

Horizon 2020 Programme

INFRAIA-02-2017

Integrating Activities for Starting Communities



SmartCow: an integrated infrastructure for increased research capability and innovation in the European cattle sector



Project ID: 730924

Deliverable number: D3.1

Deliverable title: Data Management Plan

EC version : V1

Due date of milestone	31/07/2018 (M6)
Actual submission date	31/07/2018 (M6)

EC version : V2 Updated plan

Due date of milestone	31/07/2019 (M18)
Actual submission date	31/07/2019 (M18)

EC version : V3 Updated plan

Due date of milestone	31/07/2020 (M30)
Actual submission date	15/08/2020 (M30)

1. Author(s)

Organisation name lead contractor	Agrimetrics
--	--------------------

Author	Organisation	e-mail
Anisa Aubin	Agrimetrics	anisa.aubin@agrimetrics.co.uk
Abraham Orchard	Agrimetrics	N/A
Monika Solanki	Agrimetrics	monika.solanki@agrimetrics.co.uk
Richard Tiffin	Agrimetrics	richard.tiffin@agrimetrics.co.uk
Björn Kuhla	FBN Leibniz	b.kuhla@fbn-dummerstorf.de
René Baumont	INRA	rene.baumont@inrae.fr

2. Revision history

Version	Date	Modified by	Comments
1	31/07/2018	Abraham Orchard	-
2	31/07/2019	Anisa Aubin	1 st Update
3	31/07/2020	Anisa Aubin	2 nd Update
4	15/08/2021	Anisa Aubin	Final Update

3. Dissemination level

PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

EXECUTIVE SUMMARY

Background	<p>This report documents the second updated version of the Data Management Plan for the SmartCow research project. This document forms a Deliverable D3.1 for the SmartCow project namely being the publication of an initial version of a Data Management Plan within six months of project inception, and then further updated versions of the Data Management Plan within 18 months (Milestone (MS) 26), 30 months (MS27) and 42 months (MS28) of project inception.</p> <p>SmartCow will integrate key European cattle research infrastructures and provide the academic and private research communities with easy access to 11 major research infrastructures (RIs) from 7 countries. SmartCow combines skills in animal nutrition, genetics, health and welfare and ethics in animal experimentation to facilitate up to 33 research projects involving 59 users.</p> <p>Networking activities harmonise procedures and joint research activities will refine methods to evaluate feed efficiency and emissions, develop new protocols to reduce the use of animals and produce new methods to exploit sensor data for cattle husbandry.</p>
Objectives	<p>The deliverable sets out an approach to gain commitment from project participants to managing the data within the scope of the approach. The deliverable outlines a range of activities, processes and roles that will be required to deliver a coordinated approach to data management, aligned with the principles of the partners in the project. The document will be refined and developed during the project lifecycle and this update provides a summary of progress to date.</p>
Methods	<p>This document and its subsequent updates have been developed drawing on the subject matter expertise of project participants as well as published data approaches. It uses the actual and planned results from WPs available at the time of drafting.</p>
Results & implications	<p>The Data Management Plan outlines a commitment by the SmartCow project to an approach and plan of action for the further development of data management activities within the SmartCow initiative.</p> <p>It sets out a preference for the use of open standards and models including open data ontologies developed as part of the project to simplify access and understanding of the data available. The plan encourages the use of open frameworks and technologies commonly used within academic and commercial data management to reduce proprietary skill barriers to access. The plan also prompts participants towards using licenses including Creative Commons licensing to reduce complexity in data re-use. The document commits the project to engaging with and implementing formal Data Management processes to efficiently gather, store and make research data gathered in the project available in accordance with FAIR principles wherever possible.</p> <p>Smartcow Partners and TNA beneficiaries can upload, control sharing and link their data to get additional insights and store the data beyond the life of the project.</p>

Contents

1. PROJECT ABSTRACT	5
2. DATA SUMMARY	5
2.1 Data generated by Networking Activities (NA) – WP1, 2, 3 and 4.....	6
2.2 Data generated by Joint Research Activities (JRA) – WP5, 6 and 7	6
2.3 Data generated by Transnational Activities (TNA)	6
3. FAIR DATA	8
3.1 Making data findable, including provisions for metadata:	8
3.2 Making data openly accessible:	9
3.3 Making data interoperable:	10
3.4 Increase data reuse (through clarifying licenses):	10
4. ALLOCATION OF RESOURCES	12
5. DATA SECURITY	13
6. POST-PROJECT CONSIDERATIONS	14
6.1 PLATFORM LONGEVITY & SUSTAINABILITY	14
6.2 EXTENSION OF SERVICES	14
7. ETHICAL ASPECTS	14
8. OTHER	15
9. Annex 1.....	15

1. PROJECT ABSTRACT

SmartCow integrates key European cattle research infrastructures (RIs) to promote their coordinated use and development and thereby help the European cattle sector face the challenge of sustainable production. SmartCow will provide the academic and private research communities with easy access to 11 major RIs from 7 countries (18 installations) of high-quality services and resources, covering all the relevant scientific fields and the diversity of cattle types and production systems. These RIs are needed to develop innovative and ethical solutions for efficient use of animal and feed resources that promote animal welfare and healthy livestock, as well as sustainable competitiveness. SmartCow combines strong scientific and technical skills in animal nutrition (in vivo methods for nutrient utilisation and emissions measurements), genetics (genotyped animals, phenotyping capabilities), health and welfare (sensors and automatic recordings of physiological and behavioural traits) and ethics in animal experimentation. A central promotion and management of transnational access to the RIs will provide access to around 10,000 cow-weeks and facilitate up to 33 research projects involving 59 users. Networking activities will harmonise and standardise procedures especially in animal care and measurements, the design of experiments, and data recording and analysis, thanks to a cloud-based data platform tailored for the project. Joint research activities will produce refined methods and proxies to evaluate feed efficiency and emissions, develop new protocols to reduce the use of animals, and produce new methods to exploit sensor data for cattle husbandry. Promotion of transnational access and dissemination of SmartCow outcomes and innovations for maximum impact will be supported by a Stakeholder Platform composed of pre- and post-farm gate industry, farmer organisations, NGOs and policymakers at national and EU levels.

