FINAL REPORT

SCOPING STUDY FOR CANEMAPPS DEVELOPEMENT

FINAL REPORT PREPARED BY	Radanielson A.M, Lai Y.R., Dekeyser S., Mougouei D., Pembleton K.G.
CHIEF INVESTIGATOR(S)	Dr A.M Radanielson
RESEARCH ORGANISATION(S)	University of Southern Queensland
DATE	22 April 2022
RESEARCH MISSION(S)	• Understanding data practices in sugarcane production • Developing well-informed recommendations for a stepwise approach in the development of CaneMAPPS: a tool for Cane production Mapping, Analyses for Productivity, Profitability and Sustainability.

© Copyright 2022 by Sugar Research Australia Limited.

Copyright in this document is owned by Sugar Research Australia Limited (SRA) or by one or more other parties which have provided it to SRA, as indicated in the document. With the exception of any material protected by a trademark, this document is licensed under a **Creative Commons Attribution-NonCommercial 4.0 International** licence (as described through this link). Any use of this publication, other than as authorised under this licence or copyright law, is prohibited.

This link takes you to the relevant licence conditions, including the full legal code.

In referencing this document, please use the citation identified in the document.

Disclaimer:

In this disclaimer a reference to "SRA" means Sugar Research Australia Ltd and its directors, officers, employees, contractors and agents.

This document has been prepared in good faith by the organisation or individual named in the document on the basis of information available to them at the date of publication without any independent verification. Although SRA does its best to present information that is correct and accurate, to the full extent permitted by law SRA makes no warranties, guarantees or representations about the suitability, reliability, currency or accuracy of the information in this document, for any purposes.

The information contained in this document (including tests, inspections and recommendations) is produced for general information only. It is not intended as professional advice on any particular matter. No person should act or fail to act on the basis of any information contained in this document without first conducting independent inquiries and obtaining specific and independent professional advice as appropriate.

To the full extent permitted by law, SRA expressly disclaims all and any liability to any persons in respect of anything done by any such person in reliance (whether in whole or in part) on any information contained in this document, including any loss, damage, cost or expense incurred by any such persons as a result of the use of, or reliance on, any information in this document.

The views expressed in this publication are not necessarily those of SRA.

Any copies made of this document or any part of it must incorporate this disclaimer.

Please cite as Radanielson A.M, Lai Y.R., Dekeyser S., Mougouei D., Pembleton K. (2022) Scoping Study for the development of CaneMAPPS: Final Report Project 2021/03. Sugar Research Australia Limited, Brisbane

ABSTRACT

This project scopes the development requirements of CaneMAPPS: a digital platform to facilitate growers' adoption and implementation of sustainable farming practices in sugarcane production. Consultations with SRA staff and selected industry experts were conducted via videoconference, focusing on the data use, the constraints in data use, and the type of tools available to support productivity improvement in the sugar industry. The status and needs of current SRA activities were used as a source of information for CaneMAPPS requirements. Information was cross-validated with inputs from one-to-one conversations with selected actors in the industry including growers, third-service providers and government body representatives. Data in the industry are available for a range of different purposes and in different formats. They are in separate locations and managed by different stakeholders with varying rights to access, use and sharing. These constraints reflect a lack of consistency in data collection, management and use in the industry. Recommendations for the development of CaneMAPPS are suggested to account for these constraints to enable data stewardship and governance and ensure relevance and successful engagement of different stakeholders in the industry. With the adoption of these recommendations, the core module for CaneMAPPS, which sets up the infrastructure foundation of the platform, and its first decision-support component, a nutrient management and budgeting analysis service, may be developed within a three-year project. The implementation of an agile development pipeline is a prerequisite for CaneMAPPS development to clearly articulate its scope and benefits for the industry. This will require a multidisciplinary team with expertise in human-centred software engineering, end-user engagement and sugarcane production.

EXECUTIVE SUMMARY

The early concepts of CaneMAPPS arise from the need for an efficient approach to generating spatially explicit management recommendations at a paddock level. This scoping study aimed to identify the requirements for the development of CaneMAPPS as a centralised data platform for sugarcane production allowing multilayer mapping and analyses of information to support the development of recommendations for farm management practices to improve profitability, productivity, and sustainability. To achieve this, we conducted a review of the existing data practices in sugarcane production through experts' consultations combined with a desktop study examining relevant advances in software development. This document reports on our findings which lead to the establishment of four key recommendations for CaneMAPPS development.

Consultation workshops with the different stakeholders in the industry including productivity services, governmental representatives and among SRA research teams including the near-infrared (NIR) testing, pest and disease, agronomy teams, district managers and IT teams revealed that the data collected in the industry vary greatly in their sources, formats and processing required. SRA uses farm/block productivity data primarily provided by sugar mills to inform recommendations for field and/or regional specific management. The timeliness of such data provisioning and the formats in which they are provided are often not optimal. These limit their ready integration with the different analyses and services that SRA provide since the mills collected these data for different purposes. Moreover, the processes for data collation, processing and analysis across SRA activities including nutrient budgeting and management (which contributed to the early schematics of CaneMAPPS), NIR testing, pest and diseases control are largely done manually and repeatedly, presenting significant inefficiencies in time and resources use. Historic and current data use for productivity improvement within SRA and the broader industry are project-based, with data storage, processing and codification fragmented through different applications. This problem is exacerbated by the ad-hoc use of digital platforms that are not industry-specific hence not entirely fit for purpose and the low adoption rate of digital solutions by growers. For example, farm data are often collected by different actors for different purposes and formatted for their preferred platforms creating confusion in data ownership and limited trust of growers in the use of data for broader applications beyond their original intent of collection. Legacy data are thus rarely, if not at all, used further to support future development activities and to inform other applications, which presents a loss of opportunity for data utilisation to better inform industry practices. There is thus a high degree of inconsistency in data accessibility, formats and management in the sugar industry indicating the need for an industry-specific data quality and assurance control and management standard.

After careful consideration of the data constraints, data use, storage, management, and analysis need for farm management and research in the sugar industry, four key recommendations for the development of CaneMAPPS were identified and summarised below:

Recommendation 1: A phased development approach

Development of CaneMAPPS is proposed in two stages using a phased approach: stage 1 consists of the core development which sets up the foundation of the software infrastructure of the digital platform and the first iteration of an analysis and service module development, i.e., the nutrient budgeting and management module; stage 2 encompasses the maintenance and continued expansion of the platform to include new modules and applications using the agile development to adapt to the evolving needs of the industry in a changing social, environmental and economic context. The phased approach will ensure the flexible prioritisation of the most useful features and enable sensible project scoping and budgeting. We propose that stage one be developed in a three-year project with a significant investment in staff time for human-centred software engineering, production expertise and end-user engagement. The establishment of CaneMAPPS core infrastructure based on the current data environment in the industry with adjustability via a phased approach aims to create an enabling environment to improve data stewardship in the industry. The data quality control, assurance and management implemented through its core infrastructure in CaneMAPPS supported by its expansion and evolution to cover different applications required by the sugar industry will act as a proof of concept on best data practices for the industry and hopefully disseminate such practices across the different stakeholders including the growers, researchers and service providers through its continued use by the industry.

Recommendation 2: A modular architecture design

We recommend a modular microservice-based architecture for CaneMAPPS development. This ensures scalability, flexibility, and expandability. The modular architecture facilitates continuous and independent development and expansion of different data processing and analysis applications that cater to the evolving priorities of the sugar industry and are agnostic to the programming languages used. The microservice architecture allows direct implementation of access control across modules and services of the platform and reduces vulnerability to security breaches. It is also cost effective in deployment and ongoing maintenance. The modular microservice architecture avoids costs associated with the maintenance of an evolving monolithic centralised application. It also facilitates charging individual modules' computational resources on a per-use basis (and

possibly ephemeral data storage), rather than requiring an always-on costing model for both computation and data storage. This can considerably reduce the hosting cost for the platform.

Recommendation 3: An agile process for expansion and deployment

The proposed microservice architecture of CaneMAPPS supports a high degree of flexibility and maintainability where different services can be added, upgraded, and customized, based on the preferences of the stakeholders. The use of agile software development allows for a step-by-step development of the services with the direct involvement of the stakeholders. To enable CaneMAPPS as the data-driven decision support platform it is envisioned to be, it needs to continuously expand based on the demand and priorities of the sugar industry while complementing existing platforms/digital solutions used, providing actionable solutions ranging from simple data processing and formatting applications to complex soil constraints diagnosis applications. The iterative nature of an agile methodology ensures that from the very first iteration, a working version of CaneMAPPS can be deployed, tested, and promoted/used for real-world productivity improvement. The deployed services can be improved based on user feedback in the subsequent iterations. At the end of each iteration (release) the list of the services and their qualities can be reviewed/updated. This will ensure speedy delivery of the services and give a great level of flexibility to the stakeholders to enhance and customize CaneMAPPS over time based on their priorities, and the availability of the budget/resources.

Recommendation 4: A multidisciplinary human-centric approach

CaneMAPPS is envisioned to be a digital platform that will augment and build upon existing industry initiatives such as 6ES, SPIDNet, QCANESelect, harvesting predictive tool, NutriCalc while leveraging recent advances in sugarcane science and data-driven technologies to deliver actionable solutions to growers. To achieve this, the tool will need to be at the forefront of different disciplines while being well-grounded in the day-to-day needs and operations of the industry and adaptable to the changing priorities through time. A multidisciplinary team that consists of experts in the fields of agriculture, computing (software engineering and human-computer Interaction), and extension/industry engagement is recommended to drive the development of CaneMAPPS so that it meets SRA's and the industry's expectations and vision for the platform.

