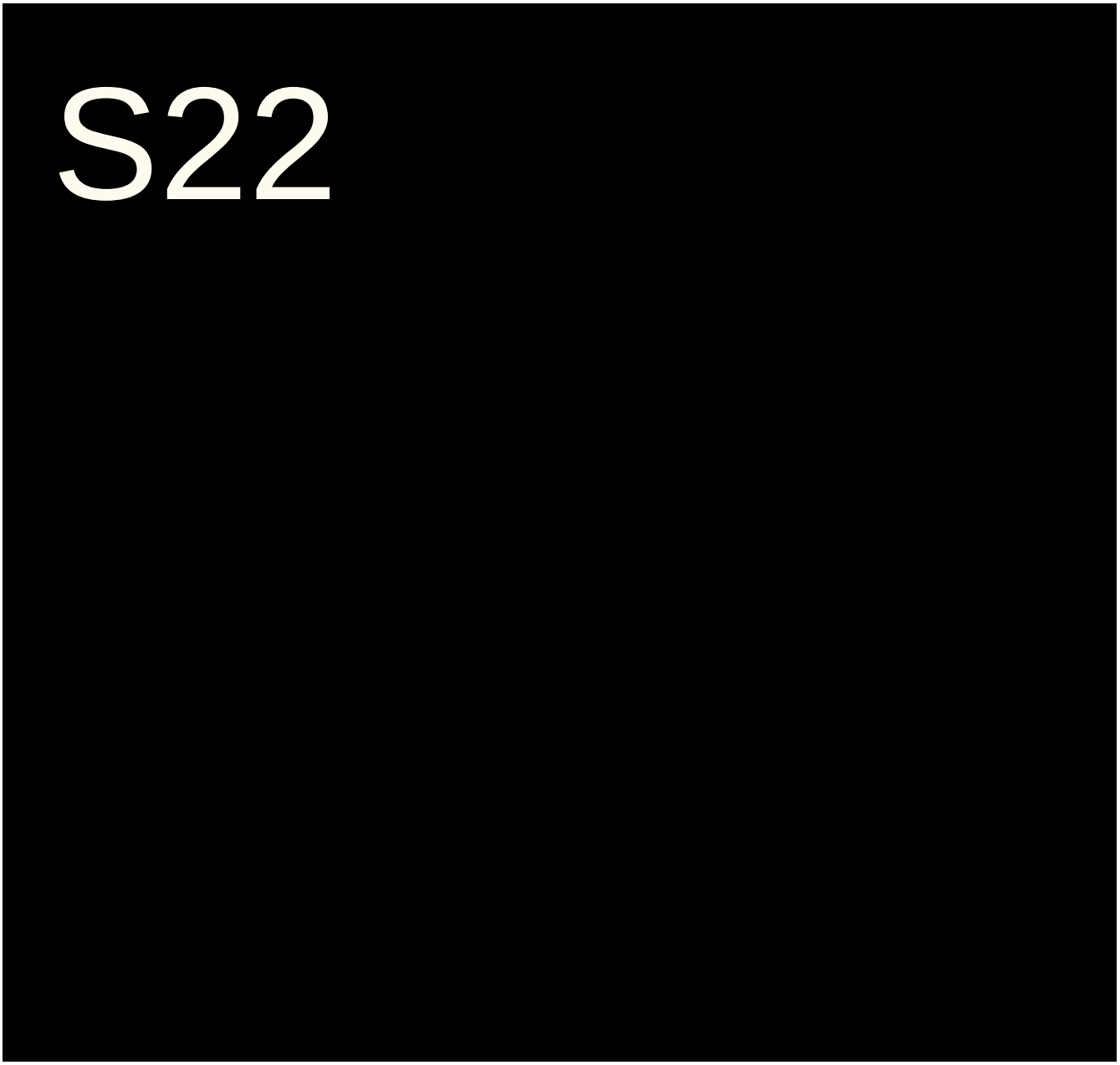
- A copy of the frame that flags units that have been selected for REACS 18/19 and
- Some characteristics of the selected sample (sample counts, overlap and despatch rate);

4. SAS programs and data used for sample design

[redacted]
S22

Frame: The frame that was used in the sample design for REACS 18/19 was based on the frame S22 [redacted] within the parent directory S22 [redacted]. This frame includes all units considered in-scope with an updated economic size threshold of $40,000 and is not stratified. The stratified frame, which also distinguishes between whether or not a unit is in scope, is called S22 [redacted] and can be found within the output directory.