2. DATA SUMMARY

The delivery of a cloud-based data platform provides a central facility for the upload, linking and retrieval of data generated in the SmartCow project. There are three types of activity in SmartCow where data is generated, and these are addressed in this data management plan.

- Networking activities (NA) generate databases describing the resources available in the RIs (animals, equipment, etc.) and the standard methods that are used for routine and experimental measurements.
- Joint Research Activities (JRA) generate data of a diverse nature, including phenotypic, metabolic, physiological and behavioural data that is being uploaded to the dedicated SmartCow cloud-based platform. The aim of this data infrastructure is to facilitate the sharing of information among project partners and ultimately the wider community of scientists and end-users, through tailored interfaces.
- Transnational Access (TNA) research projects share data on a voluntary basis, namely when the owners of the data agree to make them available and with a possible delay after the end of their TNA projects. Historical data (data obtained before the SmartCow project) used in JRA WPs could be introduced in the cloud-based data infrastructure according to rules defined by the owners.

Data is primarily numeric time series observation data stored in relational databases and individual spreadsheets, distributed across RIs. However, the data also includes:

- unstructured/semi structured textual experiment design descriptors,
- animal hereditary and genetic information,
- reference information for physical samples stored within RIs,
- animal condition imagery,
- videos of animal behavior, and
- event level accelerometer data.

Data generated by NA, JRA, and TNA activities that are needed to validate results presented in scientific publications, will be included in the data management plan. The raw data generated by these activities will be stored in the platform, either as reference for publication or future meta-analysis. The data can also be processed by the data owner in cases where the volume of data (e.g. high frequency sensor data) is prohibitive.

2.1 Data generated by Networking Activities (NA) – WP1, 2, 3 and 4

Rich metadata will be included with the data for all data presented through the cloud-based data platform. This will improve the long term value of the data in accordance with the FAIR data principles. This metadata will include provenance information to help with quality assurance. Guidelines and protocols created in WP3.1 have already been linked to Animal Trait Ontology for Livestock (ATOL) concepts during the project. These will be uploaded to the data catalogue on the cloud-based platform. ATOL concepts, in turn, allow bridging data produced during JRAs (see 2.2) and TNAs (see 2.3) with standard operation procedures and guidelines.

WP1 produces a database that maps equipment, animal resources, related techniques, animal database and animal sample banks across the consortium. This is to be updated annually and is stored on the data catalogue. This data catalogue is currently for the use of consortium members, but it can be made available and expanded upon by other researcher infrastructure not in the consortium.

2.2 Data generated by Joint Research Activities (JRA) – WP5, 6 and 7

The capabilities of the SmartCow platform have been tested using data generated by JRAs and were utilised by WP6. Data owners from WP6 created user accounts, uploaded data, linked datasets and appropriately permissioned data using the design functionality developed by Agrimetrics. Data Owners have also been enabled to export the linked data to a useable format of their choice.

The cloud-based data platform will provide an Open Access repository for all WPs to make their JRA data available through, as required by the project. Table 1 below lists these datasets and the required course of action.

2.3 Data generated by Transnational Activities (TNA)

Agrimetrics will provide TNA beneficiaries with instructions on how to use the cloud-based data platform to upload and manage their data. Access to the data can be customized (by each data owner) to reflect the individual collaboration agreements.

Standards and formats for vocabulary (measure and unit) to describe data have already been defined as part of the NA WP3, and are included in Annex 1. These standards are intended to expedite the linking of data and it will be recommended that these are followed for data generated in JRA and TNA WPs, but will not be mandated. Where adequate, data generated in SmartCow will be described according to the Animal Trait Ontology for Livestock (ATOL), Animal Health Ontology for Livestock (AHOL) and Environmental Ontology for Livestock (EOL) (task 3.3 in WP3). Where existing ontologies lack terms for observations, Agrimetrics have – where possible - created some novel ontology terms to bridge these gaps. ATOL and EOL are ongoing initiatives and SmartCow project participants will contribute to improve the ontologies for bovine phenotypic traits. Agrimetrics will integrate these newly created concepts back to ATOL, AHOL and EOL and where relevant other new concepts will be made available through the Agrimetrics marketplace.

It is expected that most data sets collected will be small, due to the nature of the studies. Exceptions are imaging files and data from instrumented behavioral and physiological monitoring. The largest amount of raw data is expected to come from behavioral sensors such as accelerometer, pedometer and temperature monitors.

The data that SmartCow will make openly available will be useful for the cattle husbandry sector, particularly for the scientific community interested in animal nutrition and behaviour, welfare, breeding and livestock production. The data will also be useful to policymakers, funders and industry who have an interest in cattle husbandry through sustainable and safe production of for the public good.