Two use cases of CaneMAPPS were illustrated using our recommendations for CaneMAPPS' development to demonstrate its comparative advantages as an industry-specific data platform for paddock-specific management and farming systems research and to establish requirements for the core infrastructure of CaneMAPPS and a template for the development of its multi-purpose, multi-layered decision support components.

Adoption of our recommendations will ensure that the development of CaneMAPPS will utilise advances in software engineering, be cost effective and sustainable in the long term and cater to the different stakeholders' needs and be applicable, customisable and expandable to address the current and evolving priorities and challenges the sugar industry faces.

TABLE OF CONTENT

TABLE OF TABLES

TABLE OF FIGURES

1 BACKGROUND

1.1 Sugar Cane Industry towards sustainable transformation

Australia is the second-largest sugarcane producer in the world with about 375 000 ha of land cropped producing approximately 31 million tonnes for the year 2020. Sugarcane is the most dominant crop along the coastal agricultural areas of Queensland and northern New South Wales (SRA, 2021), with a significant majority of production located in the environmentally sensitive Great Barrier Reef (GBR) catchments. The industry is constantly challenged to innovate for improvements in productivity and profitability to remain competitive globally while addressing the increasing constraints domestically to lower its environmental footprint to maintain social licensing to farm. The recently released SRA Strategic Plan 2021-2026 recognised the value of digital innovation and data analytics to deliver insights and support decision making to facilitate transformation in industry practices to improve productivity and profitability while ensuring sustainability.

The development of CaneMAPPS: the Cane Mapping and Analysis for Productivity, Profitability and Sustainability platform, aims to leverage this opportunity to facilitate the collation and access of available data and information and their translation into actionable solutions for growers. CaneMAPPS is envisioned to enable the optimisation of on-farm operations to improve productivity and profitability while ensuring compliance with legislation and environmental sustainability expectations.

1.2 Investment in the Industry for sustainability

Over the last decade, the sugarcane industry has made extensive investments in research and development for improving nutrient management to minimise its environmental footprint. One output from such investment is the six easy steps approach (6ES, 20145/45, 2017/04 SRA investments) which uses paddock-specific information integrated with district-level information to deliver recommendations for nitrogen (N) and phosphorus (P) budgeting to achieve more holistic management of nutrient and soil constraints. However, best practices in data collection, storage and management are required to facilitate the implementation of the proposed site-specific nutrient management to fulfil the sustainability commitment of the industry to make positive contributions to the environment.

Existing digital applications in the industry offer different services for data access and collation with information and recommendations that are unfortunately in a non-readily actionable form to farmers. Most of these digital services require additional processing and effort/skills to apply them to a specific field or subfield scale. The RP222C (Russell-Mulgrave Complete Nutrient Management Planning for Cane Farming) Project has developed an approach attempting to address these challenges leading to the early concept of CaneMAPPS (Appendix 1). This project proposed the development of CaneMAPPS as a tool to support the implementation of best management practices improving productivity at a farm level while ensuring compliance with the relevant legislation and regulations. Knowledge products such as soil maps and productivity maps developed were used to conduct one-on-one consultations for whole-farm operation planning, integrating soil constraints identification, nutrient management for productivity and water quality objectives for growers engaged in the project.

CaneMAPPS requires scalability and flexibility to enable dynamic modelling for different purposes at different spatial and temporal scales to capture the large spatial and temporal variability in sugarcane production and its associated variability in farming management. Such scalability and flexibility are also important to ensure its broad adoption and long-term success. CaneMAPPS is thus envisioned to be a one-stop digital platform for data federation (which is the enablement of multiple databases to function as one), decision support for crop and farm management, performance evaluation and reporting for the sugarcane industry. For these functions to be available in one platform, CaneMAPPS will require a stepwise staged development that the present scoping study aims to establish.

2 PROJECT OBJECTIVES

2.1 CaneMAPPS

This scoping study aims to define the development requirements for the digital platform—CaneMAPPS: Cane Mapping and Analysis for Productivity, Profitability and Sustainability—to enable cane growers to optimise their farming operations to achieve productivity and profitability at a farm level at the same time ensuring compliance with legislation and community expectations regarding environmental sustainability.

CaneMAPPS is envisaged as a central access point for farm data collection, processing and reporting across a wide range of applications from paddock-specific management to whole-farm analysis and regional/whole-of-industry performance and impact reporting. It will facilitate the transition and adoption of industry-best and sustainable practices by growers and facilitate compliance with government regulations and legislation.

While CaneMAPPS presents a single-entry point for a variety of stakeholders, the platform's implementation and distributed deployment is envisaged to leverage the flexibility, scalability, and expandability of a cloud-enabled microservices architecture.

2.2 Objectives

The project objectives were twofold:

1) understanding the data practices in sugarcane production and

2) developing recommendations for the development of CaneMAPPS that will ensure a) its scientific robustness, scalability, and applicability for different user groups across the industry, and b) the use of a stable software architecture that allows minimisation of development costs, easy maintenance and future customisation and expansion.

3 OUTPUTS, OUTCOMES AND IMPLICATIONS

The output of this scoping study is the present report documenting the findings from the expert consultation workshops on the requirements for CaneMAPPS and our recommendations for its development. Four key recommendations are proposed for the development of CaneMAPPS as an interactive and integrative digital platform for whole-of-farm sugarcane management. These recommendations are:

1. The use of a phased approach in developing the CaneMAPPS platform.
2. The adoption of a modular microservices architecture for scalability, flexibility, and expandability, which also enhances cost-efficiency.
3. The implementation of an agile software development process in defining its modules at their conception, development, deployment, and future expansion to ensure stakeholders' engagement, thereby also securing continuing relevance for industry needs.
4. The establishment of a multidisciplinary team to ensure CaneMAPPS' fit-for-purpose design and development and up-to-date relevance in technology and users' expectations to achieve better outcomes for productivity, profitability, and environmental sustainability.

These recommendations were based on inputs from group consultations conducted with key stakeholders from Sugar Research Australia (SRA). Major types of data available and used in the industry to support decision making for farm management were identified and constraints in the data environment of the sugar industry were also discussed.

3.1 Data practices in sugarcane production

In sugarcane production, growers collect and use a wide range of data from varying sources. Limited consistency was reported in how these data were recorded, stored and processed.

Within SRA, different research teams, extensions and commercial service teams have access to different types of data available in a range of formats (Appendix 1). These data were collected at different spatial and temporal resolutions. They are project- and/or service-specific with applications limited by existing agreements among project partners, collaborators, data providers, and customers.

Growers are required to establish their farm maps, conduct soil testing, establish an informed N and P budgeting, and keep a record of its implementation that can be verified by an authorised third party under the GBR catchments regulations (State of Queensland, 2019). These are mostly performed by service providers on behalf of farmers implying that data collected are rarely available for any other use. Different actors are involved in collecting and making the data available for any use leading to complexity in defining data ownership and underscoring the main challenge in data access and use in the industry. Furthermore, district-level information rather than site-specific information is often used in decision making in farm management such as in N and P budgeting leading to inefficiencies in resource management.

3.1.1 *Data collected in sugar cane production*

Four categories of data are regularly made available to characterise sugarcane crops and inform decision making for production:

- Soil data
- Variety data
- Disease risk and incidence
- Productivity data

These data are collected by the growers and/or their service providers for their specific purposes and applications. The data were then hosted and used on different platforms and in different formats. This variety of actors in data collection management presents a challenge in data use in the industry bringing complexity and limited transparency in data ownership and rights over the data access, use, storage and sharing.

For instance, soil data are generally collected by growers through soil sampling and testing performed by service providers such as NutrientAdvantage.

Variety and productivity data are available through mills.

Disease data are available through the productivity service board which conducts regular seasonal surveys and ad hoc sampling based on in-season observations. SRA is providing a commercial service in NIR analyses for the productivity service board and the mills. These data though available for research use for SRA, their storage, and processing are bound by the data privacy agreements between growers and mills and between mills and SRA.

Codification of these data varies not only among the services collecting the data but also from one season to another and among locations/production regions. These disparities were repeatedly identified as a key constraint in the use of the data available in the industry in addition to their late availability for timely planning and decision making for farm management.

Information on weeds and pests constraining productivity is collected irregularly limiting the identification of hotspots and the development of a sustainable approach for risk management.

3.1.2 *Use of available data in sugarcane production*

Time-consuming and monotonous manual processing is generally used by researchers and service providers to harmonise collected data for productivity and profitability analyses. This limits the scalability of existing data analytics and computing tools for paddock specific recommendations at the same time hinders data integration for multiscale and multidisciplinary analyses of productivity indicators for constraints diagnostics and benchmark of potential spatially and over time.

Third-party applications, internal in-house developed scripts and various spreadsheet- and database-based applications are generally used to process harmonised data to generate data products that support farm-management decision making such as nutrient application rates, soil constraint management, disease control and management, and harvesting timing. These data products mainly consist of maps, summary tables and graphs which are used to support one-on-one engagement of individuals or groups of growers to disseminate information and promote the adoption of sustainable practices that are identified from research and project outputs to enhance productivity and sustainability outcomes. This hybrid approach with large manual processing and partial automation is time consuming and does not allow proper implementation of the FAIR principles (Findable, Accessible, Interoperable and Reusable) in the data management for further applications and analyses. Data privacy and security are vulnerable when data are to be manually harmonised and assigned, moving from one platform of data processing to another. Furthermore, data quality and integrity may not be verifiable due to inconsistency in storage, management, processing and codification limiting verification and validation of the outputs, and in extreme cases leading to inadequate recommendations and loss of grower’s trust in the process.