[redacted]
S22


S22

S22

S22

REACS 18/19 Sample Allocation

Executive Summary

The REACS 18/19 sample design has returned an allocation of 27,159 total units. This allocation is expected to result in National RSEs which are less than or equal to the RSEs desired by EASB for all 50 commodities designed for at the National level. It is also to meet the desired accuracy for 63% of State-level estimates and 83% NRM-level estimates while only explicitly designing for a small number of these lower level constraints. The sub-national estimates with RSEs that are expected to be larger than the RSEs desired by PEASB are all either small contributors to the national total or where the expected accuracy is not all that much larger than the desired accuracy. There were 1267 strata, with 425 of these deliberately CEd and an additional 228 fully sampled through the allocation.

Design Process

The design process for REACS 18/19 has followed a similar process to recent cycles. It involved the following:

- An initial allocation was produced by using:
- The march production frame;
- Means and variances from the previous four cycles of REACS (which in this case included REACS 14/15 - 17/18 as well the 15/16 Ag Census);
- The Desired RSEs from the previous design; and
- National-level constraints relaxed to [redacted]% for all commodities;
  These means and variances represent the expected means and variances that would be achieved by REACS 18/19 in practice, under the assumption that recent historical data would be a good predictor of future responses. Four years of historical data was used to smooth out the impact of any unusual data. Using only national constraints prevents the design from becoming inefficient through being overly constrained;
- This initial allocation had a large sample size (greater than 32,000) so a process of outliering units was undertaken, where unrepresentative units that were having a large impact on sample size were either removed, or where the CE cut-offs for commodities were adjusted. There are a total of 37 units set as outliers currently, which is the same as last year's.

Design Constraints

The final set of Desired RSEs, or maximum allowable RSEs, for various commodities at various output levels are given in Table 1

Table 1. Desired RSEs for REACS 18/19

S47E

Design Performance

A summary of REACS 18/19 sample design, as it currently stands, is presented in the following tables.

Table 2. Summary of the initial REACS 18/19 sample design - National level

S47E

Table 3. Summary of the initial REACS 18/19 sample design - Sub-national levels

S47E

The main results from Tables 2 and 3 are that:

- All Expected RSEs for National-level commodity estimates are lower than the corresponding National-level Desired RSEs, except one commodity, i.e., pineapple whose expected RSE is S47E %;
- Approximately 70% of the State-level estimates and 83% of the NRM-level estimates have expected RSEs lower than the Desired RSEs. The state-level result is slightly higher than what was achieved in the REACS 17/18 sample design (67%), while the NRM result is similar to the 17/18's (84%);

The slightly higher-on-average achieved RSEs for State and NRM levels are not surprising given that we don't generally aim to design for these levels, and we relaxed the National RSE constraints this year. Generally, when the achieved RSEs exceed the Desired RSE, they do not exceed by much. There are a small number of State level RSEs where this is not the case - PEASB advice will be sought as to which of these should be added in as explicit State constraints.

The expected RSEs for all the National-, State- and NRM-level estimates of the required commodities are listed in the following spreadsheet. This spreadsheet also includes a stratum-, state- and sector-level breakdowns of the allocation.

Summary of Design 18_19_Final.xlsx

S22

S22

Summary Document for REACS 17/18

1. Introduction

The aim of this document is to act as a reference for all major pieces of documentation, inputs and outputs associated with the REACS 17/18 sample design. It also summarises the main differences between the 17/18 sample design and the previous cycle of REACS, which was designed in the 16/17 financial year. This document contains links to:

1. Sample design specifications;
2. Key outputs, both during and at the end of the REACS 17/18 sample design process;
3. The SAS programs used for the sample design, and documentation for those SAS programs;

2. Sample design work and specifications

Project Plan for REACS 17/18: [redacted] (Subject: Plan for REACS 2017/18 sample design (draft); Database: [redacted] S22)

This document breaks down the work that was intended to be done as part of REACS 17/18 sample, in particular it notes the major redesign objectives aimed at updating the stratification as a follow-up to the 2016 Ag Census. The last major re-stratification was 2011/2012. The two areas of focus for re-stratification were updating the size measures and investigating the treatment of multi-industry units.

Updating size measures

Since 2016 all Ag collections have moved from classifying units on the Ag frame as being in-scope if they have an economic size of at least $5,000 to the higher threshold of $40,000. Therefore, REACS 16/17 represents the first cycle of REACS with this new size scope. The method for determining which of EVAO and SBTS to use a unit's size value was also changed in order to account for the new size scope.

Preliminary results from work investigating improving the ~~size stratification of~~ SBTS model are provided. [redacted] (Subject: Preliminary Results for the Size Investigation; Database: [redacted] S22

Optimal size stratification across **regind** (region by industry) level was determined using the Lavallee-Hidiroglou (L-H) method. For this application, where appropriate, the initial endpoints for the strata were calculated by the cumulative square root method. These endpoints were then improved iteratively as guided by a Neyman allocation. The ABS macro associated with this method is flexible in that different initial endpoint methods are available, along with capability to change the parameters associated with the allocation that guides the iterations. The standard rules established from this method are provided in the table below. Note that post-allocation investigations indicated a potential for the reduction of the number of units in the size 4 strata. For example there were some reginds containing ~850 units that were adjusted manually to accommodate four size strata.

