International Partnership on Innovation

SAMS - Smart Apiculture Management Services

Deliverable N° 6.3

Transfer Study on Data Utilization

Work package N° 6 – Transferability and Partnership

Horizon 2020 (H2020-ICT-39-2017)

Project N°780755

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under **grant agreement N° 780755.** The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the EU.

Project information		
Lead partner for the deliverable	Latvia University of Life Sciences and Technologies, Faculty of Information Technologies, Department of Computer Systems.	
Document type	Report	
Dissemination level	Public	
Due date and status of the deliverable	31.08.2020	submitted 31.08.2020
Author(s)	Dr. Aleksejs Zacepins, Dr. Vitalijs Komasilovs, Dr. Armands Kviesis, Olvija Komasilova	
Reviewer(s)	Stefanie Schädlich (GIZ), Konstanza Jochim (GIZ), Magdalena Sperl (GIZ), Kibebew Wakjira (Holeta), Amanda M. (CV. PI)	

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
	Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH (Coordinator)	GIZ	Germany
	University of Kassel	UNIKAS	Germany
	University of Graz (Institute for Biology)	UNIGRA	Austria
	Latvia University of Life Sciences and Technologies	UNILV	Latvia
	ICEADDIS – IT-Consultancy PLC	ICEADDIS	Ethiopia
	Oromia Agricultural Research Institute, Holeta Bee Research Center	HOLETA	Ethiopia
	University Padjadjaran	UNPAD	Indonesia
	Commanditaire Vennootschap (CV.) Primary Indonesia	CV.PI	Indonesia

List of Abbreviations

CPU	Central processing unit
CRC	Cyclic Redundancy Check
CSV	Comma-Separated Values
DSS	Decision Support System
DW	Data Warehouse
GDPR	General Data Protection Regulation
GPIO	General Purpose Input Output
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technologies
INSA	Information Network Security Agency
MCIT	Ministry of Communication and Information Technology
MCU	Microcontroller unit
MIT	Massachusetts Institute of Technology
MoST	Ministry of Science and Technology
POST	Request method supported by HTTP
UI	User Interface

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) In Indonesia, the apiculture sector 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 Centred 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

The transfer study on data utilization addresses issues of data security and ethical data use of bee colony data and examines the usability and benefits of SAMS data for direct users (e.g. beekeepers) and third-party users (stakeholders from the scientific and economic community and the general public).

In order to define the usability of SAMS data, the study defines and analyses different user groups and their needs and examines how these needs can be fulfilled through the use of and access to SAMS data and the SAMS Decision Support System (DSS). Moreover, the study examines the different types of data produced in the SAMS system and forms of data sharing within the SAMS Data Warehouse. SAMS defines its knowledge and data transfer strategy by addressing several points (questions):

- **What should be transferred?** The SAMS DW as universal cloud system which can receive and store various types of data and was developed along a User Centred Design (UCD) approach to meet the needs of its end-users (beekeepers). In the scope of the SAMS project, bee colony data (temperature, weight, acoustics and humidity) is being transferred, stored and analysed to provide maximum usability and benefit for beekeepers

- **To whom should research knowledge and data be transferred?** Beekeepers, stakeholders from the scientific and economic society and the general public were defined as the main users that could benefit the most from research knowledge and acquired data.

- **By whom should research knowledge be transferred?** The SAMS Data Warehouse as the data storage can also act as data distributer, by allowing users to export data sets and share workspaces (created within the SAMS DW) to access real-time bee colony data.

- **How should research knowledge be transferred?** Within the SAMS Data Warehouse workspace sharing mechanism is developed. Workspace sharing can be used to share collected data to the third parties and other interested bodies.

Furthermore, the study addresses mainstreaming and further adaptation options of the Decision Support Service. Questions related to the data security, reliability and ownership as well as intellectual properties are addressed, including current national and international policies and ethical questions of data utilisation (misuse, direct (via HIVE microphone) or indirect (GPS data, moving profiles) surveillance).

After the sustainability analysis of the SAMS data warehouse it can be concluded that further adaptions are possible and this can be done after the SAMS project within the established partnership. It is worth to mention that the SAMS DW source code will be licensed under MIT license and published in a version control repository service

This transfer study is based on scientific research using literature research, and knowledge gained within the SAMS project. Moreover, consultations with Advisory Board members were conducted to find answers to specific questions.