3.2 Requirements for CaneMAPPS

The early schematic for CaneMAPPS was composed of a workflow aimed to produce farm maps to display spatially explicit field information (e.g., cane varieties, soil types) and recommendations (e.g., nutrient budgeting). CaneMAPPS was then proposed to be a digital platform facilitating data access and integration of expert knowledge to inform decision making and support growers’ transition to sustainable practices and regulations compliance.

Major requirements for CaneMAPPS, identified through the different consultations, were for a tool that:

- unifies, harmonises, and allows the interrogation of the current SRA databases (SPIDNet and QCANESelect) and external databases (e.g., productivity services databases and public data such as weather data from SILO and BOM, soil data from the Australian National Soil Archive, and data portals of each state) with the ability to manipulate data within the tool and maintain the integrity of the raw data for reproducibility and interoperability.

- imports and stores data (such as soil test data) to enable data-driven decision-making (including nutrient management and constraint assessment) at a range of scales (e.g., field, farm, region and temporally)
- enables the automation of the current data-processing procedures for timely data-driven discussion and decision making

In time, the full CaneMAPPS scope is envisioned to encompass a wide range of applications in sugarcane management (Figure 1). However, the prerequisite to ensure its relevance for broad adoption in the industry as a central digital portal will start with its ability for:

- Data harmonisation servicing different applications in use in the industry.
- Relatively uncomplicated collection of data not regularly captured in the industry.
- Mapping of multilevel and multilayer data for communication and analyses for constraints identification.
- Information management to support decision making for productivity improvement and positive outcomes for sustainability.

To achieve this, the development process of CaneMAPPS should aim to take the lead in the industry for data literacy and in engaging the industry for data stewardship. This requires long term investment to establish data ownership and management since their collection and the rights of different actors involved in the management, storage, transformation and use of the data for its original intention and further use so that the industry can benefit from the innovative data-driven farm management.

Figure 1 CaneMAPPS as a conceptual central platform for data federation enabling integrated data analyses for productivity, profitability, and sustainability in sugar cane production.

3.3 Recommendations for CaneMAPPS development

To meet the identified requirements and ensure stakeholder engagement, four key recommendations are proposed for CaneMAPPS development. These recommendations are defined for its design, development process and implementation and deployment.

3.3.1 *Recommendation 1: A phased approach to development*

The development of CaneMAPPS is proposed to follow a phased approach including three components:

- A core foundation for the platform, which is composed of the user interface and the back-end structure supporting the architecture of the platform as well as APIs, as detailed in Recommendation 2. This initial stage aims to establish the technical infrastructure of CaneMAPPS that will confer to the platform its main features for data collection, record-keeping, easy reporting and processing for analyses and data integration across platforms and source and visualisation through mapping and different analytics. The design of the CaneMAPPS user interface will be conducted during this initial stage to be friendly, interactive and transparent on CaneMAPPS value proposition for the industry and its users. The foundation of CaneMAPPS for data privacy protection and security will be developed during this initial stage and this will require an inclusive approach to engaging industry stakeholders and establishing expectations and strategies for managing these expectations. This approach may take advantage of the working model that SRA has with different actors in the industry for data sharing and use agreements and customised based on expectations and concerns for data privacy and security concerns in the industry.

- The first iteration of development of a specific module (application), establishing a template for subsequent modules and embedding a reproducible, agile methodology for the implementation, testing, integration, and deployment of further applications. This first iteration is proposed to be targeted for a module in which data requirements are well established and stakeholders and users well defined. As an example, a proof of concept using a nutrient management module for the Tully region can be considered. The data environment in the region is well described and stakeholders in sugar cane production in the region contributed to the first concepts for CaneMAPPS. Steps for this first iteration include the establishment of an inventory of data templates used in the industry and a participatory approach in sorting out templates to be offered in CaneMAPPS and the open field entry for data collection. A recurrent evaluation and test of CaneMAPPS for different features will be conducted to refine the platform for its user experiences and its value proposition in facilitating best nutrient management practices in the industry. This will include the control and treatment case studies and workshop consultation for expectations and feedback.

- Continuous development phase for additional modules (applications). The targets for these subsequent modules are proposed to be identified through a collaborative, priority-based selection process. This phase is supported through Recommendations 3 and 4.

3.3.2 *Recommendation 2: A modular architectural design*

The CaneMAPPS architecture is recommended to be a modular one, built on modern concepts of microservices which ensures scalability, flexibility, and ability to expand to future needs, while also being cost-effective in deployment and ongoing upkeep. The modular microservices architecture avoids costs associated with the maintenance of an evolving monolithic centralised application. It also facilitates per-use charging costs of individual modules' compute resources (and possibly also ephemeral data stores), rather than requiring an always-on costing model for both computation and data storage. The recommended architecture is illustrated in Figure 2.

Significant advantages of the microservices architecture as applied to CaneMAPPS include the following:

- development of new **Processing** microservices can be completed in isolation from the running system. This means new developers do not need to understand all components, and any new deployment can proceed without needing to take existing functionality offline. As an example, an early version of CaneMAPPS may only have the "XYZ" processor but can be readily expanded with processor "WXZ" using the same or other data sources through the use of the common CaneMAPPS API.

- developers can choose programming languages and frameworks that best suit a specific functional module, independent from the other components of the system. For example, some Processors will be best implemented in Python, whereas others may require more specialised tools.

- evolution of existing input templates over time, and support for new ones can be met through the creation of new **Data Importer** microservices, either connecting to existing Data Stores or in concert with the creation of new Data Store services.

- new **Data Stores** can be added dynamically to capture new data with minimal redundancy. Data Stores handle the storage of logically coherent, well-constrained sets of data, each using a database whose structure best supports the particular types of data. In many cases, these will be relational databases (such as PostgreSQL), but some Data Stores may adopt noSQL systems.

- Integration of the various Data Stores, as they are consumed by specific Processors, happens through the **Data Broker** service through a well-defined API. Mapping of new data is facilitated by the Mapping component of the Data Broker, translating between block codes, GPS coordinates, and other identifiers as required. Hence the Data Broker itself may additionally store a minimal set of data that is common to all functionalities.

- Conceptually the Data Stores combined to form a normalised collection of all data, and the Data Broker creates a centralised, consistent view. Avoiding a physical central database ensures that changes to the overall data structure are isolated such that existing functionality does not need redevelopment and can continue to operate.

- from a costing perspective, only those components that see frequent usage will incur continuous compute costs as other microservices can be spun up dynamically.

- scalability is available dynamically and at the level of individual components. This also reduces costs as other components may not need to be scaled.

Platform Considerations

The microservices architecture proposed for CaneMAPPS logically lends itself well to containerised, dynamic resourcing enabled through Cloud providers such as Microsoft Azure and others. However, on-prem deployment is also possible. A careful analysis of possible cost structures of both cloud-based and on-prem implementation of the proposed microservices architecture for CaneMAPPS would be essential but requires well-understood expected usage patterns of the various modules, as well as a detailed comparison of cloud platform vendors.

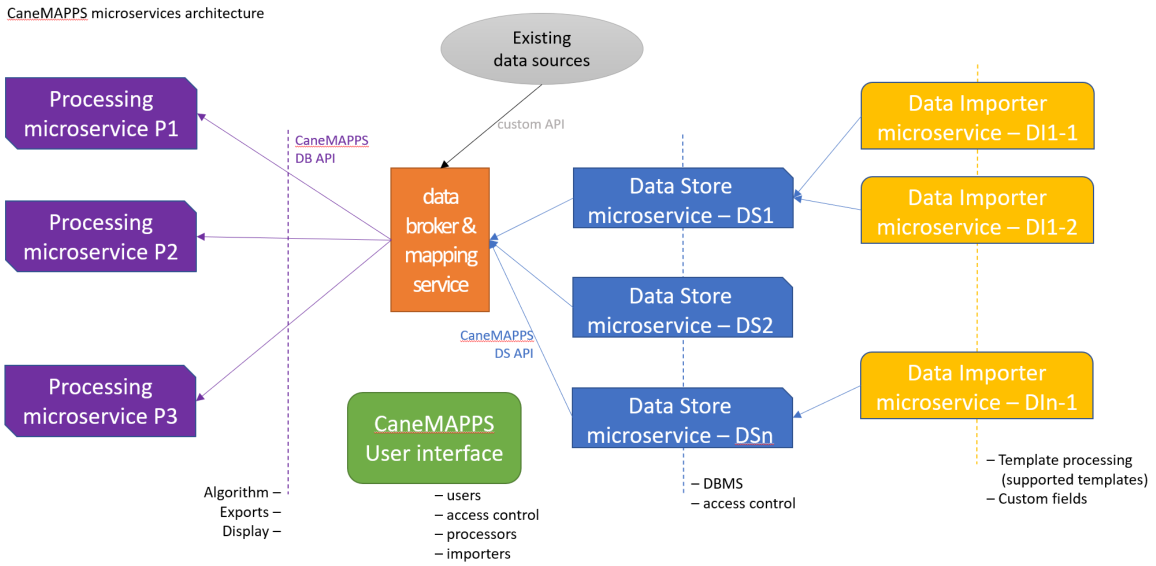

Figure 2 Draft design of the CaneMAPPS platform.

