International Partnership on Innovation

SAMS - Smart Apiculture Management Services

Deliverable N°1.10

Update Data Management Plan

WP 1 Project Management

Horizon 2020 (H2020-ICT-39-2017)

Project N°780755
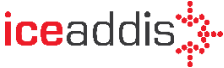
Project information	
Lead partner for the deliverable	Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH (Coordinator)
Document type	Report
Dissemination level	Confidential
Due date and status of the deliverable	31.12.2020	21.12.2020
Author(s)	Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH (Coordinator)
Reviewer(s)	UNIGRA, UNILV, CV.PI, UNPAD, HOLETA, UNIKAS, iceaddis

This document is issued by the consortium formed for the implementation of the SAMS project under Grant Agreement N° 780755.

SAMS consortium partners

Logo	Partner name	Short	Country
giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH	Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH (Coordinator)	GIZ	Germany
UNIKASSEL VERSITÄT	University of Kassel	UNIKAS	Germany
KARL-FRANZENS-UNIVERSITÄT GRAZ UNIVERSITY OF GRAZ UNI GRAZ	University of Graz (Institute for Biology)	UNIGRA	Austria
Latvia University of Life Sciences and Technologies	Latvia University of Life Sciences and Technologies	UNILV	Latvia
iceaddis	ICEADDIS – IT-Consultancy PLC	ICEADDIS	Ethiopia
IQQO Oromia Agricultural Research Institute	Oromia Agricultural Research Institute, Holeta Bee Research Center	HOLETA	Ethiopia
Universitas Padjadjaran	University Padjadjaran	UNPAD	Indonesia

	Commanditaire Vennootschap (CV.) Primary Indonesia	CV.PI	Indonesia

List of Abbreviations

Auth0	Authorization
BSON	Binary JavaScript Object Notation
DisComEx	Dissemination, Communication, Exploitation
DMS	Data Management System
DoA	Description of the Action
DSS	Decision Support System
DW	Data Warehouse
EC	European Commission
e.g.	for example
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
FP7	EU 7th Framework Programme for Research
GA	Grant Agreement
GDPR	General Data Protection Regulation
IDMP	Initial Data Management Plan
ICT	Information and Communication Technology
ITAPIC	Application of Information Technologies in precision Apiculture
JSON	JavaScript Object Notation
MIT license	Open Source license of Massachusetts Institute of Technology
MongoDB	Database Management System
MS	Microsoft
OpenAIRE	European Open Science Infrastructure, for open scholarly and scientific communication
PR	Public Relations
SAMS	Smart Apiculture Management Services
UCD	User Centered Design
UN	United Nations

Web API's   application programming interface for a web server or web browser

Summary of the project

SAMS is a service offer for beekeepers that allows active monitoring and remote sensing of bee colonies by an appropriate and adapted ICT solution. This system supports the beekeeper in ensuring bee health and bee productivity, since bees play a key role in the preservation of our ecosystem, the global fight against hunger and in ensuring our existence. The high potentials to foster sustainable development in different sectors of the partner regions are they are often used inefficient.

Three continents - three scenarios

(1) In Europe, consumption and trading of honey products are increasing whereas the production is stagnating. Beside honey production, pollination services are less developed. Nevertheless, within the EU 35% of human food consumption depend directly or indirectly on pollination activities.

(2) In Ethiopia, beekeepers have a limited access to modern beehive equipment and bee management systems. Due to these constraints, the apicultural sector is far behind his potential.

(3) The apiculture sector in Indonesia is developing slowly and beekeeping is not a priority in the governmental program. These aspects lead to a low beekeeper rate, a low rate of professional processing of bee products, support and marketing and a lack of professional interconnection with bee products processing companies.

Based on the User Centered Design the core activities of SAMS include the development of marketable SAMS Business Services, the adaption of a hive monitoring system for local needs and usability as well as the adaption of a Decision Support System (DSS) based on an open source system. As a key factor of success SAMS uses a multi stakeholder approach on an international and national level to foster the involvement and active participation of beekeepers and all relevant stakeholders along the whole value chain of bees.

The aim of SAMS is to:

- enhance international cooperation of ICT and sustainable agriculture between EU and developing countries in pursuit of the EU commitment to the UN Sustainable Development Goal (SDG N°2) "End hunger, achieve food security and improved nutrition and promote sustainable agriculture"
- increases production of bee products
- creates jobs (particularly youths/ women)
- triggers investments and establishes knowledge exchange through networks.

Project objectives

The overall objective of SAMS is to strengthen international cooperation of the EU with developing countries in ICT, concentrating on the field of sustainable agriculture as a vehicle for rural areas. The SAMS Project aims to develop and refine an open source remote sensing technology and user interaction interface to support small-hold beekeepers in managing and

monitoring the health and productivity in their own bee colonies. Highlighted will be especially the production of bee products and the strengthening of resilience to environmental factors.