1. SAMS data and status quo of the project

The SAMS Data Warehouse is used to store collected bee colony monitoring data and this data is main subject for transferring. This chapter outlines the functionality and current development status of the operational SAMS Data Warehouse and describes data types produced in the SAMS project. It provides an answer to the defined question “What should be transferred”.

1.1 The SAMS Data Warehouse

The SAMS Data Warehouse is a universal cloud system, which functions like a virtual workspace where beekeepers can store, access and assess their bee-data. Itis a universal system, which can operate with different data inputs, such as direct data transfer from bee colony monitoring devices as well as manual data upload, which has flexible data processing algorithms based on the incoming data sets. The SAMS DW can be connected not only with SAMS monitoring devices, but also with other general available bee colony monitoring systems.

By definition, the SAMS Data Warehouse is like an intermediate layer between data provider systems and data consumer systems or end-users. It is used to help beekeepers run the apiary more effectively based on recommendations and support mechanisms which are utilized by a higher amount of available data and accumulated data interpretation knowledge. The SAMS DW is developed as a cloud-based data storage and processing unit with the capabilities to combine different data sources, like existing systems and available on-apiary generated data.

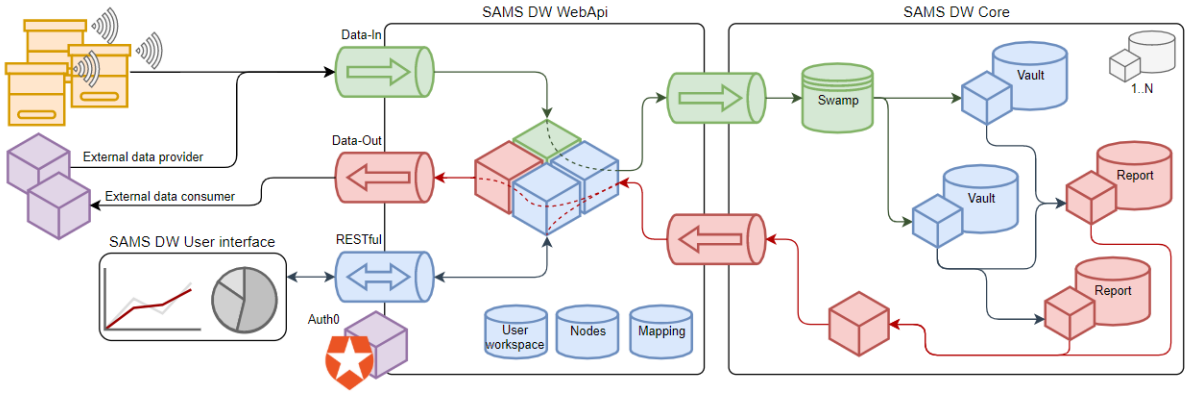

Figure 1. Architecture of the SAMS Data Warehouse

The SAMS Data Warehouse is already developed and is a fully operational platform. It can easily be accessed from any web browser (except outdated ones, like Internet Explorer). A description of the SAMS DW base and its containing parts (as well as their interaction) is provided in the deliverable on Report on Data Management (D4.1.).

The SAMS bee colony monitoring devices, or any other monitoring devices which are available on the market, can directly be connected to the SAMS DW and transfer and store the data online. The research team from the Latvian University of Life Sciences and Technologies (UNILV) has developed a basic instruction on how to send data to the DW (https://sams-project.eu/wp-content/uploads/2020/02/DW-data-sending-guide.pdf). This instruction will ease the process of sending data from any available bee colony monitoring system to the SAMS DW for storage and analysis. Users can easily sign up to enter the SAMS DW using their own

email/ password or Google account. Each user is associated with one or more workspaces, which should be configured based on apiary configuration, with their own connected monitoring systems.