Number of units in regind	Number of size strata
<120	1
120 - 300	2
301-1000	3
>1000	4

Updating region and industry

Industry level stratification was determined by identifying business activities that tend to occur

together within units (multiple-activity units). A hierarchical cluster analysis using Wards method was conducted to determine industries that could be grouped together. These were initially determined at the state level, and adjusted where needed for regions (NRMs) within states. Some comments regarding the industry groupings at the state level is provided here [redacted] (Subject: Re: Notification: Allocation summary 17/18; [redacted]). Attached below is a summary document outlining the methods and the process used to determine the region (NRM) by industry classifications. Included are links to an R script used to execute the exploratory methods.

Region by industry strata.docx

The set of regional by industry strata derived from the analysis are provided in this spreadsheet:

Regional by Industry Strata.xlsx

A review of CE bounds was also undertaken and resulted in the following conditions.
EVAO cut-offs: This cut-off was determined by finding the minimum EVAO from the following set of bounds:

- The state by industry group 99 percentile cut-off
- 20% of the total EVAO at state by ANZSIC group level
- 10% of the total EVAO at state by ANZSIC class level
- 8% of the total EVAO at regional (NRM) by ANZSIC class level
- $6M (Note: The 17/18 REACS design increased the standard EVAO cut-off from $3M to $6M)

Commodity CE cut-offs:

- For the national and state level, the cut-offs were calculated as the 99th percentile
- For NRM level, the cut-offs were the 95th percentile
- Cut-offs for 17/18 were capped to change less than 50% from the previous cut-offs
- If there were less than 20 units contributing to the calculation of the cut-off, the old cut-off was used
- The same variables and levels were used for the new cut-offs

The exception to the above is the new NRM level area of holding cut-offs, which were identified manually by looking at plots.

There was some concern from EASB regarding the size of the CE'd sector. As a result of the redesign and resetting of boundaries the CE'd sector was successfully reduced in the 17/18 sample allocation from an initial 5,487 to 3,853.

Maximum sample size: For REACS 17/18 EASB requested a maximum sample size of 30,000 units. This is lower than what was used in earlier REACS cycles and is somewhat based on earlier predictions by BSM regarding the expected sample size under the new scope. The achieved sample allocation was ~~29,166~~ 29,315.

Accuracy constraints:
The Desired RSEs are the same as those for REACS 16/17, except for the Area of Holding (AOH) constraints which were tightened from S47E % down to [redacted]% at the national level and from S47E % to S47E % at the state level. Details of the constraints are provided in this document [redacted] (Subject: Allocation summary 17/18; Database: [redacted]). The full suite of Desired RSEs, differentiated by Commodity or Priority Group, is provided in the table below.

S47E

[redacted]

^ These constraints aren't set as part of the initial allocation process and are usually only set after consultation with the BSC. Also, the constraints are typically only applied to particular commodities rather than all commodities within that Group.

There was a change in the design process for allocation in 17/18, where commodity restraints were removed during the allocation process. Important state by commodity constraints were implemented in consultation with EASB.

Design commodities: [redacted] (Subject: REACS 1718 Sample Design specs spreadsheet; Database: [redacted])
This spreadsheet contains the list of possible design commodities at the National, State and NRM levels for the REACS 17/18 sample design.

Further constraints were added to state by commodity items which were unreasonably high after national level constraints were applied are listed below.

- SA and WA: apples
- WA: dairy cattle (total)
- Vic, WA and Tas: Grapes
- WA: Total fruit & nuts (exc grapes)
- WA: Total veg

For further details on the State- and NRM-level constraints that were set the reader is directed to the Sample Allocation document in Section 2.

Stratification:
Changes to stratification as a result of the redesign are described at the beginning of this section Details on these changes can be found in the programs 'setce1' and 'Create updated cut-offs' within the sample design SAS project. The location of this project can be found in Section 3.

2. Outputs from REACS 17/18 sample design

Frame checks: EASB conducted checks of the frame The frame was modified where 76 extra units were added (1 of extra was later removed as sub-annual unit) and 64 units removed (one of them came from extra units). The final frame had 141,170 units. See the linked document: [redacted] (Subject: March 2017-18 Ag Production Frame; Database: [redacted]).