- Specific objectives to achieve the aim:
- Addressing requirements of communities and stakeholder
- Adapted monitoring and support technology
- Bee related partnership and cooperation
- International and interregional knowledge and technology transfer
- Training and behavioural response
- Implementation SAMS Business cooperation

Contents

List of tables / figures

Executive summary

This document provides an update of the Initial Data Management Plan (IDMP) for the SAMS project, funded by the EU's Horizon 2020 research and innovation programme under Grant Agreement number 780755. The I**DMP provides an analysis of the main elements of the data management policy that was used throughout the SAMS project** by the project partners, regarding all the datasets that were collected and generated by the project.

The SAMS IDMP as well as this update follows the structure of the Horizon 2020 IDMP template.[1] It describes the data management life cycle for all datasets collected, processed and/ or generated by the research project. The H2020 DMP describes, among others:

- the handling of research data during and after the project;
- the type of data that will be collected, processed, or gathered;
- what methodology and standards will be applied;
- whether and how the data will be made (openly) accessible;
- how the data are stored.

The DMP **ensures that data are transferred and used in a secure setting**; that use of the data is compliant with legal requirements (including signed informed consent and the EU data protection regulation) and that the use of both existing as well as new data occurs in agreement with the Data Owner/ Data Provider. Management of datasets that include personal information are compliant with the General Data Protection Regulation (GDPR, Regulation (EU) 2016/ 679). The GDPR is a regulation by which the European Parliament, the European Council and the European Commission intend to strengthen and unify data protection for individuals within the European Union (EU).

The purpose of this document for the SAMS project is to **present management and storing of all data used and created by SAMS in a comprehensive and detailed manner**. Like this, the project officer as well as other staff at the European Commission/ EASME may consult this document to receive information on data management and storing.

Data Management Plans (DMPs) are a **key element** of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/ or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a **DMP includes information** on:

- What is the purpose of the data collection/ generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/ collect?
- Will you re-use any existing data and how?

---

[1]   Data   Management   http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

- What is the origin of the data?

- What is the expected size of the data?

- To whom might it be useful ('data utility')?

These questions guide this document and provide a comprehensive set of information on the specific data management of the SAMS project.

Data Management Plan

Organizing, implementing and managing a project for over three years requires regular and consequent management of all kinds of data and information used or created within the project duration. Firstly, a reliable baseline was very important in order to be able to monitor and judge the results of the project. Secondly, the data generated by project activities needed to be collected, structured and presented in a way that attracts the right target groups and is conform with data management regulations.

1. Data Summary

The following data are collected and generated within the SAMS framework:

- SAMS bee colony data;
- SAMS user data (not processed or stored);
- Apiculture knowledge base;
- Project documentation: reports, deliverables and PR.

The overall purpose of such data generation and collection within the SAMS project is to have reliable data and comparative figures as baseline to research related to bees and to provide possibilities for income improvement in Ethiopia and Indonesia. Data sets are meant to be used further for scientific research.

All data is stored in a standard format to avoid the use of restricting or outdated data formats and forms of data storage. The chosen formats are based on file formats that have been recognized to be the file formats of choice for longevity and interoperability.

1.1 SAMS bee colony data

1.1.1 Purpose of data collection/ generation

The main purpose of bee colony data collection was:

- to enable the development and testing of a remote sensing technology for monitoring of the state of bee colonies;
- to gain insights into specifications of bee colony data monitoring and research aspects;
- to develop, based on data, a support tool for beekeepers;
- to identify different bee colony states and evaluate the colony status.

Therefore, it was necessary to collect data about individual bee colonies in Ethiopia, Indonesia and Europe. Data was collected to monitor the colony behavior and to develop algorithms for bee colony state recognition which is the basis for the Decision Support System (DSS) development.

All the measured data about the behaviour of bee colonies, gathered from the HIVE monitoring system, can be stored for further analysis and decision support. For the data storage a dedicated Data Warehouse (DW) is developed, which can be considered as a universal system and is able to operate with different data inputs and have flexible data processing algorithms.

Bee colony data collected during previous FP7 project "ITAPIC" was also used. This data was used to make a comparison with newly collected data in other countries that have different climate and environmental conditions. Such a comparison gave an insight about any significant differences that needed to be considered regarding honeybee behavior and the related set-up of the Decision Support System. Results gained by comparison were also a great resource for demonstration purposes and for preparing scientific publications.

1.1.2 Relation to the objectives of the project

The SAMS bee colony data generation and collection are related to the development and set-up of the Decision Support System (DSS) for local needs in Indonesia and Ethiopia. The DSS shall respond to local data management and different bee families in different regional settings in real-time and provide immediately advise solution and actions to beekeepers if necessary, to improve their actions and increase efficiency of their activities.