The SAMS DW is an open source platform and the source code will be publicly available after the end of the SAMS project. After the SAMS project end in December 2020 the SAMS DW will continue to operate. It will be maintained and supported by Latvia University of Agriculture Faculty of Information Technologies within the SAMS Partnership 2 (PS2 - International Partnership on Data Management and Utilization). By using the developed instructions and source code, any interested body (start-ups, individuals, non-profit organizations, etc.) will be able to deploy the SAMS DW on their own infrastructure or in a cloud and use all the administrative functions and maintenance options. Since they will be using a new Data Warehouse on a different server used by another party, their Data Warehouse will be empty and without data. Data collected during the SAMS project will not be included in the DW template. Basically, from the available source code, anyone can deploy a fully functioning (empty) SAMS DW without existing bee colony monitoring or any other data, such as location, apiary configuration etc. Only after connecting monitoring devices, data will be collected and stored. A description on how to deploy the SAMS DW is summarised in the deliverable on Evaluation of Responses and Support Services (D4.3) under chapter “Sustainability of the SAMS data warehouse”. This means that every party can have its own data warehouse for different kind of monitoring data (data about other insects, livestock, agricultural productivity etc.), this is not limited to only bee colony monitoring.

The SAMS DW Web API’s interfaces are protected by Auth0 authentication and authorization service. A non-interactive machine-to-machine authentication flow was implemented to allow the monitoring devices to send data.

An important aspect that should be emphasized is that DW user accounts are also managed by Auth0 service and SAMS does not store the user account information or any other private and personal details.

1.2 Data types produced by SAMS

Within the SAMS project remote monitoring of bee colonies is performed by using a specifically developed hardware (SAMS HIVE unit), which can measure several parameters of the bee colony. There are several versions of SAMS bee colony monitoring systems available (low-fidelity prototype, high-fidelity prototype, different versions of ESP based systems) that are described in more detail in the deliverable “Evaluation of HIVE Prototype Designs” (D3.4) and “Manual on HIVE Construction and Operation” (D3.5). Main measurements include weight and temperature of the bee colonies as well as recording of sound of the colony. Moreover, environmental temperature and humidity is measured to gain data about the ambient environment. The system can be equipped with further sensors and numbers highly depends on the sensor type itself. For example, if it is a DS18x20 type sensor (temperature sensor), theoretically there can be up to 70, 80 or more sensors connected (depending on the connection type and wire length), because they use a specifically designed protocol, called 1-Wire®, that uses only one data line. Therefore, only one GPIO (General Purpose Input Output) pin of the MCU/ CPU is being used. Other sensors might use more GPIO's, hence the number of available/ free GPIOs also limits the sensor count.

All collected SAMS bee colony monitoring data is stored in the developed SAMS DW and can solely be accessed by the users who are owning the linked SAMS hardware. Data can be shared between DW users, if the original user provides the access. During the process of creating the sharable workspace link, the original user is notified to agree, that his username and apiary data will be visible to everyone, who will receive the link and use it (whom it will be sent out) (see Figure 2Figure 2. User workspace sharing). The sharable link can be deactivated afterwards, meaning that it won't be active and new persons won't be able to use it. But the one who was supposed to have the link initially (and joined while the link was active), can continue to use the workspace until the original user deactivates it also for this user.