Frame quality improvements: Additional improvements specific to 2017/18 sampling and future quality of the frame were:

- Placing the 75+1 large units nominated by the BSC (see Frame Checks above) into the CEd sector. The benefit of this was that their information could be fed back to the frame in future years.
- There were 23 strata added to the fully sampled set. These strata had high proportions of units using SBTS and/or had zero EVAO on the frame, as well as high sampling fractions. This added about 400 additional units to the sample, but would enable feedback about EVAO information at the end of this cycle. [redacted] (Subject: Re: Notification: Allocation summary 17/18; Database: [redacted])

Variance Outliers: [redacted] (Subject: Tracking design progress; Database: [redacted])
Units that reported values that were considered to be unrepresentative both within a particular year of design data and across all four years of design data compared to those reported by other units in the population were set as outliers. The design data for REACS 17/18 was REACS data for the 13/14, 14/15 and 6/17 cycles in addition to the 15/16 Ag Census data.

The tracking document provides a list of all units set as surprise outliers during the sample design process. There is also documentation of what commodity values caused these units to be set as

outliers. This outliering process was done over several stages, which differed by whether the intention was to simply lower the overall sample size or to get the Expected RSEs of certain estimates below the Desired RSEs. For further details on the different stages of outliering the reader is directed to the sample allocation document provided below.

Sample allocation & final accuracy constraints for REACS 17/18:
[redacted] (Subject: Re: Notification: Allocation summary 17/18; Database: [redacted] S22 [redacted])
This document contains details regarding the sample allocation for REACS 17/18 that ended up being approved for use by EASB. The set of RSE constraints that were used to generate this allocation can also be found within this document.

A spreadsheet containing the stratum- and state-level allocations as well as the Expected RSEs for all potential design commodities at the National, State and NRM levels can be found at the bottom of this document.

Overlap control strategy:
[redacted] (Subject: Summary Document for REACS 17/18; Database: [redacted] S22 [redacted])
Historically, REACS has only looked maximise overlap with the previous cycle and this was in service of having a good quality of movement estimates. The 16/17 cycle of REACS looked to minimise overlap with previous REACS cycles in an attempt to lower overall respondent burden. For the 17/18 REACS we have continued to aim for a balance in terms of having about a two-thirds overlap in sample with the 16/17 REACS.

For the REACS 17/18 there were sets of selections for which overlap was controlled:
I. Maximise overlap with units selected in REACS 16/17;
II. Minimise overlap with units selected in both the REACS 16/17 and REACS 14/15;
III. Minimise overlap with units selected in any of the REACS 14/15, REACS 13/14 or REACS 12/13

Approval of sample allocation by REASB: [redacted] (Subject: Allocation summary 17/18; Database: [redacted] S22 [redacted]) Written approval of the sample design from EASB is contained in this link above.

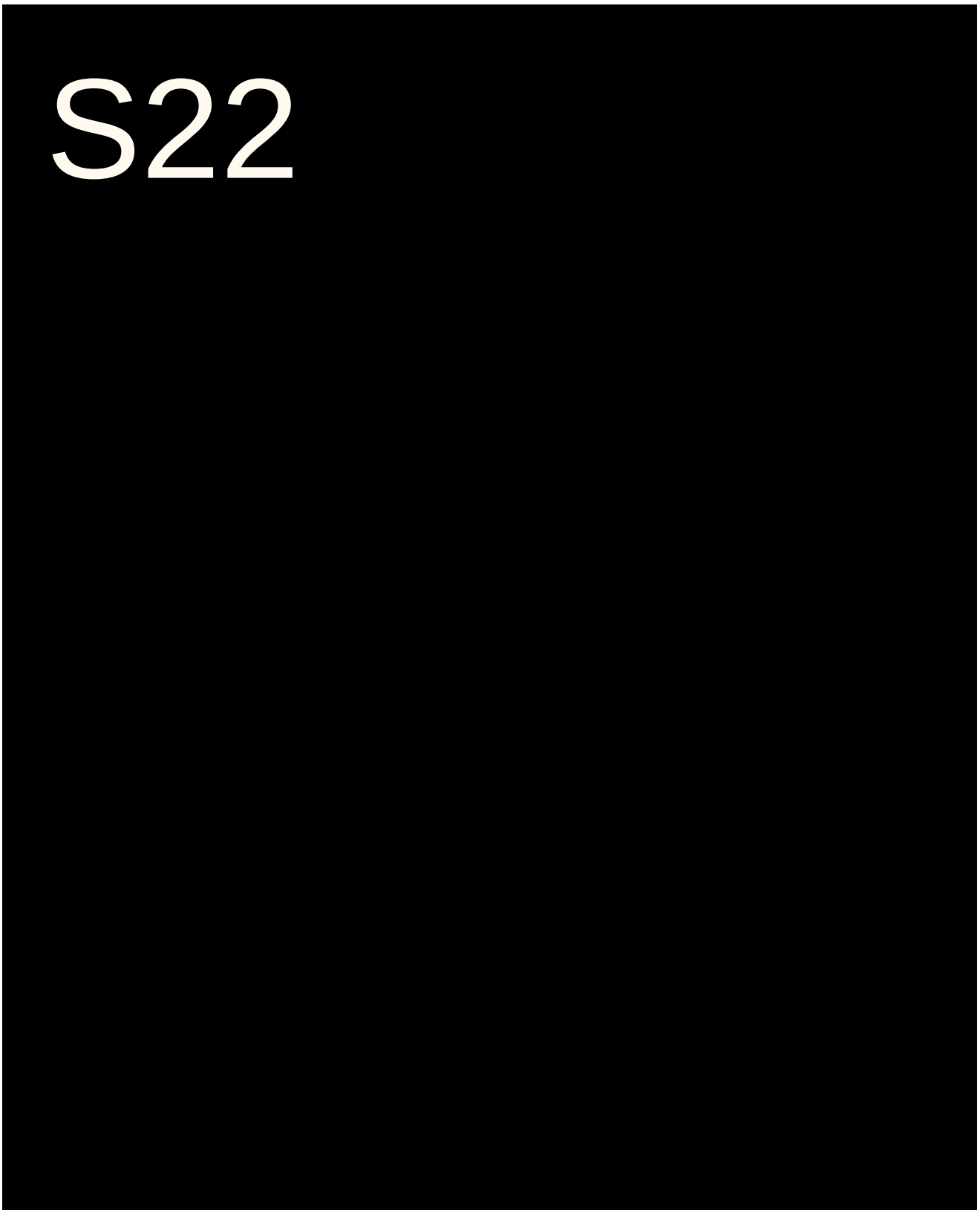
Sample Selections: [redacted] (Subject: REACS 2017/2018 Selections; Database: [redacted] S22 [redacted])
This document contains:
- A copy of the frame that flags units that have been selected for REACS 17/18 and
- Some characteristics of the selected sample (sample counts, overlap and despatch rate);