In green the user interface front end components. Different types of microservices are grouped by functionality. So-called "Processing" microservices are associated with different applications of the platform and are unlimited in number as the platform expands. A common data broker and mapping service (orange box) accesses data sources DS1 to DSi through the CaneMAPPS API. It also integrates existing data sources such as SPIDNet and others (grey oval). Finally, "Importer" services allow data entry through supported templates, which can be added dynamically.

3.3.3 *Recommendation 3: An agile process for expansion and deployment*

CaneMAPPS will require access to data and to deliver meaningful data products for different purposes and particularly for constraints for productivity identification and the development of recommendations for profitability and productivity improvement. These will be supported through different services in the proposed microservice architecture of CaneMAPPS. The proposed microservice architecture supports a high degree of flexibility and maintainability where different services can be added, upgraded, and customized, based on the preferences of the stakeholder, while the users continue to benefit from the available services at any given time. To implement such architecture and maintain it, a compatible methodology is required to facilitate implementing and maintaining CaneMAPPS based on the priorities of the stakeholders. In this regard, we propose the use of Agile Methodologies (e.g., Scrum) which have been widely used in similar contexts and allow for a step-by-step development of the services with the direct involvement of the stakeholder to ensure the quality of the services meets the expectations of the users. The iterative nature of an Agile methodology allows for creating a backlog of the services which can be implemented in through different iterations based on the priorities of the stakeholders. It also ensures that from the very first iteration, a working version of CaneMAPPS can be deployed, used, and tested (Figure 3). The deployed services can be improved based on user feedback in the subsequent iterations. At the end of each iteration (release) the list of the services and their qualities can be reviewed/updated. This will ensure speedy delivery of the services and gives a great level of flexibility to the stakeholders to enhance CaneMAPPS over time based on their priorities, and the availability of the budget/resources.

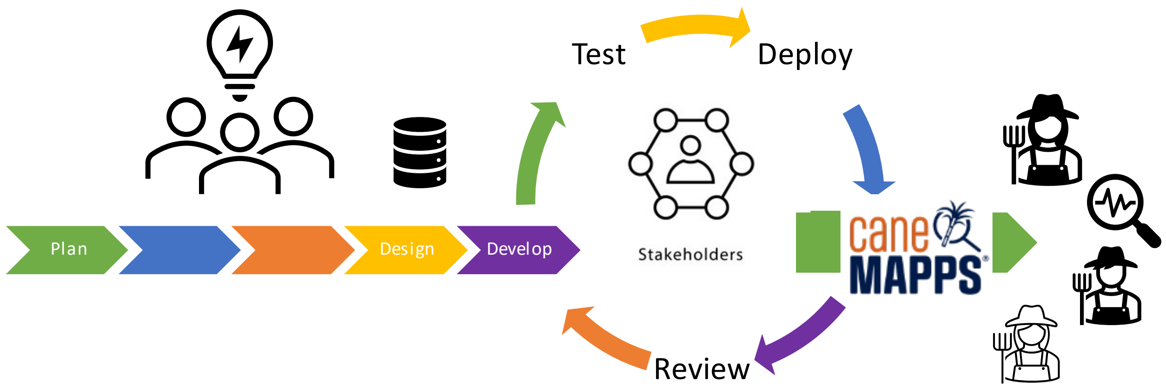

Figure 3 Agile methodology for CaneMAPPS to ensure fit for purpose development and stakeholders' engagement at any phases and stages.

3.3.4 *Recommendation 4: A multidisciplinary human-centric approach*

CaneMAPPS is a digital platform that will augment and build upon existing industry initiatives such as 6ES, SPIDNet/QCANESelect, and Maturity tool, at the same time, will leverage recent advances in sugar cane crops science and data-driven technologies to deliver actionable solutions to growers. To achieve this the tool would need to be at the forefront of different disciplines at the same time grounded in the reality of the day-to-day needs of the industry, adaptable to its continuous change of priorities with time.

The development of CaneMAPPS will require a dedicated multidisciplinary team that will encompass not only its technical complexity but the human aspects of the system, which is essential for its acceptability and usability. It is important to understand the values and preferences of different stakeholders and provide mechanisms through which disagreements can be resolved. The values (e.g., privacy, security) and preferences of the users need to be integrated into the architecture, design, and implementation of the system as early as possible in the development process. This is important to ensure a high level of usability, I.e., the users use the services provided by CaneMAPPS and are willing to contribute to its improvement, for instance through providing feedback/information, entering data and so on. To achieve this, a multidisciplinary team from Agriculture, Human-Computer Interaction, and Social Sciences need to be formed to investigate the values and preferences of the stakeholders and devise the requirements concerning human aspects of CaneMAPPS. These requirements can then be integrated into the services provided by CaneMAPPS during different iterations (stages) of the project - by a team of Software Developers.

With the complexity and the long-term undertaking of CaneMAPPS development, a dedicated professional software development team is required to deliver an optimum software development approach and ensure consistency in the tool development and maintenance as it evolves with time. The software engineers tasked with developing CaneMAPPS will need detailed experience in the implementation of Microservices architectures, cloud and native experience, well established DevOps, Continuous Development/Deployment and Continuous integration and architecture-as-code practices. Experience with data science will also be advantageous and lead to an efficient development pipeline.

As CaneMAPPS will be an SRA branded tool with SRA providing oversights to its application and management, a dedicated SRA resources personnel overseeing the development will be required. This will ensure in house capability and capacity for CaneMAPPS exists within SRA to facilitate long term development and longevity.

3.4 Outcomes and Implications

3.4.1 Documentation of needs of digital platforms for SRA's research and development endeavour

The present project identified immediate gaps that a digital platform such as CaneMAPPS can address and documented existing needs for a digital application at different scales in the different services that SRA is delivering to the industry (see section 6 and Appendix 2). These are valuable information to guide future initiatives for digital solutions development in the industry.

3.4.2 Recommendations supporting an inclusive environment in the industry

Adoption of the recommendations outputs of the project will ensure the availability in the industry of a tool with a robust and end-user focused design. The agile approach proposed in the continuous development of CaneMAPPS aims for an inclusive design informed by relevant stakeholders in the industry, and their different expectations of uses and applications. This will ensure its broad adoption and uptake across the industry, leading to improved productivity and environmental outcomes.

3.4.3 Recommendations for a cost-effective platform maintaining quality and fit for purpose objective

The project recommends a modular architecture for CaneMAPPS that represents a cost-effective software design and deployment.

The proposed phased approach ensures the development of CaneMAPPS as an efficient tool that augments and builds upon existing industry initiatives to deliver actionable solutions to the industry and to position the Australian sugarcane industry as a leader in profitability, environmental sustainability, and resource-use efficiency. The initial phase of CaneMAPPS development will focus on establishing the microservices architecture, data protocols, and addressing data ownership, consent and capture issues. To achieve this in addition to developing the initial service modules over a three-year period will have a cost of approximately $696,000 p.a. This cost includes fixed costs/corporate overheads and potential staffing costs that can be allocated directly to SRA staff, consequently, the cash cost for this development activity could be less depending on the contractual organisation. An itemised breakdown of the costs is provided in table 1 below.

TABLE 1: COSTS PER ANNUM (ASSUMING A THREE-YEAR DEVELOPMENT EFFORT) TO ESTABLISH THE BASELINE INFRASTRUCTURE AND THE INITIAL SERVICE MODULES FOR CANEMAPPS.

PROJECT TEAM (SALARY PLUS ON COSTS)	$ P.A. (ASSUMING A THREE-YEAR DEVELOPMENT EFFORT)
Project leadership	96000
Cane production and data science expertise (to be split across relevant experts)	52000
Human-centred software engineering expertise (to be spilt across relevant experts)	76000
Software development	135000
Extension/industry engagement expertise	27000
SRA oversight	16000
Operational	
Travel for industry engagement	16000
Office costs (ICT, software licences, printing phones)	8000
Hosting costs (e.g., Azure)	25000
Administration, legal, insurance, office space, library	245000
Total	**696000**

Table 2 provides a breakdown of the budget based on four areas of activity: 1) Development of microservices architecture, 2) Establishing CaneMAPPS data protocols, 3) Addressing consent, privacy, security and data ownership challenges, and 4) development testing and refinement of the first service components (nominally the nutrient budgeting components). Indicative duration of each component is also provided. As part of the agile development methodology, the development testing and refinement of the first service components is expected to identify and correct issues with the micro services architecture and the data protocols. This expectation has been accounted for in the proposed budget. After the development of initial service module subsequent service module development are not expected to be as costly. A potential budget for the development of further analytical components with a similar level of complexity to a nutrient budgeting component is provided in table 3.