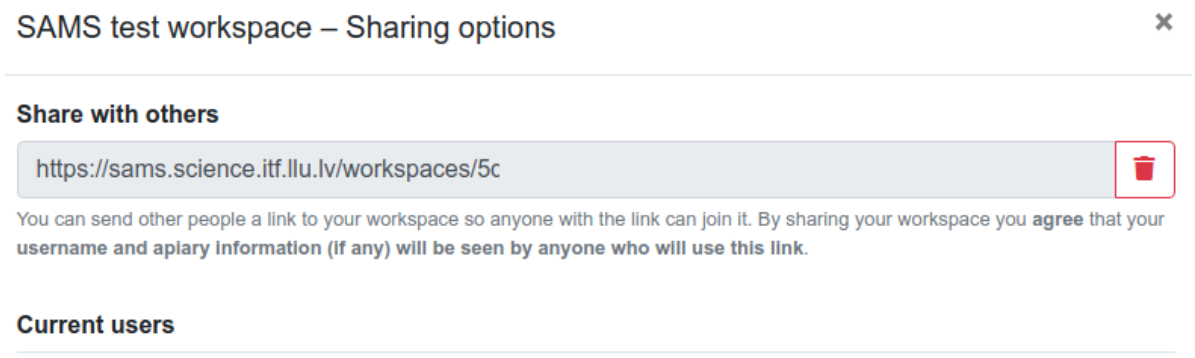

Figure 2. User workspace sharing

As it is part of the concept of the SAMS project to provide all data collected during the project time as open source, a data export feature has been included in the Data Warehouse. This allows for all collected data to be exported to .csv files and to be disseminated freely. The exported .csv file includes a timestamp and the name of the object(s) whose data is being exported (e.g., sensor name, device name). If the object's name includes a region name (depending on how the beekeeper has named its hive, sensors etc.), then the .csv file can contain information about the region as well. But this is up to the beekeeper, if he wants to share such data (he can also edit the column names if necessary). Data can be exported either by system admins or the beekeeper himself. "Outsiders" interested in data should contact system administrators. Any interested person can ask the system administrators for collected data to make further analysis and investigations. As individuals outside the system do not know the amount of data available and the beekeepers whom to contact, the contacts of system administrators can be found at the SAMS DW introduction page.

Data types currently used in SAMS are summarized in the table below:

Table 1. Data types used in SAMS

Parameter	Units	Measuring interval	Sensor	Sensor specifics
Bee colony temperature	°C	20 min by default 2 min for scientific purposes	DS18B20	Precision: +/- 0.5°

Bee colony weight	kg	20 min by default 2 min for scientific purposes	BOSCHE H30A single point load cell	Accuracy class C3, Y=7.500
Bee colony acoustics	Frequencies with its intensity	20 min by default	MEMS SP0645	Operating range 50Hz - 15KHz
Environmental temperature	°C	20 min by default 2 min for scientific purposes	DHT22	Precision: +/- 0.5°
Environmental humidity	% RH	20 min by default 2 min for scientific purposes	DHT22	Precision: +/- 2% RH
Humidity inside the hive	% RH	20 min by default 2 min for scientific purposes	DHT22	Precision: +/- 2% RH

1.3 Status quo of the SAMS DW

The SAMS DW is implemented and is a fully operating platform. It is deployed on a local UNILV server and maintained by faculty administrators. The SAMS hardware and any general available hardware for bee colony monitoring can be linked to the DW for data transfer, storage and analysis.

In addition, other functions (like workspace sharing, data export, apiary templates) have been implemented based on the requests of the target user groups (beekeepers, scientists, general public, etc). The user interface (see Figure 3, Figure 4) of the web version is updated and modified to be more user-friendly. The interface is built to be more in line with the current developed user interface design of the mobile app (more details D4.3. Evaluation of Responses and Support services) and is now more compatible with mobile devices. Similar colour schemes and interface elements (buttons, lines, list elements) are used.

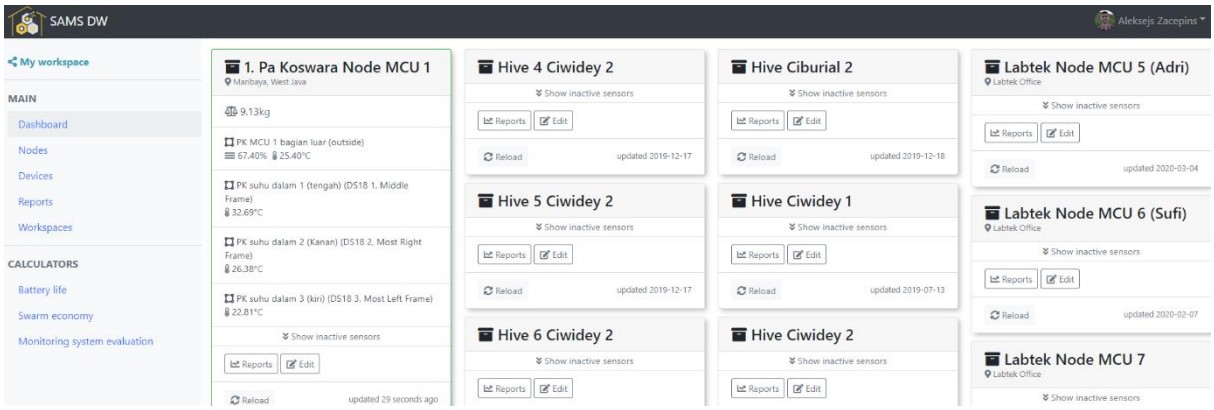

Figure 3. Screenshot of the SAMS DW user interface (desktop view)