3. SAS programs and data used for sample design

[redacted] S22

Frame: The frame that was used in the sample design for REACS 17/18 was based on the frame [redacted] S22 within the parent directory [redacted] S22). This frame includes all units considered in-scope with an updated economic size threshold of $40,000 and is not stratified. The stratified frame, which also distinguishes between whether or not a unit is in scope, is called [redacted] S22 and can be found within the output directory.

[redacted] S22


S22

S22

S22

REACS 17/18 Sample Allocation

S22

- The addition of some state-level constraints for selected commodities
- Increasing the minimum sample in each stratum from 10 to 15
- Tighten Area of Holding constraints at the National (to [redacted]%) and State level (to S47E %)
- Top-up selected strata if there is a large reliance on SBTS to permit feedback of EVAO
- Have a final look at strata with large sample size and sampling fraction to check for outlier candidates

Incorporating these updates raised the sample size to 29,315. A summary spreadsheet containing details of the sample allocation is attached:

Summary of Design 17_18 - Final.xlsx

Executive Summary

The REACS 17/18 sample design has returned an allocation of 26,303 total units. This allocation is expected to result in National RSEs which are less than or equal to the RSEs desired by EASB for all 50 commodities designed for at the National level. It is also to meet the desired accuracy for 67% of State-level estimates and 84% NRM-level estimates while only explicitly designing for a small number of these lower level constraints. The sub-national estimates with RSEs that are expected to be larger than the RSEs desired by EASB are all either small contributors to the national total or where the expected accuracy is not all that much larger than the desired accuracy. There were 1293 strata, with 452 of these deliberately CEd and an additional 175 fully sampled through the allocation.

Design Process

The design process for REACS 17/18 has followed a similar process to recent cycles. It involved the following:

- An initial allocation was produced by using:
- The production frame;
- Means and variances from the previous four cycles of REACS (which is this case included REACS 13/14 - 16/17 as well the 15/16 Ag Census);
- The Desired RSEs from the previous design; and
- National-level constraints for commodities but not any sub-national constraints;
  These means and variances represent the expected means and variances that would be achieved by REACS 17/18 in practice, under the assumption that recent historical data would be a good predictor of future responses. Four years of historical data was used to smooth out the impact of any unusual data. Using only national constraints prevents the design from becoming inefficient through being overly constrained;
- This initial allocation had a large sample size (greater than 32,000) so a process of outliering units was undertaken, where unrepresentative units that were having a large impact on sample size were either removed, or where the CE cut-offs for commodities were adjusted. There were a total of 25 units set as outliers, which is a decrease on the last sample which had a total of 36 outliers;

Design Constraints

The final set of Desired RSEs, or maximum allowable RSEs, for various commodities at various output levels are given in Table 1

Table 1. Desired RSEs for REACS 17/18

S47E

Design Performance

A summary of REACS 17/18 sample design, as it currently stands, is presented in the following tables.