1.1.3 Types and formats of data generated/ collected

The developed SAMS monitoring system gathers data only about honeybee colonies and only this data is being stored and processed in the SAMS DW. Such data is temperature, humidity, weight and sound. SAMS DW users can store and access their bee-data in the DW workspace and use the DW for the interpretation of measured data and support for decision making processes. Moreover, the SAMS DW user can chose to share their data with third parties or to keep their data private.

The data is gained from various sensors, which are placed in the hive (temperature, acoustic) or beneath the hive (scales), as well environmental conditions (temperature and humidity) are monitored

Table 1 Bee colony data

Parameter	Type of data	Storage format	Place of storage
Bee colony temperature (°C)	Numeric data	Binary JavaScript Object Notation (BSON)	Cloud data warehouse
Bee colony weight (kg)	Numeric data	Binary JavaScript Object Notation (BSON)	Cloud data warehouse
Bee colony acoustics (intensity frequencies)	Numeric data	Binary JavaScript Object Notation (BSON)	Cloud data warehouse

Environmental temperature (°C)	Numeric data	Binary JavaScript Object Notation (BSON)	Cloud data warehouse
Environmental humidity (% RH)	Numeric data	Binary JavaScript Object Notation (BSON)	Cloud data warehouse
Humidity inside the hive (% RH)	Numeric data	Binary JavaScript Object Notation (BSON)	Cloud data warehouse

Why was this data format chosen?

Data format was chosen based on sensor specification and output values. Storage format is automatically chosen by the specific database used; in the SAMS case it was MongoDB that stores data in BSON format.

Do the chosen formats and software enable sharing and long-term validity of data?

Yes, the chosen formats and software enable sharing and long-term validity of data.

1.1.4 Utility of the data

All the data and knowledge gathered during the SAMS project time are provided to interested third parties for research purposes and to enable national and international knowledge exchange on bee data and bee management. Providing open access to SAMS data and results, shall help to better understand bee colony behaviour in Ethiopia and Indonesia, and to achieve minimisation of honeybee losses and increase of honey production and quality in the mentioned regions.

All the collected data and linked hardware to the Data Warehouse will continue to operate and users will not lose their connection to the Data Warehouse. Third parties will be able to use the collected data for the research purposes as well.

Data collected within the SAMS project can be of use to several target groups and communities. The SAMS project team has distinguished four main user groups of SAMS data: beekeepers, scientists, economic communities and general public.

More information is provided in the Report D.6.3. Transfer Study on Data Utilisation.

1.1.5 Reuse of the data

Are there any existing data or methods that can be reused?

All collected data can be reused to apply other analysis algorithms by other scientists.

Is it needed to pay to reuse existing data?

No, data collected during the SAMS project is open and free. Interested parties can request collected data e.g. for reuse for scientific research aspects from the Latvia University of Life

Sciences and Technologies, Faculty of Information Technologies. Different options are possible afterwards (1) sharing single workspace or (2) provide all existing data.

Are there any restrictions on the reuse of third-party data?

Data from any general available bee colony monitoring system and therefore third-party can directly be connected to the SAMS DW to transfer and store the data. There are no restrictions for the reuse of such data. By default, collected bee colony data within the SAMS project is intended to be open and can be shared with third parties after the project end. All involved beekeepers accept this approach and agree to share data about their colonies.

Can the data created (which may be derived from third-party data) be shared?

Yes, data can be shared.

1.1.6 Origin of the data

Data was gained through the installation process of the SAMS Hive monitoring systems in Ethiopia and Indonesia together with local beekeepers, and in Germany as well as in Latvia by test installation through project partners. Recently another beekeeper in Romania installed the SAMS hive monitoring system, as part of the partnership 3 - *Apiculture Technologies and Services.* In comparison to the SAMS project generated data, some data was used from previous FP7 project "ITAPIC".

1.1.7 Expected size of the data

The SAMS DW as data exchange format is using JavaScript Object Notation (JSON) format. It is text based, can be human readable, but not so verbose as XML. Usual incoming requests contain data about a single measurement and are composed of source identifier, measurement timestamp and one or more measured values. In total, SAMS DW successfully received and processed 3 112 679 data-in packages. The average sizes of data-in packages are summarized in the table below.

Table 2 Bee colony data type and size, bytes

Data type	Size*, bytes	Comments
Acoustics	48 587	The package contains 2048 real number values corresponding to different sound frequencies.
Temperature	678	Single measurement: timestamp and value
Humidity	682	Single measurement: timestamp and value
Weight	675	Single measurement: timestamp and value
Combined	960	The package sent by ESP8266 (NodeMCU) devices combines outside temperature and humidity, weight and temperature

		measurements from 3 in-hive sensors, does not contain timestamp.