Table 1 Datasets generated through the NA, JRA and TNA that will be hosted on the Agrimetrics platform for SmartCow and associated plan of action.

Work Package	Description	Accessible from	Desired outcome [plan]	Public or Private to Smartcow Consortium
NA Data				
1	Listing of facilities, equipment, and sample banks of the SmartCow consortium	Dataset on Catalogue. Agrimetrics - Data Catalogue: Smartcow WP1 Data capture of facilities and equipment	Reference for SmartCow consortium members post funding.	Consortium only
3	Book of Methods	Dataset on Catalogue. Agrimetrics - Data Catalogue: Smartcow WP3 Book of Methods	Attract people to contribute to and extend existing guidelines.	Public
JRA Data				

5	Data on dry matter intake, milk production/composition, nitrogen balance, and CO2 and CH4 emissions of dairy cattle	Dataset on Catalogue. Agrimetrics - Data Catalogue: SmartCow WP5 Wageningen University and Research	Potential for use by TNA partner and other Consortium members and publication.	Consortium only
6	Sample rumen, milk, methane and faeces data	Agrimetrics - Data Catalogue: Smartcow WP6 Sample Data	Detail pending	Consortium only
7	Data of health scores and biochemical parameters of different cows during the transition period and Data of drinking and resting behaviour and mastitis events	Data Catalogue Agrimetrics - Data Catalogue: TDS evaluation of dairy cows during the transition period	Continued use by the community.	Public
3	Demonstration linked livestock data.	Data Catalogue Agrimetrics - Data Catalogue: Demonstration linked-livestock data	Open access sample dataset to enable anyone to learn how interoperable data can be made available.	Public
TNA Data				
2	Methane Measurement and Feed efficiency data	Pending decision	Further analyses, and potential for scientific publication.	Detail pending
2	Dairy skin sensor data with animal feed and, production and health data collected from 3 sites.	Pending decision	Scientific publication.	Consortium only

3. FAIR DATA

3.1 Making data findable, including provisions for metadata:

Datasets will initially be made findable through their dataset descriptions in the Agrimetrics Data Catalogue. The Agrimetrics Data Catalogue enables dataset owners to control who has visibility of the Data Catalogue entry. If the owner enables public visibility of the Data Catalogue entry then it will be findable through web search engines as well as discoverable within the Data Catalogue itself. If the owner does not enable public visibility then it will only be findable withing the Agrimetrics Data Catalogue by those who the data catalogue owner has permitted access to find it. The Data Catalogue adopts the [Data Catalogue Vocabulary](#), DCAT, standard. This facilitates data discovery by providing a



an integrated infrastructure for increased research capability and innovation in the European cattle sector

rich set of metadata types to which data catalogue owners can provide annotations that those searching for data can search and filter by.

The ontologies defined in the SmartCow project will be indexed in the AgroPortal and Linked Open Vocabularies (LOV) ontology repositories to enable their discovery and wider reuse outside the project. A suitable project-wide versioning scheme will be designed.

A Vocabulary of interlinked dataset (VOID) or Data Catalogue (DCAT) description of the datasets will be provided by each partner and made available to describe the latest version and provide links to the earlier VOID/DCAT versions of the dataset.

The scope for metadata creation will be the various datasets curated in the project. Wherever possible, suitable existing metadata models will be used. Where models are not available, metadata will be created using Semantic Web standards such as RDF, RDFS and OWL

3.2 Making data openly accessible:

During the course of the project, regular discussions have been organised with data owners and curators to assess the licensing terms for the datasets. This will facilitate the decision on which of the datasets can be made openly available and outline the rationale for closed datasets.

The collection of data is the responsibility of each partner involved in the JRA and TNA WPs. A specific task in NA3 (WP3), led by Agrimetrics, has created the cloud-based data infrastructure based on the needs of the SmartCow partners, who are all potential users. Agrimetrics has taken steps to ensure the data infrastructure makes data management and integration straightforward.

Various open source software is used in JRA and NA WPs: Symphony (for ATOL and EOL), SQL Server (to extract subsets of data from partners' databases), Kubios HRV (a heart rate variability analysis software), the R package (for statistical analysis and modelling) and others.

Agrimetrics uses well-understood data management access technologies that are appropriate to the different types of users who will make use of the data collected within the project. These technologies include REST APIs (application programming interfaces) and a Graph QL API, documented using the Swagger specification to facilitate easy data access by data engineers and data analysts using commonly available data tools.

Cloud-based services are used to store the data. Metadata and associated analytical code has been made available through online code repositories such as GitHub. Raw data will be preserved by project partners and in the SmartCow cloud-based data infrastructure. Final results will be stored in the data infrastructure or in publicly available standard data repositories (e.g. Zenodo, HAL, etc) for at least ten years following the end of the project. Upon request, the SmartCow cloud-based data infrastructure will be made accessible to the public for reevaluation and further integration on their own systems (e.g. integration with other databases). Agrimetrics will have the possibility to charge customers beyond the project duration to ensure the maintenance and sustainability of the database. In the medium to long term, the SmartCow cloud-based data infrastructure will be a research facility that can be used in future projects. In addition, partners are encouraged to publish research data as supporting material with their publications to facilitate preservation of data for future reuse by other projects or research initiatives.