Table 2. Summary of the initial REACS 17/18 sample design - National level

S47E

Table 3. Summary of the initial REACS 17/18 sample design - Sub-national levels

S47E

The main results from Tables 2 and 3 are that:

- All Expected RSEs for National-level commodity estimates are lower than the corresponding National-level Desired RSEs;
- Approximately 67% of the State-level estimates and 84% of the NRM-level estimates have expected RSEs lower than the Desired RSEs that would have been. The state-level result is lower than what was achieved in the REACS 16/17 sample design (87%), while the NRM result is similar (85%);

The expected RSEs for all the National-, State- and NRM-level estimates of the required commodities are listed in the following spreadsheet. This spreadsheet also includes a stratum-, state- and sector-level breakdowns of the allocation.

(Note: This allocation summary has been superceded by the final allocation summary attached at the top of this document)

Summary of Design 17_18.xlsx

S22

S22

Summary Document for REACS 16/17

Update July 2017 - It was discovered that there were issues with the stratification (changes in NRMs were not accounted for though a re-optimisation of the industry and size stratification) and so the stratification was less optimal than it could have been. For further details please see the 'Stratification' below.

0. Introduction

The aim of this document is to act as a reference for all major pieces of documentation, inputs and outputs associated with the REACS 16/17 sample design. It also summarises the main differences between the 16/17 sample design and the previous cycle of REACS, which was designed in the 14/15 financial year. This document contains links to:

1. Sample design specifications;
2. Key outputs, both during and at the end of the REACS 16/17 sample design process;
3. The SAS programs used for the sample design, and documentation for those SAS programs;

1. Sample design work and specifications

Project Plan for REACS 16/17: [redacted] (Subject: REACS 16/17 Project Plan; Database: [redacted] S22 [redacted])
This document breaks down the work that was intended to be done as part of REACS 16/17 sample design into two key parts:

1. An investigation into possible changes to the REACS sample design based on the performance of past cycles;
2. The sample design process and selection process for REACS 16/17;

The investigation into possible changes to the sample design based on the performance of past cycles ended up not being incorporated into the 16/17 sample design. However, for the sake of reference, the results of this investigation can be found here: [redacted] (Subject: Past Permormance of REACS; Database: [redacted] S22 [redacted])

A summary of the major tasks assigned with each of these two key parts are described within the project plan.

Size Scope Threshold and Scoping Method: [redacted] (Subject: Confirmation of the use of the new size scoping method for REACS 16/17; Database: [redacted] S22 [redacted])
Since 2016 all Ag collections have moved from classifying units on the Ag frame as being in-scope if they have an economic size of at least $5,000 to the higher threshold of $40,000. Therefore, REACS 16/17 represents the first cycle of REACS with this new size scope. The method for determining which of EVAO and SBTS to use a unit's size value was also changed in order to account for the new size scope.

Maximum sample size: For REACS 16/17 EASB requested a maximum sample size of 30,000 units. This is lower than what was used in earlier REACS cycles and is somewhat based on earlier predictions by BSM regarding the expected sample size under the new scope.

Accuracy constraints: The Desired RSEs for REACS 16/17 are the same as those for REACS 14/15, except for Priority 2 estimates at the State level which were tightened (from [redacted]%) in order to improve the accuracy of some important published estimates. The full suite of Desired RSEs, differentiated by commodity or Priority Group, is provided in the table below.

^ These constraints aren't set as part of the initial allocation process and are usually only set after consultation with the BSC. Also, the constraints are typically only applied to particular commodities rather than all commodities within that Group.

Design commodities: (Subject: REACS 16/17 design constraints spreadsheet; Database: S22)
This document contains a spreadsheet that has the list of possible design commodities at the National, State and NRM levels for the REACS 16/17 sample design. However, in order to not over-constrain the design explicit constraints were only set for a subset of the National and State commodity estimates. At the National level constraints were only set for commodities that had had National constraints in earlier cycles of REACS; this meant that constraints were not set for Apricots, Broccoli, Brussels Sprouts, Cauliflower, Peas, Pumpkin and Sweet Corn.

For further details on the State- and NRM-level constraints that were set the reader is directed to the Sample Allocation document in Section 2.

Stratification: Most details of the REACS stratification are unchanged from the 13/14 cycle, which is described in the following document: (Subject: Sample Design Report for REACS 2013-14; Database: S22).

An important change that should be noted is that because the frame delivered to BSM still uses the $5,000 size value threshold units on the frame that are Out of Scope for REACS 16/17, under the tighter size scope, are given a size stratification value of '0'. In other words, the frame used to take selections for REACS 16/17 doesn't exclude Out of Scope units; it places those units into strata from which no units are selected.