TABLE 2: ACTIVITY BASED BUDGET FOR THE INITIAL PHASES OF CANEMAPPS DEVELOPEMENT

ACTIVITY	DEVELOPMENT OF MICRO SERVICES ARCHITECTURE	ESTABLISHING CANEMAPPS DATA PROTOCOLS	ADDRESSING CONSENT, PRIVACY, SECURITY AND DATA OWNERSHIP CHALLENGES	DEVELOPMENT TESTING AND REFINEMENT OF THE FIRST SERVICE COMPONENTS
Duration (months)	6	12	12	18
Salaries plus on-costs				
Project leadership	24,000	48,000	72,000	144,000
Cane production and data science expertise (to be split across relevant experts)		60,000	24,000	72,000
Human-centred software engineering expertise (to be spilt across relevant experts)	45,000	36,000	45,000	102,000
Software development	130,000	60,000	15,000	200,000
Extension/industry engagement expertise		17,500	35,000	28,500
SRA oversight	4,000	8,000	12,000	24,000
Operational costs				
Travel for industry engagement and input		12,000	12,000	24,000
Office costs (ICT, software licences, printing phones)	2,000	4,000	6,000	12,000
Hosting costs (e.g., Azure)	37,500			37,500
Administration, legal, insurance, office space, library	152,000	134,000	120,000	329,000
Total	**394,500**	**379,500**	**341,000**	**973,000**

TABLE 3: POTENTIAL BUDGET FOR FURTHER CANEMAPPS SERVICE COMPONENT DEVELOPEMENT AFTER THE COMPLETION OF THE INITIAL DEVELOPMENT PHASE (12 MONTHS DURATION DEVELOPEMENT)

SALARIES PLUS ON-COSTS ITEMS	COST
Project leadership	10,000
Cane production and data science expertise (to be split across relevant experts)	20,000
Human-centred software engineering expertise (to be spilt across relevant experts)	10,000
Software development	96,000
Extension/industry engagement expertise	10,000
SRA oversight	4,000
Operational costs	
Travel for industry engagement and input	12,000
Office costs (ICT, software licences, printing phones)	8,000
Hosting costs (e.g., Azure)	12,000
Administration, legal, insurance, office space, library	153,000
Total	**335,000**

4 INDUSTRY COMMUNICATION AND ENGAGEMENT

4.1 Industry engagement during course of the project

During the ten weeks duration of the project, the project team conducted five group consultations and one to one interviews and informal discussions with SRA staff and some industry participants. These engagement activities represent the first iteration in the agile approach for CaneMAPPS development.

SRA staff were invited to provide their aspirations for a potential digital platform to use supporting their research, development, extension and industry service activities.

With the short duration of the project and its nature, limited engagement and direct communication to the industry as a whole was undertaken and was limited to one-on-one discussions.

An early email informing on the objective of the project was communicated to different stakeholders. These were to introduce the project to the industry and to invite broader larger engagement and input into the development of CaneMAPPS.

4.2 Industry communication messages

With the concept of CaneMAPPS validated, the dot points below were identified as requirements for CaneMAPPS for its first iteration of development:

- CaneMAPPS will be a central digital portal, specific for the cane industry for data access, collection, storage, record keeping and reporting.
- Data harmonisation will be a required service for interoperability with different applications used in the Industry

- Facilitation of easy collection of data not regularly collected in the industry will be required
- Mapping of multilevel and multilayers of data would be made available to facilitate communication and engagement with industry and analyses for constraints identification to support decision making for productivity improvement.

As CaneMAPPS is adopted within the industry, the above requirements will undoubtedly evolve. The agile development approach and microservices architecture that is proposed for CaneMAPPS will facilitate adaption to evolving requirements.   selling points would shift and change for the stakeholders engaged during its development and for its use.

5 METHODOLOGY

To scope the requirements for CaneMAPPS a combination of survey and desktop review was planned in the early stage of the project. This approach was implemented in a two-step process: a targeted group discussion within SRA and an open informal one to one interview with selected industry participants.

This revised approach allowed the project team to define the CaneMAPPS scope for preidentified target users in SRA as they provide service and support to growers.

A desktop review was conducted to overview existing relevant project reports conducted in the industry. These documents were selected as the project's objective were within CaneMAPPS scope namely productivity improvement through farm management in sugar cane production and management practices in sugar cane production (e.g., soil constraints management, soil management, nutrient management) for sustainability outcomes (e.g., water quality, optimum resources use). A review of selected tools used in the industry and other industry were also conducted.

5.1 Survey through groups and one to one interview consultation

Six group consultations were planned with input from the project steering committee. The first of these consultations was used to introduce the project to SRA staff ensuring engagement from the SRA team in the CaneMAPPS development process and invite contributions from teams operating across SRA's remit. The following five consultations were conducted to map the data environment within which SRA groups that are working in supporting growers with their decision making focused on improving productivity, profitability, and sustainability.

The group consultations were conducted as online meetings using teams group meeting platform. They were designed as an open conversation around three main topics:

- The data used and data practices: type, format, source, storage, and management
- The current constraints in data use including adaptative strategies
- The objective of the data use and form of data products used to support and achieve the purpose

The one-to-one interview consultation was initiated from email exchanges between the project team introducing the scoping study objective and inviting the selected key industry participants to provide support and inputs in collecting further information for the scoping study. The key industry participants were selected based on their role and their expertise in sugar cane production and the use of data-driven technologies in crop management. They were representatives of government bodies, growers, productivity service board and third-party service providers.

5.2 Overview of project reports

Several project reports were reviewed to establish an overview of data-driven solutions developed by SRA to support growers in adopting management practices improving productivity and resulting in positive outcomes towards sustainability, particularly in terms of water quality and compliance to environmental regulations.

These reports were screened to identify the types of data available and generated in the industry and how they were made available for use and processing to deliver improved crop management increasing productivity at the paddock level. The different use of these data was identified and their display for research outputs communication and recommendations for extensions services.

5.3 Overview of selected digital tools in the industry

Overview of digital tools used and available in the industry were conducted using information available describing the tool from its current website when available (Appendix 3). The project team focused particularly on tools that were suggested from the open one-to-one discussions including backpaddock (**www.backpaddock.com.au**), production wise (**https://productionwise.com.au/**), AgTRIX (**https://agtrix.com/products/farming-3-0/**) and SST Software (**www.agxplatform.com**). Information available from the comprehensive review on existing decision support tools for soil constraints conducted by Pembleton et al., 2018 was also used with a focus on tools that are used in sugar cane management or were developed specifically for the industry.

5.4 Stakeholder consultation for validation of study outputs and recommendations

Validation of the study outputs was conducted through on-going interaction with the project steering committee and through cross validation of information during the group consultation meetings. Presentation of the project findings through the progress report and the final project presentation were used to validate and finalise the recommendations for CaneMAPPS development. These recommendations were direct outputs of the project which the present report documents.

6 RESULTS AND DISCUSSION

6.1 Data practices in SRA and in sugar cane whole farm operations

Groups consultations and one to one interview were conducted between 31st January to 4 March 2022 to capture inputs from the SRA team's representative of the project developing the early schemer for CaneMAPPS. These teams were the SRA IT team, SRA pest and disease research unit, SRA agronomy team, SRA extensions and industry service team.

Data collected on farm were used to inform planning for next seasons but mainly for evaluation of the farm performance for revenue estimate. Thus, farm data were very sensitive in the industry particularly when it has ability to infer/estimate farm income.

Two main actors were identified determinant in collecting and in making the data available for use in the industry. These are the productivity services board and the mills who directly collected data for growers and/or organise and facilitate the data collection for growers in different format. The most common format seemed to be a paper-based data format.

Mills have most of the required productivity data; however, these data are not available in a timely manner. Data formats also vary between mills. Data labelling is also a key challenge with block numbers used. This format will not be suitable in CaneMAPPS so a harmonizing method will need to be developed.

Farm data for SRA research use and to inform recommendations for field and/or regional specific management are majorly provided by millers who collected the data for different purposes.

For nutrient management, an existing database has been developed as part of the prototype schemer for CaneMAPPS. This database schemer provides a good starting point for CaneMAPPS.

For near-infrared testing (hereafter referred as NIR), the NIR online proprietary system provides service to mills, with spectral results of up to 30 different analyses stored, within which only a small percentage is reported back to growers. The spectral data is currently stored in a binary format with Access databases. A typical workflow requires the processing of 2-5 GB of data with DLL modules used to apply spectral libraries and calibrations.

For weed management and incursions monitoring, SRA currently have historical survey data from 2013-2014 with GPS information and GIS data in Manifold, some aerial mapping and trial data and weed herbicide-resistant data in excel format with GPS locations.

For diseases management, SRA maintains a soil born disease pathology Access database that contains assay data and analysis conducted for the past 25 years together with samples submitted through standardized sample submission sheets with recommended survey methods and sampling strategies.

Soil tests are regularly performed by growers as suggested by the 6ES recommendations (**https://sugarresearch.com.au/growers-and-millers/nutrient-management/**) and the reef regulations (State of Queensland,2019). Soil samplings are generally undertaken at the block level and included analyses for macro and micronutrients content, soil chemical properties such salinity and pH (Appendix 1 samples of soil sampling test). These data are made available to growers in pdf and or in excel format from the service providers (i.e. the soil laboratories). They are often labelled by block and when possible, referenced by geocoordinate of the sampling locations.

6.2 Experiences in digital tool and software development in SRA and for sugar cane crop management

General farm management data are manually encoded in regular computer applications such excel in most of the case or into commercial software for record keeping such provided by AgTrix and SST.

Computer and mobile based applications are rarely used directly by growers. About 20-25% of growers in MacKay region are reported to use computer application facilitated by third party providers using AgTrix and SST platforms.

Encoded data are exported to be processed using excel based applications and customised analytics in SST which provide averaged information and summary display in farm layout for paddock specific outputs.

These data were associated to the mills data which provide indicators of block and farm productivity. They are presented in tables or in maps available in pdf or shape file and made available to growers, accessible to their services providers.