SmartCow: an integrated infrastructure for increased research capability and innovation in the European cattle sector



This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

3.3 Making data interoperable:

The FAIR principles have advocated the importance of using standard vocabularies/ontologies to describe FAIR data. In accordance with these principles, SmartCow datasets have been semantically annotated with ontological data models. By using ontologies as metadata models, we rigorously address the Interoperability and Reusability principles of FAIR, which encourages reusing established vocabularies whenever possible (Interoperability) and using as many of these elements as possible (Reusability). Using ontologies as metadata enables the publication of datasets as contextually connected linked data.

Partners from WP3, 5, 6 and 7 participating at VoCamp have agreed on common terminology, key words, units and formats (see Annex 1). It is recommended that these concepts are applied to the data and other documents to be incorporated in the consortium's database as part of the development of these common schemas.

Ontologies that are World Wide Web Consortium (W3C) recommendations, as well as other well-known and widely reused ontologies, will be used as vocabularies where suitable. Semantic mappings to relevant external ontologies will be provided from the ontologies defined within the project.

3.4 Increase data reuse (through clarifying licenses):

Most datasets and metadata will be published under an attribution and share-alike license such as Creative Commons Attribution (CC BY), Creative Commons ShareAlike (CC BY-SA) and Open Data Commons Open Database Licenses (ODbL).

Most datasets will be made available towards the end of the project, however the SmartCow Executive Committee will organise regular discussions with data owners and curators to assess the suitability of releasing the datasets openly during the interim.

It is the goal of SmartCow to share final research data (i.e. cleaned data on which summary statistics are based). Access is provided to allow data to be submitted to the platform and subsequently accessed using automated procedures. The platform incorporates credential management so that access to data can be managed in accordance with the rules defined in the Consortium Agreement. Access to and permissions associated with the dataset is discussed further in Section 5. Access could be given to historical data according to rules defined by the data owners. Both the database and the dedicated data-exchange server will be hosted and managed by Agrimetrics. Sharing with members not involved in the SmartCow consortium will be subject to a data sharing agreement established by the SmartCow members. The agreement will indicate the conditions of use, criteria for access, and acknowledgements in case of publication. Project partners who wish to withhold patentable or proprietary data can do so, and advice on this will be given by the Executive Committee and Intellectual Property (IP) experts from partner organisations. Project partners are committed to give public access to all data that will not be subject to a patent application one year after the end of the

project at the latest, and for a length of 10 years unless a separate agreement is negotiated between the partner and Smartcow members.

Most datasets should be suitable for being used by third parties, however the SmartCow Executive Committee will organise regular discussions with the data owners and curators to assess any restrictions on data reuse.

Appropriate measures are in place to ensure that datasets are curated while maintaining the six core dimensions of data quality: completeness, uniqueness, timeliness, validity, accuracy and consistency. Data providers will need to define data quality for their dataset within their metadata, which can be stored on the cloud-based data platform. Strategies will also need to be defined by the data owner for routine monitoring of curated datasets so that measurement errors, transcription errors, estimation errors and data processing errors are minimised.

The length of time for which a dataset is reusable will vary amongst the datasets curated in the project. The SmartCow Executive Committee will organise interim discussions with data owners and curators to assess the reuse timelines of their datasets. Knowledge generated in the SmartCow project will be managed with the support of project partners, the Executive Committee (composed of WP leaders) and when needed by the IP experts appointed by the partner organisations. Knowledge management will follow the data management plan, as well as the dissemination and exploitation plan, and will respect the rules established in the signed Consortium Agreement (CA).

The management of Background and Results Intellectual Property Rights (IPR) will be regulated in detail through the CA. The CA regulates the process of obtaining IP protection, exploitation and revenue sharing between partners. The CA was prepared and signed at the project start. SmartCow follows the rules for IP set out by the European Commission, specifically:

- “Background” i.e. partners’ pre-existing know-how, while remaining the sole property of their owners, will be made available to other partners when needed for the project implementation (e.g. historical data). List of partners’ Background needed to implement the project will be annexed to the CA, and specific restrictions regarding its use can be added.
- “Results” i.e. knowledge developed through the project, will be owned by the partners who have directly contributed to its creation. In case of joint ownership, a separate contract will be drawn up and signed by the co-owners to determine their rights and obligations and settle the IP management and exploitation rules.
- Traceability of Background and Results information is sought throughout the project. SmartCow generates a constant flux of Results between the partners, and each partner’s contribution to the Results will be recorded.
- Access rights to Results for in-house research or for teaching activities will be granted on a royalty-free basis. Access Rights to Results and Background brought to the project, if needed for Use of a Beneficiary's own Results including commercialization or for third-party research, will be granted on fair and reasonable conditions. Specific secrecy agreements will be signed among partners involved in tasks with sensitive IP and commercial issues, if required. Confidentiality for external guests will be managed through secrecy agreements.

Prior to the dissemination of the project results, the relevant partners will have to submit for approval an abstract to the Executive Committee and to other partners involved in producing the results. The period for reviewing the abstract will be fixed in the CA according to the type of results and dissemination channels. Partners and the Executive Committee will provide an opinion on whether

it is acceptable for dissemination as it is, or if some modifications are required, or if the dissemination should be delayed to implement protection measures. When needed, IP experts will be appointed by the partners' organisations to provide advice on protection measures.

SmartCow will follow the guidelines on open access for scientific publication and research data in Horizon 2020. SmartCow will actively support the publication of results in open access, peer-reviewed scientific journals, with budget allocated to each partner for 'gold' open access (i.e. where open access publications are provided by the publisher with direct access to the underlying research data associated with the publication).