* provided sizes also include technical payload needed for HTTP protocol, such as Authorization and encoding headers

After receiving and processing incoming data, it is stored in SAMS DW core. Data storage supports data point resolution of 1 minute, meaning that for a given object it can store one measurement per minute.

Table 3 Bee colony data type, documents, points and size

Data type	Documents	Data points	Size, kbytes
Temperature	119 521	1 479 044	14 016
Temperature (daily summary)	7 238	119 514	2 008
Humidity	41 749	529 745	5 140
Weight	61 950	716 007	7 612
Acoustics**	29 364	189 130	4 680 864
Modelling results	1 108	28 157	441

** – single data point consists of 2048 values for corresponding frequencies

Overall SAMS DW database stores additional auxiliary information and metadata. SAMS DW WebApi module mostly stores user node descriptions, source mappings, device configurations, and other metadata. SAMS DW Core is main storage for measurements. Database stats are summarized in table below.

Table 4 Bee colony database stats

	WebApi module	Core module
Collections	7	15
Objects	1 465	3 375 065
Average object size	918 Bytes	2554 Bytes
Data size	1 313 KBytes	8 417 391 KBytes
Storage size	3 436 KBytes	4 787 988 KBytes
Indexes	13	18

Index size	572 KBytes	49 196 KBytes

1.2 SAMS user data

No user-specific (including e-mail address) information is being processed and/ or stored within the SAMS DW. Users access their hive data by using Auth0 service. Moreover, no geolocation services are being used to define the location of the hives.

1.3 Apiculture knowledge base

1.3.1 Purpose of data collection / generation

Contextualizing of local systems (Ethiopia and Indonesia) was done by a literature review compiled with Ethiopian and Indonesian beekeeping experts. The data was collected to create an open-source database, which can be accessed by everyone from around the globe. Further, the database was created in order to be self-sustainable. Interested parties (beekeepers, people along the honey value chain, business experts, bee interested people etc.) are invited to create an own account and to further feed the database with new data. The freely accessible data covers all kind of topics related to beekeeping and SAMS: bee races, bee health, bee management tips, beehive systems, information on funding and licensing, labels and certifications for honeybee businesses and also SAMS project documents like deliverables.

1.3.2 Relation to the objectives of the project

In line with the SAMS bee colony data generation and collection, the developed beekeeping data base (SAMSwiki) shall respond to beekeepers needs and provide immediately advise solution and actions to beekeepers if necessary, to improve their actions and increase efficiency of their activities. Further, information on business development and SAMS-system related content was provided for people along the honey value chain.

1.3.3 Types and formats of data generated / collected

A mixture of quantitative and qualitative textual data was collected from existing scientific and non-scientific publications and from individual expert opinions from Ethiopia and Indonesia as well as surveys and UCD actions.

Table 5 Apiculture knowledge base dataset

Dataset	Type of data	Storage format	Data size	Place of storage	Usability
Beekeeping knowledge	Textual data	Online and pdf		SAMSwiki or data up to 31.12.2019 available in SAMS reports (D5.3)	

Survey results	Textual data	Online and PDF	5.4 MB (market survey)	SAMS website and personal partner storage	available in SAMS reports (D2.3)
UCD interviews	Textual data	Online and PDF		MS Teams and personal storage	Results available in SAMS reports

Why was this data format chosen?

The original idea was to create a numeric bee health and bee management database. But as we had no/not enough real data from beehives in 2018/2019 to develop a reliable bee health and bee management database in this format, we worked out the plan to conduct a large-scale literature research aiming to contextualize the beekeeping sectors in mainly Indonesia and Ethiopia. We also gathered some information for Europe, but as it is no problem to find beekeeping literature in Europe, we focussed on Indonesia and Ethiopia as target countries. In the early stage of our research, we found that so far, beekeeping knowledge from Indonesia and Ethiopia was not collected and made available for the public.

Further, we found that especially for Indonesia almost no literature regarding bees and beekeeping exists. Therefore, we decided to provide the information found within the literature research in a textual database similar to Wikipedia (the SAMSwiki). With this open source, wiki-like approach we encourage interested people from the two target countries and further regions to let the database grow by sharing their beekeeping knowledge with the rest of the world. This kind of knowledge exchange makes the database format, we have chosen, sustainable. In the end the SAMSwiki can still serve as bee health and bee management platform which can relate to the DSS and provide advice to beekeepers.

Do the chosen formats and software enable sharing and long-term validity of data?

Yes.

1.3.4 Utility of the data

The utility of the data serves especially beekeepers, scientists, the economic society, business experts and people along the honey value chain or people who are interested in beekeeping topics or the SAMS project.