Spatial analyses of these data using map overlays is performed in QGIS or ARCGIS are among SRA service to growers to provide paddock specific information for constraints identification (e. g., block and sub block comparison, varieties performance comparison, hotspot for disease and pest monitoring) and for management decision making (e.g., nutrient management, variety choice).

The above are the workflow for the regular data collection and collation in sugar cane production for productivity improvement processing particularly for nutrient management planning.

Main constraints in data management and processing for the early concept of CaneMAPPS were the limited consistency in templates for data collection and the time-consuming data harmonisation and the requirement of large manual processes data for map production. This will require a data harmonisation approach to be developed as part of the broader CaneMAPPS development.

Inputs collected from the group consultations for the requirements for CaneMAPPS were a tool that:

- unifies and allows the interrogation of the current SRA databases (SPIDNet and QCANESelect) and external databases (e.g., productivity services databases and public data) with the ability to manipulate data within the tool.
- imports and stores data (such as soil test data) to enable data driven decision making tool (including nutrient management and constraint assessment) at different scales
- enables the automation of the current SRA data-processing procedures to facilitate existing services provided by SRA.

The below points are key information for the development of CaneMAPPS and its scope:

- The current preferred and recommended system for ICT development used by SRA is Azure which facilitates collaboration with external parties and provides the option to ensure long-term maintenance. Data collection templates in the industry are diverse. Despite commonalities, they are continuously evolving. The evolution if these templates at time does not seem to serve a purpose and at other times it is important to collect new data sources that were not previously available or utilised. CaneMAPPS will need to help both stabilise the data templates while facilitating changes in template when they are needed.
- Many farmers make decisions based on district level information which is largely accessible and reliable to benchmark potential and risk. However, it is low level resolution information meaning opportunities may to increase productivity at the farm and block level are likely missed.
- CaneMAPPS may provide opportunities to benchmark paddock specific information for validity and diagnostics against potential and actual performance information to facilitate and encourage more farmers to move towards paddock-specific management
- Data aggregation and presentation in an appropriate format, e.g., as thematic maps, at the subdistrict and the district level and temporally (i.e., enabling comparison over time) based on different decision-support purposes/objectives, e.g., productivity benchmarking are important not only for correct data interpretation and for ensuring data privacy.

- CaneMAPPS is envisioned to integrate data for existing and past research projects that SRA is a part of, (e.g., the irrigation project and the remote sensing mapping project) and for industry services SRA currently provides (e.g., NIR, pathology, nutrient budgeting etc.)
- Where existing tools are available, CaneMAPPS is envisioned to complement and enhance the use of such tools to streamline decision-making, such as existing SRA tools including NutriCalc and the harvesting predictive tool for sugarcane. Consequently, CaneMAPPS will need to connect to those tools (most likely via API's) where appropriate.

6.3 User cases examples for CaneMAPPS

The user cases for CaneMAPPS vary due to the large diversity that exist in data users, data type, use, processing, and management in sugarcane production. The two cases below were selected to illustrate how CaneMAPPS can offer opportunities to 1) fill a gap in the industry for monitoring purpose such as for weed control and management and 2) to innovate in the well-established nutrient management recommendation in the industry under 6ES for a data-driven paddock-specific recommendation. These user cases scenarios were defined from the expert's consultation in the industry. For actual implementation, these scenarios can serve as draft plan to initiate discussion and engagement, using the agile approach during CaneMAPPS initial stage and further development, for a broader and inclusive design.

6.3.1 *Weed monitoring application through crowd sourcing of weeds photo*

This is a scenario of application for CaneMAPPS proposed for free to the industry that will allow continuous and ad hoc weed occurrence through photo capture.

An application with log in access for data entry

Through CaneMAPPS application portal, a general user can log in and select weed monitoring icon to check on data ownership rights and licensing approval, then use the application to upload a taken geo-referenced photo or take an actual picture.

These two data entry systems will be available through the user interface of CaneMAPPS and managed through the data import service collecting the file type, the file location and date of capture and the access control provided and approved by the user defined as the user inputting the data in the system.

User interface will specify approval and access rights to the data for different purposes in a given period of time and the conditional storage of the data. These access controls will need to be defined for different end-users such as growers, service providers and researchers with options of change of access and removal of access of data.

Platform service for data storage

Storage and codification of the data is managed through the data store and management of CaneMAPPS which attribute to the photo its location, associated to the season and the year of cropping.

This will require the establishment of a protocol for data harmonisation which will be mainly driven by the data location in space and in time and the data owner. These will be the basis to ensure the data can be findable, understandable, and reusable in the future.

Location of the picture will be encrypted to ensure data privacy. Access to the data will be only available to the data owner and identified users given access and use for the data, for instance for research purposes.

User interface for information on weed species identification and risk to crop productivity

Upon data upload, user can access information to identify the weed species and the level of risk known for yield loss associated to the crop stage and its resistance level to herbicide and initiate consultation with extension agent and service provider in the industry for effective control and management as needed.

This will be provided through a data-processing microservice of CaneMAPPS using computer vision technology such as google lens and other machine learning technologies for image identification against internal image database and cloud-based image databases.

The consistent approach in weed monitoring through CaneMAPPS as proposed will provide a cost-effective methodology to improve the understanding of weed distribution in space and time and its effects on productivity in sugarcane production. Aggregated occurrence in space can map risk and hotspots, this can lead to the evaluation of efficiency of recommendation for weed management and trends of resistance over time. These will support development of well-informed recommendations, improved over time as users continue to upload and capture weed photos and access information on the species identification, risk for herbicide resistance and to productivity for further consultation from extensions and service providers.

6.3.2 *User case scenario for CaneMAPPS to support data driven decision making for Nutrient management*

For CaneMAPPS to be a farm management decision support tool in sugarcane production, it needs to facilitate data collection, record keeping, integration of available data for analyses and reporting as per different regulations requirements. As a digital platform it is also expected to provide capacity for analyses and for information display e.g., as maps and other visualisation means that can be easily incorporated in information processing by growers for their management decisions. For nutrient management, the information processing and analyses needed are to support the recommendations for nitrogen and phosphorus budgeting for the coming cropping season. These are to comply with the reef regulations to ensure licensing for production by implementing the nutrient management 6ES recommendations. 6ES allows establishment of a nutrient management plan for the season, crop in-season monitoring, evaluation for adjusting nutrient management in season, and post season productivity and efficiency evaluation and identification of constraints limiting productivity of the system for management improvement.

Users and user rights in CaneMAPPS to promote and build trust for data stewardship

Potential users of CaneMAPPS in this scenario encompasses different actors involved in collecting, providing the data and make them available for use and actors that are using the data products for specific applications.

For nutrient management decision making, primary data used are associated with the farm characterisation: location and layout, and the paddock characterisation for production: soil test, variety information, crop stage, farm productivity data.

These data are regularly collected by the growers or by service providers for the growers and are available in different format in soft (e.g excel sheet, csv, pdf, shape file, png) and hard copy (i.e paper filling form, pdf print out).

Rights (i.e. for access, use, sharing) over these data are not solely with growers, the primary owner of the data, but also associated with different third parties providing services and equipment that exercise rights on the data by collecting and processing the data in a format ready to use for a given application.

User interface for CaneMAPPS will thus include different layers of access from an open access general public interface with general information on CaneMAPPS value proposition and capability and a next level of access that is a customised user interface defined by user log in.

Each user role is assigned to customised and pre-defined data import, data processing and analyses services. These are specific to customised and pre-defined level of restrictions for data access, use and visualisation.

An example of this using the importing of data into CaneMAPPS for visualisation and input to Nitrogen and Phosphorus budgeting is illustrated in the figure 4 below.

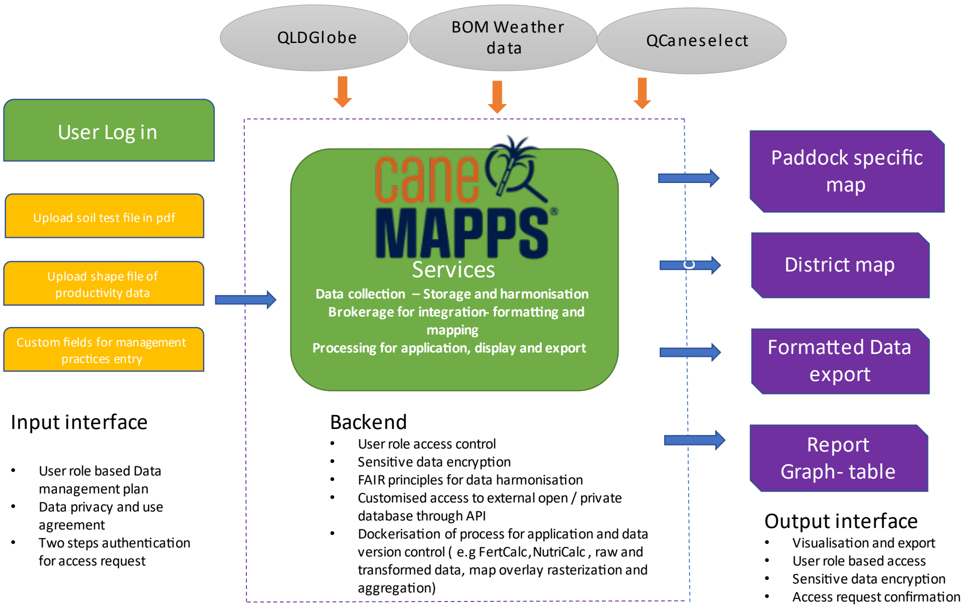

Figure 4: Example of the functions of CaneMAPPS for data import, storage, brokerage, processing for mapping and visualisation through its interface and backend components capability.