4. ALLOCATION OF RESOURCES

Initial costs for the development of the data infrastructure are covered within the Programme budget and allocated funding.

Sustainability of any solution will be achieved through cost sharing agreements with ongoing programme activities and programme participants. Revenue streams may be explored through commercial routes.

The data infrastructure has been developed to enable ongoing access to data for programme participants and the wider academic community, with costs set appropriately and in agreement with the programme members. Data management is the responsibility of all participants in the programme. Work package leads have been allocated roles for data audit, ontology and data standardisation.

Platform implementation and the design and implementation of practical data management processes has been allocated to Agrimetrics. Agrimetrics works within the programme framework and agrees with programme coordinators and participants (as data owners, consumers and custodians) on any requirements around access control to data within the platform.

While implementation of this plan will be coordinated by Agrimetrics, the specific decisions around the precise order and details of what will be implemented will be decided by the consortium. This is particularly true with regard to data publication and the extent to which the data are published according to FAIR principles.

Agrimetrics remains responsible for drafting the final update to the Data Management Plan and will act as a subject matter expert to engage, educate and inform programme stakeholders on data management processes.

The programme has identified the value of transnational reuse of research data. However the evaluation of the value of the specific data sets will only become apparent once operational data management processes are implemented within the project. Increasing availability of research conducted within the programme alongside any historical data available is likely to increase the likely long-term value of retaining data. The cloud technologies will also allow the exploitation of data storage with lower cost/availability levels to maximise data preservation and minimise ongoing costs. Working with stakeholders, Agrimetrics will implement appropriate evaluation procedures to assign appropriate retention policies for data, in line with its existing data protection policies and procedures.

5. DATA SECURITY

Ensuring the security of data – whether it is commercial, research or personal in nature - is fundamental to Agrimetrics’s business. Most partners in the project are ISO 9001 certified.

Agrimetrics has implemented an Information Security Management System (ISMS): a framework – overseen by an Information Security team – of systems, controls, policies and procedures that are employed to manage and protect the information assets stored on the company’s infrastructure. Agrimetrics’ ISMS is ISO27001-accredited which means the strict international standards for best practice in information security management have been met.

Data security has been of paramount importance throughout the development of the cloud-based platform for SmartCow. Agrimetrics has utilised its existing data security model to:

- Agree defined processes to identify and minimise sensitive data (including any personally identifiable information (PII) stored and processed within the programme.

Agrimetrics’s policies and procedures for handling data are fully compliant with the General Data Protection Regulation (GDPR) and UK Data Protection Act.

Some, limited, personal data is collected when SmartCow researchers create a login to upload data or access data on the SmartCow platform. Agrimetrics uses the information provided, usually an email address, in order to create a user account and provide access to the platform.

Agrimetrics’s Privacy Policy, accessible on Agrimetrics’s website, describes the technical information collected when users connect to the platform. This is in line with the information collected when any of our customers or users access our services and data via the Platform.

- Define, identify and assign appropriate levels of security to data sets.

As the data owner, the user uploading data to the SmartCow platform is in control of the permissions associated with the data at the point of upload and beyond. Permissions can be granted to view or edit both the data and the description of the data. Once uploaded to the platform via the Data Catalogue, the data is stored securely in the Cloud and, depending on the permissions assigned, can be viewed and downloaded by other users of the Data Catalogue. The data owner may choose to have the dataset visible to all users browsing the Catalogue, or to a subset of users (perhaps those involved with a particular Work Package of SmartCow) or may keep the dataset entirely private. The permissions can be revisited at any time. It is not necessary to give Agrimetrics viewing or editing access to the data in order to store it on the SmartCow platform.

- Identify and implement appropriate service levels for storage and backup services for use within the platform including appropriate data replication and disaster recovery capabilities

Information security continuity is embedded in Agrimetrics' business continuity management system. Agrimetrics ISMS includes policies on information security aspects of business continuity management, operations security, and a business continuity and disaster recovery planning procedure. The use of Microsoft Azure Cloud data infrastructure provides multiple levels of system redundancy plus automated back-up and recovery of data. Hosting in Azure minimises the chance that one incident within the infrastructure will cause a loss of data or the loss of service.

Moreover, Agrimetrics is fully-enabled as a remote working business. All tools are cloud-based so can be operated from any device, anywhere, and virtually all documentation is stored electronically. This means that in the event of the loss of our offices or services to the offices the whole team can work from home, with little or no interruption to business activities.

6. POST-PROJECT CONSIDERATIONS

6.1 PLATFORM LONGEVITY & SUSTAINABILITY

The data uploaded to the platform will be stored on the Cloud indefinitely, unless the data owner wishes to remove it. The data owner may remove the data from the platform at any time, either by logging into the Data Catalogue and deleting the dataset, or by contacting Agrimetrics to do this on the data owner's behalf. Data Owners are advised that, while data is available on the platform any user with the appropriate permissions (assigned by the Data Owner) can download the data or link it with other data. Any locally stored versions of the data, which are outside Agrimetrics' control, will remain available in the event that the data is deleted from the platform.

From the outset of the SmartCow project the intention has been for the SmartCow Platform to remain accessible beyond the project duration. This is to enable further evaluation of the data for research purposes and, potentially, further integration (e.g. with other databases). Long-term maintenance of the data platform has an associated cost, the value of which will depend on what is required of the platform. If further interrogation of the data on the Platform is required for future research purposes, researchers should factor the licence cost in to grant applications. Agrimetrics will confirm this cost at the time of submission of such grant applications.