1.3.5 Reuse of the data

Are there any existing data or methods that can be reused?

As the SAMSwiki is an open-source knowledge platform, no restrictions occur in relation to the use of the provided data and information.

The literature data for the SAMSwiki was compiled of existing scientific and non-scientific publications.

The data acquired within the UCD process is based on different interview methods and surveys and is also open available with no restrictions.

Is it needed to pay to reuse existing data?

No. The generated data is open source and for free publicly available.

Are there any restrictions on the reuse of third-party data?

There are no restrictions on the reuse of third-party data. The database works similar as Wikipedia. Every interested person can become an author and create articles for the wiki. References to the SAMSwiki or to the original publications, which were used to generate the data for the SAMSwiki, should be given by the authors. Nevertheless, with this kind of approach, it is not possible to retrace, if authors acted accordingly.

Can the data created (which may be derived from third-party data) be shared?

Yes. It is possible to copy the relevant SAMSwiki page URL to share it.

1.3.6 Origin of the data

The data for the SAMSwiki has its origin in existing scientific and non-scientific publications compiled with expert opinions from Ethiopian and Indonesian beekeepers. The data within the SAMSwiki has no classic numeric format, instead the data is based on a literature review and the contextualizing of local systems for Ethiopia and Indonesia. Furthermore, data and information provided on the SAMSwiki was generated within the SAMS project scope.

1.3.7 Expected size of the data

So far (21.12.2020), the SAMSwiki has 13 main chapters and 75 sub-chapters in English. It is planned to fully translate them into Bahasa (Indonesia) and Amharic (Ethiopia) which triples the amount. Besides, the total number of articles comprises of 157. Additionally, 184 pdf files and pictures are found to be accessible through the SAMSwiki equals a total file size of 80.24 MB.

1.4 Project documentation: reports, PR and personnal data

1.4.1 Purpose of data collection / generation

Data collected and generated within the above-mentioned sections (1) SAMS bee colony data, (2) SAMS user data (not processed or stored) and (3) Apiculture knowledge base are partly processed within reports and PR activities.

The purpose of data collected and/ or generated data within that scope of the SAMS project, is to (1) enhances international cooperation, (2) promote and strengthen the apiculture sector worldwide by providing knowledge and tools, (3) supporting management of bee health and bee productivity, (4) to increases production of bee products, (5) creates jobs and (6) establishes knowledge exchange through networks for mutual learning.

In addition to the above mentioned 3 project sections, the following data are generated and collected within the SAMS project for reporting and evidence purpose towards the European Commission as funding institution:

- Data on SAMS Dissemination activities (e.g. events attended, social media posts, audio-visual material from events);

- Personal data (e.g. list of participants of the Steering Committee Meeting, CVs).

1.4.2 Relation to the objectives of the project

The collected and/ or generated project data relates to the overall goal of the SAMS project: enhance international cooperation of ICT and sustainable agriculture between EU and developing countries and ease the use of open source ICT applications in developing countries. Such data collection and report generation allows the SAMS project team and other individual stakeholders, that the concept applied within the project scope can be replicated to other settings and expertise can be exploited to addresses requirements of end-user communities on beekeeping, researchers and third parties like start-ups or SMEs to increases production of bee products, creates jobs (particularly youths/ women), triggers investments and establishes knowledge exchange through networks.

Data collection of personal data is required in terms of proving actions and project staff involvement.

1.4.3 Types and formats of data generated / collected

The nature of the data is both qualitative and quantitative.

Data is collected and generated in the following way within the SAMS project:

- MS Excel, MS Word, MS PowerPoint compatible files;

- Data are requested, stored and transferred via file sharing (e.g. Microsoft Teams);

- statistics, e.g. twitter activities;

- Monitoring actions such as:

  o DisComEx Monitoring Tool;

  o Work Package Monitoring Tool;

  o Capacity activity monitoring;

  o Business related monitoring;

- Documentation sheets, e.g. timesheets;

- Pictures (e.g. jpg format) and recordings (e.g. mp4 format).

1.4.4 Utility of the data

The utility of the data is described in the scope of the SAMS project: (1) enhances international cooperation, (2) promote and strengthen the apiculture sector worldwide by providing knowledge and tools, (3) supporting management of bee health and bee productivity, (4) to increases production of bee products, (5) creates jobs and (6) establishes knowledge exchange through networks for mutual learning. Therefore, the data and information collected and generated is most of all useful to:

- SAMS consortium;

- Researchers;

- Economic society;

- Beekeepers;

- General public;

- Cooperating projects;

- European Commission services and European Agencies;

- EU National Bodies;

- Policy Makers;

Through communication and dissemination activities carried out through the project´s lifetime, the results and project outputs are exploited, and knowledge and data gained are made available to all interested organisations and partners around the world.