Also, some of the size stratification boundaries were adjusted to take account of the impact of updated frame information following the 15/16 Ag Census on the 16/17 allocation. Details on these changes can be found in the programs 'setce1' and 'Create updated cut-offs' within the sample design SAS project. The location of this project can be found in Section 3.

NOTE: It was discovered that the stratification used for this design was less than optimal, because the industry and size stratification was based on NRM stratifications that had since been changed. This has meant that the sample size is probably larger than it could have been under a more optimal stratification and also, the Expected RSEs provided below are likely to be less accurate predictions of the quality of estimates. For further details see the following document: (Subject: Impact of not accounting for change in NRM definitions for REACS 16/17 design; S22)

2. Outputs from REACS 16/17 sample design

Test frame checks: (Subject: Test Frame Checks; Database: S22)
Production frame checks: (Subject: Production Frame Checks; Database: S22)
These documents outline the results of some checks performed on the test and production frames. One check that was of particular importance for REACS 16/17, given that this was the first cycle after the Ag Census, was to see whether the relevant variables had had their frame values updated to reflect more information about units. The list of variables that were checked can be found here: (Subject: Variables with feedback from Ag Census; Database: S22)

Variance Outliers: [redacted] (Subject: Tracking design progress; Database: [redacted])
Units that reported values that were considered to be unrepresentative both within a particular year of design data and across all four years of design data compared to those reported by other units in the population were set as outliers. The design data for REACS 16/17 was REACS data for the 12/13, 13/14 and 14/15 cycles in addition to the 15/16 Ag Census data.

The tracking document provides a list of all units set as surprise outliers during the sample design process. There is also documentation of what commodity values caused these units to be set as outliers. This outliering process was done over several stages, which differed by whether the intention was to simply lower the overall sample size or to get the Expected RSEs of certain estimates below the Desired RSEs. For further details on the different stages of outliering the reader is directed to the sample allocation document provided below.

Sample allocation & final accuracy constraints for REACS 16/17: [redacted] (Subject: REACS 16/17 Sample Allocation; Database: [redacted])
This document contains details regarding the sample allocation for REACS 16/17 that ended up being approved for use by EASB. The set of RSE constraints that were used to generate this allocation can also be found within this document.

A spreadsheet containing the stratum- and state-level allocations as well as the Expected RSEs for all potential design commodities at the National, State and NRM levels can be found at the bottom of this document.

Overlap control strategy: [redacted] (Subject: Possible overlap control strategies for REACS 16/17; Database: [redacted])
When selecting samples REACS has previously only looked to maximise overlap with the previous cycle's selections. However, in a move to try and get REACS to be more like other business surveys for REACS 16/17 the decision was also made to minimise overlap with units that would be expected to move rotate out REACS since they had been in sample for the last three years. This document presented various options for implementing both overlap control requirements with the options differing based on the relative priorities assigned to the two requirements. The decision was made to slightly prioritise maximising overlap with REACS 14/15 over forcing units that had been in sample for too long out of the sample.

Approval of sample allocation by REASB: [redacted] (Subject: BSC approval for REACS 16/17 sample design; Database: [redacted])
Written approval of the sample allocation and overlap control strategy (see above) by EASB.

Sample Selections: [redacted] (Subject: REACS 2016/17 selections; Database: [redacted])
This document contains:

- A copy of the frame that flags units that have been selected for REACS 16/17; and
- Some characteristics of the selected sample (sample counts, overlap and despatch rate);

3. SAS programs and data used for sample design

[redacted]

Frame: The frame that was used in the sample design for REACS 16/17 was based on the frame [redacted] within the parent directory. This frame includes all units considered

in-scope under the old $5,000 threshold and is not stratified. The stratified frame, which also distinguishes between whether or not a unit is in scope, is called S22 and can be found within the output directory.

S22

S22

Update: The BSC ended up choosing Option 2, which controlling for overlap with both sets of previous REACS selections while preferring to maximise overlap with the previous cycle of REACS selections. This decision still allowed for a reasonable amount of units which had been sampled for each of the last three annual Ag collections (including the Ag Census) to be rotated out of sample for REACS 16/17.