Features and processes capability in CaneMAPPS for flexible options

As data collection tool CaneMAPPS is required to provide an interface that allows upload of data in different formats and inform its users on the different options for data protection and access applicable to the data to be used and shared for the user application and CaneMAPPS application.

As a data storage platform facilitating record keeping, CaneMAPPS data storage and management system applies FAIR principles on the data and takes consideration of the rights provided by the users for multilevel data protection and integrity through different encryption methods and security technologies.

Sensitive data such as farm location and farm productivity data will have a high level of encryption and protection in addition of restricted access for different users and for the CaneMAPPS admin. Request for data access can be implemented in several steps based on user role to the recognised data owner and the party with rights on the data. In relational database systems (e.g. PostgreSQL) most of the security requirements can be met through use of row-level security (RLS); for NoSQL datastores equivalent strategies will be employed. This will reduce risk of breach of data access and improve data privacy protection.

Access to the platform will require two-factor authentication (2FA). For instance, growers (and other users) accessing data on CaneMAPPS will require log in with username and password and a verification code through SMS.

SRA researchers conducting interrogation of farm data for a given location in CaneMAPPS will require to submit access requests to the data providers uploading the data to CaneMAPPS (in most of the case third party providers such productivity boards or SRA district managers) and to the primary owner of the data—the grower.

Data harmonisation and spatial capability in CaneMAPPS *for interoperability across platforms*

As CaneMAPPS works mostly with farm data that can be easily linked to spatial features, the geolocation attributes of data will be used to drive data codification and harmonisation. Attributes associated with the data are aggregated and disaggregated spatially.

This enables CaneMAPPS to not be limited with data identifiers used by data providers such as the mills such the inconsistent block identification issues for data sorting. Block labelling became data attributes for a given season, a given year at a given location.

Spatial data management in CaneMAPPS can be processed using open-source data analytics software as R, Python and QGIS, however, when needed it can also be integrated with third-party licensed software.

The microservice-based architecture proposed for CaneMAPPS will allow the use of different options when needed to facilitate the use of third-party licensing software to specific application and the associated legal implications of making it available in an open software.

User, based on their level of access, can then sort out and interrogate data using the location of interest and the common label for farm layout (e.g. block number for a season, year and location).

Apart from primary farm and SRA research project data stored by CaneMAPPS, CaneMAPPS will use secondary spatial data that are readily available from public domain databases, e.g., BOM, ASRIS, Geoscience Australia and TERN databases. These data are accessed only when needed using APIs. User can elect to display such secondary data such as climate and soil classification openly through CaneMAPPS at different spatial scales such as regional, district, farm, block and sub-block levels.

Users, based on their level of access, can access their primary sensitive data (i.e. variables of interest e.g. soil N content or yield) spatially e.g. on the farm layout), and in tabular and graphical forms for the attributes of aggregated data of interest (e.g., average per block, per variety, average or quantiles in the district).

Data processing to inform nutrient management

The microservice-based architecture proposed for CaneMAPPS and its spatial visualisation capability will allow the use of different applications for analyses and data processing on a single platform. For nutrient management, each application e.g., the 6ES, can have its microservice tab available. Potential microservice tabs can include : 1) the estimate of plant N requirement (based on yield potential of the district and the actual prediction of the season potential and Leaf N content through leaf N samples and NDVI remote sensing analyses) , 2) estimate of N rate to be applied (based on soil N and the previous N requirement), and 3) identification of fertiliser type to be applied (based on soil test and available information on different commercial fertilisers and the cropping season).

The data broker service of CaneMAPPS will ensure **formatting of the data into a ready to use format** for these applications and scripts in the data broker service will ensure their aggregation for the spatial representation that are customised and restricted based on the user.

Data microservice processing including application such NutriCalc and FertiCalc can be used then to generate application for N and P recommendation rate and customised script of data analytics tools such R and Python to deliver outputs for display and export and to format data based on user request for instance graph, tabular, Map. User role-based access control on data will remain valid for the outputs access, visualisation and export. For instance, paddock specific information mapping (e.g yield map) will be accessible to data owner and those they grant access for while their data may be used to generate information at district level (i.e average seasonal yield at district level trends over time) that SRA researcher may select to display on the platform for yield gap analyses against the district potential yield.

7 CONCLUSIONS

This scoping study has been conducted to identify the data practices within Australian sugarcane production to inform the scope for a digital tool for data management and analyses supporting decision making for productivity and profitability improvement while considering sustainability factors. In the industry a range of data are collected for a range of objectives. Data collection is completed by different actors with limited cross-sectional applications and standardised data practices. Emerging technologies in software development and in data analytics presented opportunities to unify and harmonise such diverse practices. Doing so will provide a platform to facilitate data driven innovation in sugar cane farming operations and in developing and refining recommendations for best management practices. CaneMAPPS presents an opportunity then to define the data collection, storage, management, record-keeping standards and analysis/processing with not only for sugar research Australia but as well for growers and beyond to preserve institutional and industry knowledge to ensure productivity and profitability improvement and support decision making for adoption of management changes towards positive outcomes for sustainability.

Basic requirements and functionalities for CaneMAPPS development are then defined as:

- a consistent data import, storage, and harmonisation approach to ensure interrogation and manipulation of data from existing databases such as SPIDNet and QCANESelect and enable data quality assurance

- a unified user interface (the IT platform) supported by independent modules (individual interacting microservices containers, Figure 2) that ensure the following services:
  - Data-import microservices (enable the import of different types of farm data in a wide range of formats and templates)
  - Data-management microservices (enable internal data storage and harmonisation), Data-broker microservices (enable data quality consistency assurance through the fetching and interrogation of publicly available data via existing application programming interfaces (APIs) (such as soil data from the Australian National Soil Archive, and from data portals of each state, weather data from SILO and from BOM, etc.) and of SRA internal databases including QCANESelect and SPIDNet while maintaining the data integrity of existing database, and interactions with the data-management microservices)
  - Processing microservices (interact with the data-broker microservices to source relevant data for designated analysis needs (e.g., nutrient budgeting or soil constraints mapping) and output analysis in required formats (e.g., shapefiles/pdfs).

8 RECOMMENDATIONS FOR FURTHER RD&A

Development of CaneMAPPS is proposed as a phased an agile approach. The adoption of an agile approach would allow proper engagement of the tool users, their requirements, and to ensure positive end-user experience. The agile approach would use an iterative approach in defining the applications and analytics to be developed in the tool and the design of its interfaces. Such an approach will ensure adaptability to industry needs.

The first phase of development of CaneMAPPS can be planned in a three-year project and would follow with several phases as the tool evolves and new needs for its applications are identified and required.

Four main recommendations were identified by this project for the development of CaneMAPPS. These recommendations will ensure the tool is fit for purpose, relevant for the industry and developed in a cost-effective manner.

Next phase of the present study is then the initial stage of development with CaneMAPPS (Figure 5) which would start with:

- Validation and refinement of the user case scenario through inclusive stakeholders' consultation in sites with contrasting data environment
- Testing of different stakeholder engagement approaches for the agile methodology development of the core foundation of CaneMAPPS

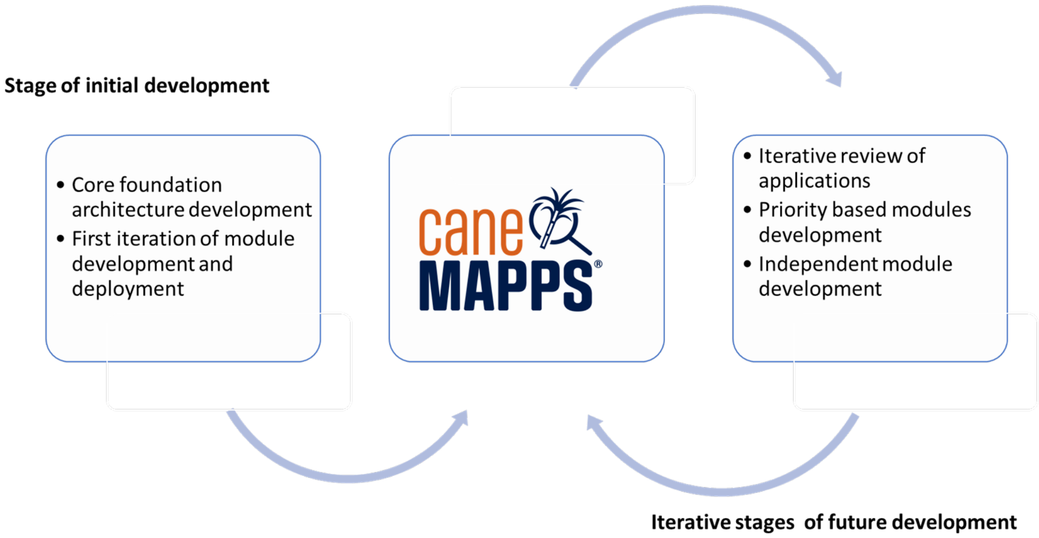

Figure 5: Phased approach for CaneMAPPS development.