2. Fair Data

The term 'FAIR' data is short for saying that all data should be Findable, Accessible, Interoperable and Re-usable. These principles do not necessarily suggest any specific technology, standard, or implementation-solution, but they highlight the main elements that should guide data management.

2.1 Making data findable, including provisions for metadata

Are the data produced and/ or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

Not applicable.

What naming conventions do you follow?

Generally, all text documents follow the same logic: Grant Agreement number, name of the project, number of the deliverable, name of deliverable, status/date (As applicable with the initials of a person who added comments), example: "GA_780755_SAMS_D1.1.SCM 1_draft"

Regarding generated SAMS bee colony data sets it was agreed with developers (UNILV and UNIKAS) that data payload will be sent in a JSON format and each data package contains a sourced and one or more measurements, where each measurement has a timestamp (or measurement interval in seconds) and one or more numeric values. Data flows in more detail are described in D4.1 Report on Data Management. Data related to SAMS Dissemination follow the naming convention of the respective event. Processed personal data are also related to the specific employee or event.

Will search keywords be provided that optimise possibilities for re-use?

SAMS uses the tags in the „metadata", e.g. for putting PDF documents online. Therefore, it added keywords in the properties.

Do you provide clear version numbers?

Not applicable. There are no data versions because each data row is supported by timestamps.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

In SAMS DW there are metadata created describing the hive itself and devices installed in the hive for making measurements. Metadata includes the type (device, hive, hive element, apiary, group, other), name of the object, which is defined by the Data Warehouse user and location which is optional.

Furthermore, as mentioned above SAMS uses keywords to optimise possibilities for the reuse of SAMS collected and generated data and information provided, e.g. in reports.

2.2 Making data openly accessible

Which data produced and/ or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

All data produced and used in the project is openly available and can be accessed via the [SAMS project website](#), the [SAMSwiki](#), the [SAMS github](#) or the [SAMS Data Warehouse](#).

Only the Minutes of Steering Committee Meetings and other project meetings are confidential and therefore only accessible for members of the consortium (including the Commission Services). Generated and collected raw data of beehives is currently mainly used internally/ by the respective beekeeper for interpretation purposes, but if requested the access can be granted to other individuals such as researchers by contacting the Latvia University of Life Sciences and Technologies, Faculty of Information Technologies.

To ensure the sustainability of SAMS results, a partnership agreement was signed with the Latvia University of Life Sciences and Technologies, Faculty of Information Technologies to further allow the hosting of the SAMS Data Warehouse and SAMSwiki after the SAMS project end. This means that all the collected data and linked hardware to the Data Warehouse will continue to operate and users will not lose their linkage to the SAMS data warehouse. Furthermore, third parties will be able to use the collected data for e.g. research purposes.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

Not applicable.

How will the data be made accessible (e.g. by deposition in a repository)?

The data is accessible to the general public on the [SAMS project website](#), [SAMSwiki,](#) the [SAMS Data Warehouse](#) and the repository OpenAIRE.

What methods or software tools are needed to access the data?

There are no special software or methods needed to access and/ or export the data, as the format of exported data is ".csv". It can be loaded/ imported into different software such as spreadsheet type software, like Excel as well as R Studio or "pandas" library ("Python data analysis library"). It mainly depends on the user's preference but as mentioned a commonly used software could be used for that as well. Generated data from beehives are e.g. not stored as audio but as numbers. The reuse of such data is done via web browser, to see the data summary and charts.

Is documentation about the software needed to access the data included?

There is no need to have a specific documentation to access data as no special software is required.

Is it possible to include the relevant software (e.g. in open source code)?

Not applicable.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

SAMS hive data is stored in a database, which is located on a server, placed at the SAMS partners UNILV at Latvia University of Life Sciences and Technologies. There are no restrictions on use/ access. The source code of the SAMS DW and SAMS HIVE system is published in a version control repository service GitHub. A data access committee is not required.

Have you explored appropriate arrangements with the identified repository?

The appropriate arrangements with the identified repository were made from the Latvia University of Life Sciences and Technologies. A partnership agreement was signed with the Latvia University of Life Sciences and Technologies, Faculty of Information Technologies to further allow the hosting of the SAMS Data Warehouse and SAMSwiki after the SAMS project end.

If there are restrictions on use, how will access be provided?

All bee colony data is stored in the [SAMS Data Warehouse](#) and is open available and for free. Interested parties can request collected data e.g. for reuse for scientific research aspects from the Latvia University of Life Sciences and Technologies, Faculty of Information Technologies. Different options are possible afterwards (1) sharing single workspace or (2) provide all existing data.

All respective public available reports and results are stored and provided on the [SAMS project website](#), [SAMS wiki](#) and the repository OpenAIRE.