6.2 EXTENSION OF SERVICES

The Executive Committee will engage with the SmartCow partners to define the services that might be required longer-term. Agrimetrics will then establish the options for enabling the extension of services. Specific costs of services will be determined before the end of the project.

7. ETHICAL ASPECTS

The programme's objectives have identified the efficient reuse of existing research, the standardisation of research protocols and the most efficient use of research facilities as an opportunity to minimise the requirement for animal testing while maximising the impact of testing

undertaken (implementation of the 3R principles of ethical research: Replace, Reduce, Refine). Efficient access to data will therefore have positive ethical impact in animal science.

8. OTHER

Agrimetrics will work with programme participants to ensure that country-level and other requirements around data management including privacy are adhered to.

9. Annex 1

Table 2 Outcome of the agreed terms to be used and their mappings to the Agrimetrics Ontology and/or ATOL from VoCamp (2020).

Finalised term to be used as data header for data recording	Ontology Entity (Entities highlighted in green already exist in the Agrimetrics ontology)	ATOL identifier (WP3 to decide the mapping relationship)
Animal Information		
Cow_number	animalNumber	
Cow_ID_XXXX	animalID	
Breed	breed	
Gender	gender	
Birth date	dateOfBirth	
Mother_ID_XXXX	motherID	
Father_ID_XXXXX	fatherID	
Parity	parity	
LactationNumber	lactationNumber	http://opendata.inra.fr/ATOL/ATOL_0005296
Calving date	calvingDate	
Body weight_XXXX	https://data.agrimetrics.co.uk/ontologies/animal/BodyWeight	
Body condition score	bodyConditionScore	
Lameness score	lamenessScore	
Days of pregnancy	daysOfPregnancy	
Production unit structure	productionUnit	
Farm location	Location (lat, lon as attributes)	
Pen_ID	penID	
Observation date	observationDate	
Observation start date	observationStartDate	
Observation end date	observationEndDate	
Observation datetime	observationDateTime	
Observation start datetime	observationStartDateTime	
Observation end datetime	observationEndDateTime	



an integrated infrastructure for increased research capability and innovation in the European cattle sector

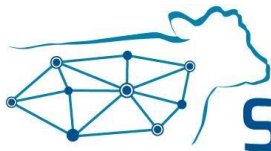
Animal Behaviour		
HeatBehaviour_MinutesPerDay	HeatBehaviour	http://opendata.inra.fr/ATOL/ATOL_0000874
HeatBehaviour_MinutesPerDay	HeatBehaviour	
HeatBehaviour_NrPerDay	HeatBehaviour	
HeatBehaviour_NrPerDay	HeatBehaviour	
HeatBehaviour_StartEndTime	HeatBehaviour	
HeatBehaviour_StartEndTime	HeatBehaviour	
Active_YesNoPerMinute	Active	
Active_YesNoPerMinute	Active	
Active_MinutesPerDay	Active	
Active_MinutesPerDay	Active	
Active_StartEndTime	Active	
Active_StartEndTime	Active	
Inactive_YesNoPerMinute	Inactive	
Inactive_YesNoPerMinute	Inactive	
Inactive_MinutesPerDay	Inactive	
Inactive_MinutesPerDay	Inactive	
Inactive_StartEndTime	Inactive	
Inactive_StartEndTime	Inactive	
Highactive_YesNoPerMinute	Highactive	
Highactive_YesNoPerMinute	Highactive	
Highactive_MinutesPerDay	Highactive	
Highactive_MinutesPerDay	Highactive	
Highactive_StartEndTime	Highactive	
Highactive_StartEndTime	Highactive	
Walking_YesNoPerMinute	Walking	http://opendata.inra.fr/ATOL/ATOL_0000805
Walking_YesNoPerMinute	Walking	
Walking_MinutesPerDay	Walking	
Walking_MinutesPerDay	Walking	
Walking_NrPerMinute	Walking	
Walking_StartEndTime	Walking	
Walking_StartEndTime	Walking	
Eating_YesNoPerMinute	Eating	http://opendata.inra.fr/ATOL/ATOL_0000363
Eating_YesNoPerMinute	Eating	
Eating_MinutesPerDay	Eating	
Eating_MinutesPerDay	Eating	
Eating_MinutesPerDay	Eating	
Eating_NrVisitsPerDay	Eating	
Eating_StartEndTime	Eating	
Eating_StartEndTime	Eating	
Eating_StartEndTime	Eating	
Eating_KgIntake	Eating	
Eating_NrPerMinute	Eating	
Ruminating_YesNoPerMinute	Rumination	http://opendata.inra.fr/ATOL/ATOL_0000779
Ruminating_YesNoPerMinute	Rumination	
Ruminating_YesNoPerMinute	Rumination	
Ruminating_MinutesPerDay	Rumination	
Ruminating_MinutesPerDay	Rumination	
Ruminating_MinutesPerDay	Rumination	
Ruminating_StartEndTime	Rumination	
Ruminating_StartEndTime	Rumination	
Ruminating_StartEndTime	Rumination	
Ruminating_KgIntake	Rumination	
Ruminating_NrPerMinute	Rumination	
Lying_YesNoPerMinute	Lying	http://opendata.inra.fr/ATOL/ATOL_0000837
Lying_YesNoPerMinute	Lying	
Lying_MinutesPerDay	Lying	
Lying_MinutesPerDay	Lying	
Lying_NrBoutsPerDay	Lying	
Lying_NrBoutsPerDay	Lying	
Lying_StartEndTime	Lying	
Lying_StartEndTime	Lying	
Standing_YesNoPerMinute	Standing	http://opendata.inra.fr/ATOL/ATOL_0000835
Standing_YesNoPerMinute	Standing	
Standing_MinutesPerDay	Standing	
Standing_MinutesPerDay	Standing	
Standing_NrBoutsPerDay	Standing	
Standing_NrBoutsPerDay	Standing	
Standing_StartEndTime	Standing	
Standing_StartEndTime	Standing	
Walking time	walkingTime	
Standing time	standingTime	