S47E, 47F

Research Officer

Business Statistics Methodology | Methodology and Data Management | **Australian Bureau of Statistics**

(P) S22 **(M)** S22 **(F) (02) 6252 7784**

(E) S47E, 47F **@abs.gov.au (W) www.abs.gov.au**

----- Forwarded by S47E, 47F /Staff/ABS on 11/05/2017 01:45 PM -----

From: S47E, 47F /Staff/ABS
To: S47E, 47F Staff/ABS@ABS, S47E, 47F /Staff/ABS@ABS,
Cc: S47E, 47F /Staff/ABS@ABS, S47E, 47F /Staff/ABS@ABS
Date: 11/05/2017 10:47 AM
Subject: Possible overlap control strategies for REACS 16/17

Hello S47E, 47F and S47E, 47F,

In earlier discussions between EASB and BSM EASB had requested that the selections for REACS 16/17 should look to:

- Maximise overlap with REACS 14/15 (ie. the previous REACS sample); and
- Minimise overlap with units that have been selected in each of Census 15/16, REACS 14/15 and REACS 13/14;

S22 Some possible overlap control strategies that BSM would like EASB to consider are presented below. These strategies differ in the relative priorities BSM have placed on controlling overlap between the two sets of REACS selections;

S22

We can talk about these results and thoughts for moving forward with the selections in the meeting later today. In particular, we'd be interested to hear what you think about:

- How to prioritise between maximising overlap with REACS 14/15 and minimising overlap with units that have been selected in each of Census 15/16, REACS 14/15 and REACS 13/14;
- S22

Previous selections details

Previous selections details

The number of units selected and the number of units on the frame of in-scope units for REACS 16/17 (ie. the units that we can actually control overlap for) are presented in Table 1 below.

Table 1. Number of in-scope units and total units for various sets of selections

Set of selections	Selected units that are in-scope in 16/17	Total units selected at the time of collection	In-scopes as a Total u
S22			
REACS 14/15	25,149	34,620	73%
Units selected for last three years of Ag collections	17,197	17,386	99%

Overlap control options

Table 2 presents some possible overlap control strategies for the REACS 16/17 selections. One thing to note when looking at these options is that units that were selected for the last three years Ag collections (with which we are trying to minimise overlap) is a subset of units that were selected in REACS 14/15 (with which we are trying to minimise overlap). Given that these are competing priorities BSM have presented options that represent a continuum in terms of controlling for overlap between these two sets of selections:

- Options 1 and 5 provide the overlap when controlling the overlap with only one set of REACS selections is considered;
- Options 2 and 4 provide the overlap when controlling the overlap with one set of REACS selections is slightly prioritised over controlling for overlap with the other set of selections; and
- Option 3 provides the overlap when controlling the overlap with each set of selections is given equal priority;

For the sake of brevity references to the REACS 14/15 selections are given by 'R15' and references to common selections in the last three Ag collections are given by 'RP3'.

The overlap control strategies and selections were undertaken by a SAS program developed by BSM a number of years ago, and which has been used for multiple business surveys. This program decides on what units to select based on those units' selection histories and the user's preferences for overlap control. These preferences are specified by the relative priorities, or weights, BSM gives to controlling overlap with those previous selections. These priorities are provided in the results below. .

Table 2. Results from controlling overlap with R15 and RP3 selections at varying priorities

Option	Description	Overlap with R15 selections (% and number of R15 sample sector selections)	Overlap with RP3 selections (% and number of RP3 sample sector selections)
1	Only control overlap with R15	55% (12,052)	58% (8,222)
2	R15 (+1.5), RP3 (-1)	54% (11,644)	52% (7,306)
3	R15 (+1), RP3 (-1)	45% (9,862)	38% (5,449)
4	R15 (+1), RP3 (-1.5)	35% (7,628)	25% (3,497)
5	Only control overlap with RP3	22% (4,806)	19% (2,746)

Some things to note about the results from this table are that:

- Under the scenarios where overlap control is only performed for one set of selections then it possible to achieve an overlap rate for 55% for R15 or an overlap rate of 19% for RP3. However, once overlap control is attempted for both sets of selections at the same time the results become less optimal for both surveys;

- The maximum possible overlap rate for R15 (55%) is lower than the overlap rate of 71% achieved between REACS 13/14 and REACS 14/15, which only looked to maximise overlap between the two surveys. This reduction in the overlap rate is likely due to the fact that it has been two years since the last REACS and so there was more scope for births and deaths to have impact on that overlap figure;
- Once maximising overlap with R15 has a reasonably large priority over minimising overlap with RP3 then the amount of overlap achieved for R15 is not all that different from the case when only overlap with R15 is controlled. However, the amount of overlap with RP3 selections is still sensitive to how strongly that is prioritised over controlling for overlap with R15;
- S22

For the sake of reference I've included some results for the amount of overlap in the CE'd sector, which we cannot control for, below.

	Overlap with			
#(CE'd units)	R15 selections	RP3 selections	S22	#(Unique overlapping selections)
5,487	3,425	3,020		3,856 (70%)

Regards,

S47E, 47F

Research Officer

Business Statistics Methodology | Methodology and Data Management | **Australian Bureau of Statistics**

(P) S22 **(M)** S22 **(F) (02) 6252 7784**

(E) S47E, 47F **@abs.gov.au** **(W)** **www.abs.gov.au**