Is there a need for a data access committee?

Not applicable. A data access committee is not required.

Are there well described conditions for access (i.e. a machine-readable license)?

Not applicable.

How will the identity of the person accessing the data be ascertained?

Not applicable.

2.3 Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

The generated data are interoperable, meaning that data exchange and re-use is possible between members of the SAMS consortium as well as for external stakeholders and individuals. As previously described, there is no special software or methods needed to access and/ or export the data.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

The generated data is already interoperable. Nevertheless, a long-term cooperation with BeeXML is planned to use the standardised data approach for bee colony data exporting and importing as well as for interpretation of the data.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Standard vocabularies are used for all data types in the data set.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Not applicable, as standard vocabularies and ontologies are used.

2.4 Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

The SAMS DW source code and related SAMS hive monitoring system codes will be licensed under MIT license and published in a version control repository service GitHub.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Not applicable.

Are the data produced and/ or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Project data is useable by third parties, during and after project end. No restrictions apply.

How long is it intended that the data remains re-usable?

SAMS project website: up until 4 years after project end, 2024, according to the Grant Agreement, Article 28 – Exploitation of Results.

SAMS wiki: in the scope of the partnership with Latvia University of Life Sciences and Technologies the operation of the SAMSwiki will be at least ensured for 4 years after project end, until 2024.

SAMS Data Warehouse: as partnership with Latvia University of Life Sciences and Technologies is established, there is no strict deadline when the Data Warehouse will end its operation.

OpenAIRE repository: as long as OpenAIRE is operated

Are data quality assurance processes described?

Data quality assurance processes are described in the deliverables.

Further to the FAIR principles, the IDMPs should also address:

Not applicable.

3. Allocation of resources

What are the costs for making data FAIR in your project?

There are no immediate costs anticipated to make the dataset produced FAIR. Costs for data management and data storage are not defined separately. Any costs are covered by the partner's own budgets.

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Any costs are covered from project direct costs.

Who will be responsible for data management in your project?

The coordinator of SAMS, GIZ, is responsible for the overall data management in collaboration with Latvian University of Life Science (UNILV) who are storing most of the data on their servers. All SAMS partners that were collecting and generating data related to project activities, such as events, are responsible for this respective data management. GIZ will not accept any liability at any time for any kind of damage or loss that might occur to anybody from partners related data handling.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

Resources for long term preservation are discussed in the Grant Agreement, Article 36: During implementation of the action and for four years after the period set out in Article 3, the parties must keep any confidential data, documents or other material (in any form) that is identified as confidential at the time it is disclosed ('confidential information'). If a beneficiary requests, the Agency may agree to keep such information confidential for an additional period beyond the initial four years.

In addition, as previously described long-term preservation will be applied via the [SAMSwiki](#) and the SAMS Data Warehouse which contain all relevant and main project results and data. After the project's end it will be managed by partnerships established within the SAMS project.

Further long-term preservation lies in the responsibility of each partner and their country-specific and company-specific legal framework.

4. Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

For the duration of the project, data backups are made automatically by the server. Within the project, sensitive data transfer is secured by coded file transfer.

The developed SAMS monitoring system gathers data only about honeybee colonies and only these data are being stored and processed in the SAMS DW. Such data are temperature, humidity, weight and sound data. When registering a SAMS hive within the DW, beekeepers can also enter voluntary information about the location of the beehive. Location information can be added as a free text in the corresponding field. Nevertheless, no geolocation services are being used to define the location of the hives. No personal data (including e-mail address) of any individual beekeeper is being processed and/ or stored within SAMS DW. Only DW administrators can see the usernames of connected users (where the usernames are requested from Auth0 service and not stored within DW database).

To access the bee colony data stored in the DW, an individual need to log in to the system. This process is being handled by [Auth0](#) service, furthermore, to log in (and during sign up), user is being asked only for a username (e-mail address, in this case; it should be noted that the e-mail address can also be a non-existing one, but in such way user risks of losing access

to the DW, if he forgets his password) and password, or the user can use his existing Google account, if he prefers to do so. By signing up, the user is notified that he agrees to the terms of service and privacy policy of Auth0 (more details about [Auth0 privacy policy](#)). To comply with the EU General Data Protection Regulation, Auth0 provides the option to store data on EU servers. No additional personal (sensitive) information (like name, surname, age, gender, home address etc.) that could lead to an identification of an identifiable person, is required/ asked to be filled in during log in or sign up steps. From a security point of view, since access to the DW Web API's interfaces are protected by Auth0 authentication and authorization service, non-interactive Machine-to-Machine authentication flow is required (this is needed so only the known devices (with specific device IDs whom we trust) could access DW and send data). During such flow Auth0 service provides access token (a credential) that is issued to an authorized device and must be included into each HTTP request to DW endpoints. The access token has an expiration time (for example, 24 hours) and should be eventually renewed by the device. Such tokens are also acquired via SAMS DW. Remote measurement system is sending HTTP POST request to DW data-in interface, including authentication token within request header and JSON formatted data as a body of the request.