SmartCow: an integrated infrastructure for increased research capability and innovation in the European cattle sector



This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

Gas Emissions		
enteric CH4 production_g/d	https://data.agrimetrics.co.uk/ontologies/animal/CH4Emission	http://opendata.inra.fr/ATOL/ATOL_0001391
enteric CH4 production_ppm/m	https://data.agrimetrics.co.uk/ontologies/animal/CH4Emission	http://opendata.inra.fr/ATOL/ATOL_0001391
CO2 production_g/d	https://data.agrimetrics.co.uk/ontologies/animal/CO2Emission	http://opendata.inra.fr/ATOL/ATOL_0001392
O2 consumption_g/d	https://data.agrimetrics.co.uk/ontologies/animal/O2Consumption	http://opendata.inra.fr/ATOL/ATOL_0005315
enteric H2 production_g/d	https://data.agrimetrics.co.uk/ontologies/animal/H2Production	http://opendata.inra.fr/ATOL/ATOL_0005314
rumen fluid acetate concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenAcetateConcentration	http://opendata.inra.fr/ATOL/ATOL_0001349
rumen fluid n-butyrate concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenButyrateConcentration	http://opendata.inra.fr/ATOL/ATOL_0001351
rumen fluid propionate concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenPropionateConcentration	http://opendata.inra.fr/ATOL/ATOL_0001350
rumen fluid total VFA concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenTotalVolatileFattyAcid	http://opendata.inra.fr/ATOL/ATOL_0005301
rumen fluid iso-butyrate concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenIsoButyrateConcentration	http://opendata.inra.fr/ATOL/ATOL_0005317
rumen fluid iso-valerate concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenIsoValerateConcentration	http://opendata.inra.fr/ATOL/ATOL_0005318
rumen fluid n-valerate concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenNValerateConcentration	http://opendata.inra.fr/ATOL/ATOL_0001352
rumen fluid caproate concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenCaporateConcentration	http://opendata.inra.fr/ATOL/ATOL_0001353
rumen fluid NH3 concentration_mmol/L	https://data.agrimetrics.co.uk/ontologies/animal/RumenNH3Concentration	http://opendata.inra.fr/ATOL/ATOL_0001314
rumen fluid pH	https://data.agrimetrics.co.uk/ontologies/animal/RumenPH	http://opendata.inra.fr/ATOL/ATOL_0001284
Nutrient utilisation		
fecal NIR spectra_cm-1		
total fecal output_kg/d	TotalFecalOutput	http://opendata.inra.fr/ATOL/ATOL_0005125
total fecal output_kg DM/d	TotalFecalOutput	http://opendata.inra.fr/ATOL/ATOL_0005125
fecal dry matter content_g/100g	FecalDryMatterContent	http://opendata.inra.fr/ATOL/ATOL_0001395
fecal organic matter content_g/kg DM	FecalOrganicMatterContent	http://opendata.inra.fr/ATOL/ATOL_0001409
urine volume_L/d	UrineVolume	http://opendata.inra.fr/ATOL/ATOL_0005132
concentrate intake_kg/d	ConcentrateIntake	
concentrate dry matter intake_kg/d	ConcentrateDryMatterIntake	
name of forage	forage	http://opendata.inra.fr/EOL/EOL_0000351
name of forage batch	forageBatch	
daily forage intake_kg/d	ForageIntake	
forage dry matter content_g/100g	ForageDryMatter	
daily forage dry matter intake_kg/d	ForageDryMatterIntake	
dry matter intake_kg/d	https://data.agrimetrics.co.uk/ontologies/animal/DryMatterIntake	
metabolizable energy_MJ/kg DM	MetabolizableEnergy	http://opendata.inra.fr/EOL/EOL_0001934
diet dry matter_g/kg	DietDryMatter	http://opendata.inra.fr/EOL/EOL_0000067
diet Ash_g/kg DM	DietAsh	http://opendata.inra.fr/EOL/EOL_0000068
Diet Crude protein_g/kg DM	DietCrudeProtein	http://opendata.inra.fr/EOL/EOL_0000124
Diet NDF_g/kg DM	DietNDF	http://opendata.inra.fr/EOL/EOL_0000172
Diet ADF_g/kg DM	DietADF	http://opendata.inra.fr/EOL/EOL_0000115
Diet Starch_g/kg DM	DietStarch	http://opendata.inra.fr/EOL/EOL_0000120
Diet Water Soluble Carbohydrates_g/kg DM	DietWaterSolubleCarbohydrate	http://opendata.inra.fr/EOL/EOL_0000122
Diet OilB_g/kg DM	DietOilB	
Diet Lactic Acid_g/kg DM	DietLacticAcid	
Diet Acetic Acid_g/kg DM	DietAceticAcid	http://opendata.inra.fr/EOL/EOL_0001455
Diet Propionic Acid_g/kg DM	DietPropionicAcid	http://opendata.inra.fr/EOL/EOL_0001456
Diet n-Butyric Acid_g/kg DM	DietNButyricAcid	http://opendata.inra.fr/EOL/EOL_0001457
Diet total volatile fatty acids_g/kg DM	DietTotalVolatileFattyAcids	http://opendata.inra.fr/EOL/EOL_0002000
Diet total volatile compounds_g/kg DM	DietTotalVolatileCompounds	
Diet organic matter digestibility_g/kg	https://data.agrimetrics.co.uk/ontologies/animal/DigestibilityOfOrganicMatter	http://opendata.inra.fr/ATOL/ATOL_0001234
Diet N_g/kg DM	DietNitrogen	http://opendata.inra.fr/EOL/EOL_0000123
diet gross energy_MJ/kg DM	DietGrossEnergy	http://opendata.inra.fr/EOL/EOL_0000164



SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Milk Production		
milk lactose concentration_g/100g	LactoseContentInMilk	http://opendata.inra.fr/ATOL/ATOL_0000619
milk lactose concentration_g/kg		
milk lactose concentration_g/L		
milk protein concentration_g/100g	https://data.agrimetrics.co.uk/ontologies/animal/ProteinContentInMilk	http://opendata.inra.fr/ATOL/ATOL_0001521
milk protein concentration_g/kg		
milk protein concentration_g/L		
milk nitrogen concentration_g/100g	https://data.agrimetrics.co.uk/ontologies/animal/NitrogenInMilk	http://opendata.inra.fr/ATOL/ATOL_0000617
milk nitrogen concentration_g/kg		
milk nitrogen concentration_g/L		
milk fat concentration_g/100g	https://data.agrimetrics.co.uk/ontologies/animal/FatContentInMilk	http://opendata.inra.fr/ATOL/ATOL_0001520
milk fat concentration_g/kg		
milk fat concentration_g/L		
milk yield_kg/day	https://data.agrimetrics.co.uk/ontologies/animal/MilkYield	http://opendata.inra.fr/ATOL/ATOL_0001518
milk yield_kg/milking		
milk yield_L/day		
milk yield_L/milking		
milk yield_kg/suckling		
milk yield_L/suckling		
milk fat yield	FatYieldInMilk	http://opendata.inra.fr/ATOL/ATOL_0000549
milk protein yield	ProteinYieldInMilk	http://opendata.inra.fr/ATOL/ATOL_0000550
milk nitrogen yield	NitrogenYieldInMilk	http://opendata.inra.fr/ATOL/ATOL_0001750
milk lactose yield	LactoseYieldInMilk	http://opendata.inra.fr/ATOL/ATOL_0000618
milk somatic cell count_X1000/mL	MilkSomaticCellCount	http://opendata.inra.fr/ATOL/ATOL_0000991
milk urea concentration_mg/L	UreainMilk	http://opendata.inra.fr/ATOL/ATOL_0000727
milk casein concentration_g/100g	CaseinInMilk	http://opendata.inra.fr/ATOL/ATOL_0000612
milk casein concentration_g/kg		
milk casein concentration_g/L		
milk monounsaturated fatty acid concentration_g/100g	MonoSaturatedFattyAcidConcentrationInMilk	http://opendata.inra.fr/ATOL/ATOL_0000703
milk polyunsaturated fatty acid concentration_g/100g	PolySaturatedFattyAcidConcentrationInMilk	http://opendata.inra.fr/ATOL/ATOL_0000704
milk unsaturated fatty acid concentration_g/100g	UnSaturatedFattyAcidConcentrationInMilk	http://opendata.inra.fr/ATOL/ATOL_0000637
milk saturated fatty acid concentration_g/100g	SaturatedFattyAcidConcentrationInMilk	http://opendata.inra.fr/ATOL/ATOL_0000636
milk ketone body content_mmol/L	KetoneBodyContentInMilk	http://opendata.inra.fr/ATOL/ATOL_0005288
milk beta hydroxybutyric acid content_μmol	BetaHydroxybutyricAcidInMilk	
milk citric acid content_mmol/L	CitricAcidInMilk	http://opendata.inra.fr/ATOL/ATOL_0001765
stage of lactation_month	LactationStage	http://opendata.inra.fr/ATOL/ATOL_0005297
stage of lactation_days		
milking duration_month	MilkingDuration	http://opendata.inra.fr/ATOL/ATOL_0000143
milking duration_days		
milking rate_L/min	RateOfMilking	http://opendata.inra.fr/ATOL/ATOL_0000142
milking rate_L/s		
milking rate_kg/min		
milking rate_kg/s		
mean milk flow rate_L/min	MeanFlowRateOfMilk	http://opendata.inra.fr/ATOL/ATOL_0000138
mean milk flow rate_L/s		
mean milk flow rate_kg/min		
mean milk flow rate_kg/s		
milk free fatty acid concentration_g/100gfat	MilkFreeFromFattyAcidConcentration	http://opendata.inra.fr/ATOL/ATOL_0001753
milk conductivity_mS	ConductivityInMilk	
milk freezing point depression_celsiusDegree	FreezingPointDepressionInMilk	