9 REFERENCES

Pembleton, K., Ghahramani, A., Kodur, S., Kath, J., Dahlhaus, P., Whatmuff, M., Liu, D., Saravanamuthu, K. 2018. Final Project Report: Soil models, tools and data: Current state of play, future directions and setting up for longevity and a legacy from the CRC for high performance soils [Project 4.3.01], Cooperative Research Centre for High Performance Soils.

SRA. 2021. Sugar Research Australia. Strategic plan 2021-2026.

State of Queensland. 2019. Queensland reef water quality.Farming in the reef catchment, Prescribed methodology for sugarcane cultivation. Office of the Great Barrier Reef, Environmental Policy and Programs, Department of Environment and Science.

10 APPENDIX

10.1 Appendix 1 METADATA DISCLOSURE

TABLE 4 METADATA DISCLOSURE 1

Data	Project reports reviewed
Stored Location	USQ OneDrive
Access	Restricted and by specified users
Contact	Project manager

TABLE 5 METADATA DISCLOSURE 2

Data	Samples of data collection templates
Stored Location	USQ OneDrive
Access	Restricted and by specified users
Contact	Project manager

TABLE 6 METADATA DISCLOSURE 2

Data	Samples of database for nutrient management
Stored Location	USQ OneDrive
Access	Restricted and by specified users
Contact	Project manager

10.2 Appendix 2: Consultation workshops details and inputs

Date	Consultation objectives	Data practices	Constraints	Opportunities and options
28 January 2022	Project activities review			SRA expectations: a tool that (1) unifies and allows the interrogation of the current SRA databases (SPIDNet) and external databases (mill databases and production services databases) with the ability to manipulate data within the tool; (2) import and store farm data (such as soil test data) to provide decision-support (nutrient management and constraint assessment). The target audience is SRA as an organization to use as an extension tool for farmers and automate the current SRA processes. (key variables: The types of data SRA is interested in, the analysis they take.)
31 January 2022	Project introduction to SRA teams			potential scope of CaneMAPPS: an SRA tool to support engagement of growers in adopting research outputs formulated by the translation team, to automate current data processing to allow link among existing SRA tools and other platform that are relevant to SRA activities
2 February 2022	Early concept of CaneMAPPS	Farmers data from several sources	Limited automation in data	Existing database as template

		harmonised into a database manually Processed using excel applications and mapping tool with limited capability for scaling Processed data exported and displayed in overlayed map for different purposes: nutrient budgeting, constraints identification **Resources:** Project report Database for nutrient management Layouts for outputs for nutrient management Russel Mulgrave project Soil test sampling	management and storage Inconsistent data labelling and formatting	Direct application and use of CaneMAPPS with ongoing regular activities in the far north district.
7 February 2022	SRA IT Team	Preferred system for development with Azure DevOps. Large variability of data template in the industry SPIDNet - supply variety disease susceptibility SPIDNet and QCANESelect are going to be re-developed. Growers getting the data from millers and sharing it with SRA or directly accessing the mills' data (Mackay) Access to SRA system with authentication layer for the platform - integrated with Azure ID. Row-level security on data and on access options based on user groups for SPIDNet. Current regulations on data privacy that SRA is currently bound are by Miller agreements.	Mill cannot share growers' data with others due to no agreement in place, however, it is possible for other millers to share, though we need to provide incentives for them). SPIDNet and QCANESelect is the same platform accessed through different interface based on the purpose of the user There are more than 13 templates just for one region. Yield and productivity	Prepopulate farmers crop to one crop earlier to understand the disease conditions and the soil test results, the labs -- e.g., nutrient advantage. Friendly mill group for concept testing and validation e.g., the Tully mill

		Mill data is coming as shapefile but yield and productivity data is available in pdf. Template for soil data is almost standard as soil NAdvantage is the most used soil test provider. **Resources:** Azure database sample for the Nutrient management	data is coming per block and when block is composed of sub-blocks of different varieties yield obtained is a mixed of that. When the block is composed of small different factors such variety, harvesting may not be done per subblock but across block generating average yield for the whole block disregarding the subblock design- this can happen but not systematically	
15 February 2022	SRA Pest and Disease research	Two years of survey data on weed occurrence and status from 2012-2013 25 years of soilborne disease data stored in an Access database Disease monitoring data are available through mills with regular data and ad-hoc sampling analyses collected by productivity services board for growers **Resources:** Poster presentation of weed research outputs and data survey report. Sample of weed data monitoring in excel sheet	Limited tool for data collection for regular monitoring and ad hoc reporting on weeds occurrence and status. Limited data integration for analyses and constraints diagnostic	How to integrate data that are collected for different purposes. capability to provide spatial information at different scales, paddocks, districts Farmer decision making is based on district level information with objective to use paddock specific data
16 February 2022	SRA Agronomy	Remote sensing data for biomass estimate. Productivity data obtained from the mills.	Limited standard practices in data management and storage to leverage	Improving existing tools for crop management such NutriCalc Customised display for different

		Soil test Fertiliser application rate Data collection are project based.	data collected from mills and from different projects.	information presentation associated with access restrictions Benchmarking for farm performance
18 February 2022	SRA extensions and industry services	Data are collected for growers and from the mills Online data systems to support service to Industry with restricted use of data and regulated by agreements between SRA and the mills and the mills and the growers. Few percentages of data were reported back to growers. Spectral data with more than 30 variables collected and stored Farm data import and export using excel applications. **Resources:** Description of current NIR process and options for automation	Proprietary data from different equipment requiring centralised data Limited automation in data processing due to inconsistent data identifier	Platform for augmenting the data collected for the industry to inform research and decision making at different levels Platform for version control of calibration models Foundation data layers for harmonisation and promoting standard data practices for the industry

10.3 Appendix 3 Summary of relevant digital tools informing CaneMAPPS design

Name	Description	Application	Access	Strength	Gaps
AgTrix- Farm 3.0	Agriculture data management platform with customised data processing options	Data storage and visualisation	Paid subscription with different levels for access to services	Experiences across different industry, Customised yield mapping and farm layout for management application features, data visualisation tool	Limited data analyses and automate information integration for management recommendation, required a certain level of skills for ease integration of information for decision making
DecipherAg	Agriculture data management platform with remote sensing service for crop monitoring	Data storage and visualisation	Paid subscription with different levels for access to services	Remote sensing monitoring and customised farm mapping for high resolution NDVI data, visualisation tool	Limited data analyses and require a certain level of skills for ease integration of information for decision making
Climate Corporation Nitrogen Advisor	Decision supporting tool for crop nitrogen management	Monitoring of crop Nitrogen and critical timing for application	Open with registration and log in information. Data collected by the tool is submitted by user with approval and agreement for the tool provider for storage and use to improve the platform data modelling.	Nitrogen dynamic prediction is based on robust modelling and long-term historical data. Simple to use and very focus single application	One application in one platform
AgX Platform	Agriculture data management platform for customised data formatting facilitating interoperability across third party platforms and equipment	Data storage and formatting for further processing and visualisation on different applications in agreement with the AgX Platform	Paid subscription with different level for access to services	Experiences across different industries, Focus on data interoperability and facilitating availability for further use	Required a certain level of skills for use and direct application for farm management. Not industry specific
6th grain	Decision supporting tool for crop management	Grain industry	Paid subscription with different levels for access to services	Integrated different remote sensing data for in season crop growth status monitoring	Specific for grain Industry. Grower only tool for information management than recommendation development.

QCaneselct	Platform for sugar cane variety recommendation and selection	Sugar cane variety information access	Open with registration and log in information	Robust information on variety characteristics with update for the most recent developed varieties for the industry	No link with adaptative and targeted variety specific recommendation for management, Low resolution of variety recommendation with limited integration of paddock specific information
SPIDnet	Platform for sugar research data management with focus on variety evaluation and performance	Sugar cane research data integration	Open with registration and log in information	Sugar cane specific database across research disciplines of SRA	Limited function for data processing and integrated analyses, research-based information with limited application to farm management
Productionwise	Platform for access to varieties of decision supporting tools for grain crop management and agricultural data management	Grain industry	Paid subscription with different level for access to services	Well design platform for data management and information visualisation, Varieties of application in one platform	Limited integration of information processing, required a certain level of skills for ease integration of information for decision making
Ferticalc	Decision supporting tool for crop nutrient management	Establishment of fertiliser application rate for sugar cane crop	Open	Supported and recommended by the industry with specific user recommendation	Recommendation based on soil test with limited account for farm history. Tedious process at scale and limited automation to account for previous farm performance
Susfarms	Decision supporting tool for crop management for Sugar cane crop	Best Management Practices checker for Sugar Cane production in South African environement	Open	Guides to implement best management practices with step-by-step blanket recommendation	Required additional information processing for user specific management
FEAT	Farm economic analysis tool	Evaluation of management practices to guide practices changes	Open with registration and log in	Customised scenarios of cost benefit analyses integrated with industry benchmark. Informative for pre-season planning	Limited application for constraints diagnostic and recommendation for improvement, required next step of information processing for application

Terranis	Agriculture data management platform with remote sensing service for crop monitoring	Remote sensing monitoring for crop growth status monitoring	Paid subscription with different level for access to services	High resolution of imagery for in season monitoring of pest and disease across different crop industry	No link to recommendation and at stage of development for application to Sugarcane in Australia
OZData Australian food data exchange	Agriculture data platform enabling sharing of data across stakeholders	Data management for interoperability and sharing	Open with registration and log in specific to user	Broad application with Cross industry data stewardship and governance implementation opportunity	Broad application with focus on facilitating data access and ready to use for further application for impact