Is the data safely stored in certified repositories for long term preservation and curation?

SAMS data are stored in a database, which is located on a server, placed at the SAMS partners UNILV at Latvia University of Life Sciences and Technologies. At the Latvia University of Life Sciences and Technologies, data are stored and managed based on the developed guidelines of the data privacy ([Guidelines on the data privacy in Latvia University of Life Sciences and Technologies](#)). The SAMS DW source code and related SAMS hive monitoring system codes will be licensed under MIT license and published in a version control repository service GitHub.

Furthermore, all SAMS partners must keep any data, documents or other material confidential during the implementation for the project and for at least four years after the period according to Article 36 of the Grant Agreement. Individual data storage regulation of each partner organization/ institute applies in case they remain longer.

Data related to the overall Data Management of GIZ, are safely stored on secured GIZ servers within the GIZ's Data Management System (DMS). Furthermore, all SAMS partner are responsible to ensure that data is stored safely. GIZ will not accept any liability at any time for any kind of damage or loss that might occur to anybody from partners related data handling.

5. Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

There are no ethical or legal issues that can have impact on data sharing. The SAMS partners comply with the ethical principles as set out in Article 34 of the Grant Agreement, which states that all activities must be carried out in compliance with:

- The ethical principles (including the highest standards of research integrity e.g. as set out in the European Code of Conduct for Research Integrity, and including in particular: avoiding fabrication, falsification, plagiarism or other research misconduct);

- The Commission recommendation (EC) No 251/2005 of 11 March 2005 on the European Charter for Researchers and on a Code of Conduct for the Recruitment of Researchers (OJ L 75, 22.03.2005, p. 67), the European Code of Conduct for Research Integrity of ALLEA (All European Academies) and ESF (European Science Foundation) of March 20112;

- Applicable international, EU and national law.

Regarding ethical misuse of data, there is no such possibility, since only temperature inside the bee colony, weight of the colony, temperature and humidity outside the hive and acoustics of the bee colony (frequency spectrum and not the actual audio recording) are being stored in the SAMS DW. By the definition data misuse is the inappropriate use of data. This can be defined as a legal violation or actions that go against a certain corporate policy, and it can be done either intentionally or accidentally. SAMS partners are not responsible for any potential data misuses, falsification and selling. The SAMS DW user is the primary owner of the data and is responsible for any decisions on data dissemination.

As the SAMS project team practiced the User Centered Design (UCD) approach within the research phase, collected form of personal data through user interviews are kept as anonymous. No names, addresses or any other personal data is stored, as the purpose of such data generation lied with the users behaviour, and motivation to ensure system adaption related to needs and requirements overall. Solely pictures of actual persons were kept, with the consent of those persons, for the purpose of documentation of the user research process.

The SAMS project does not involve the use of human participants or personal data in the research of the bee colony data and therefore there is no requirement for ethical review.

Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

Informed consent for data sharing and long-term preservation is included not only in questionnaires but also in all related document and formats dealing with personal data. The SAMS consortium agreed that each consortium member have and store their own forms.

6. Other Issues

Do you make use of other national/ funder/ sectorial/ departmental procedures for data management? If yes, which ones?

This Initial Data Management Plan is the only procedure for data management procedure SAMS is using. For data protection issues, the DSGVO is being considered.

---

[2] https://www.allea.org/wp-content/uploads/2015/07/Code_Conduct_ResearchIntegrity.pdf

7. Further Support in Developing the IDMP

The Research Data Alliance provides a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools. The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include:

- registry of Research Data Repositories;
- some repositories like Zenodo, an OpenAIRE and CERN collaboration allow researchers to deposit both publications and data, while providing tools to link them.

Other useful tools include DMP online and platforms for making individual scientific observations available such as ScienceMatters.

Table 6 History of Changes of IDMP

History of Changes		
Version	**Publication date**	**Changes**
1.0	June 2018	Initial version
1.1	December 2020	Updated version

Project website: www.sams-project.eu

Project Coordinator contact:

Stefanie Schädlich
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
Wielinger Straße 52
82340 Feldafing, Germany
stefanie.schaedlich@giz.de

DISCLAIMER
Neither GIZ nor any other consortium member nor the authors will accept any liability at any time for any kind of damage or loss that might occur to anybody from referring to this document. In addition, neither the European Commission nor the Agencies (or any person acting on their behalf) can be held responsible for the use made of the information provided in this document.