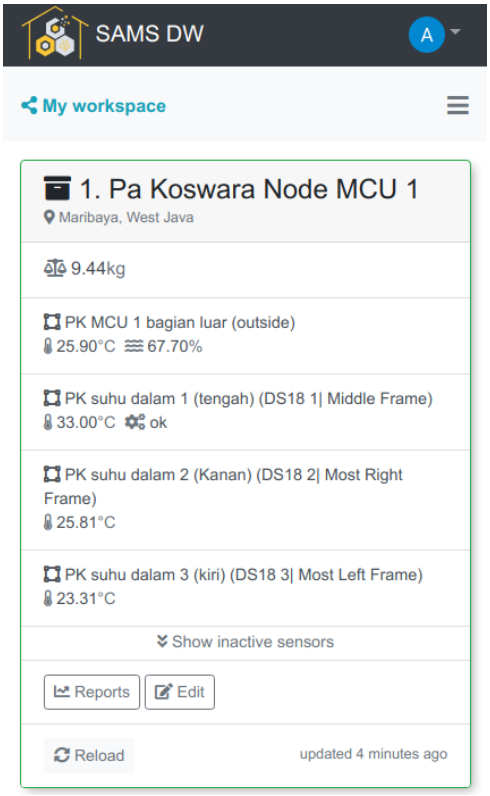

Figure 4. Screenshot of the SAMS DW user interface (mobile view)

Due to the COVID-19 influence and limitations in the implementation of systems, the operation of the SAMS hardware data sets are limited in the DW and not fully representative, thus data analysis stage and implementation of DSS to the DW is on hold. In addition, infrastructural issues in Ethiopia and Indonesia influence the provision of datasets for the SAMS DW (further information will be provided in D3.6. Transfer on Data Utilisation). Potential data analysis algorithms and models are developed and ready for implementation as described in the report Data Analysis and Interpretation in the Decision Support System (D4.2) as well as in the scientific article "Application of fuzzy logic for honey bee colony state detection based on temperature data".

2. Usability and benefits of SAMS data for DW users and third parties

All the data and knowledge gathered during the SAMS project time will be 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. In a best-case scenario, the provision of collected datasets could serve to upscale the knowledge about bee colonies and its behaviour in other countries and regions by providing important insights. After the project end data will not be open-source as default, and can be shared only if data owner (beekeepers) allows it.

Within this chapter analysis of possible SAMS users and third parties who can make use of SAMS data is done.

To make the SAMS results sustainable partnership agreement will be signed with Latvia University of Life Sciences and Technologies, Faculty of Information Technologies and this will allow of hosting of the SAMS data warehouse even after the SAMS project end. This means that all the collected data and linked hardware to the 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.

2.1 Types of SAMS users and use roles within the SAMS DW

Currently there are two user roles in the SAMS DW – administrators and users. Administrators can be considered super users, who have access to the all existing workspaces in the DW. During the SAMS project duration UNILV team members are assigned the administrators rights. After project end, it is not planned to grant anyone the administrator access rights as this role is used mainly for development purposes during SAMS project duration. SAMS DW users, on the other hand, own a specific SAMS bee colony monitoring hardware or other type of general bee colony monitoring hardware which they connect to the SAMS DW. By doing so, the SAMS DW user can store, access and assess their bee-data in the DW workspace and use the DW for the interpretation of measured data and support for decision making processes. This is of particular benefit, as in some cases, beekeepers have sensors to measure specific parameters from their bee hives, but lack a software that allows them to interpret and analyse the collected data. Moreover, the SAMS DW user can chose to share their data with third parties or to keep their data private. The image below (Figure 5) demonstrates the concept of user roles and workspaces (private/ shared) within the SAMS DW.

A third type of users are third parties, for example scientists and the general public. They can receive access to the hardware data of DW users if the owner shares their workspace with them.

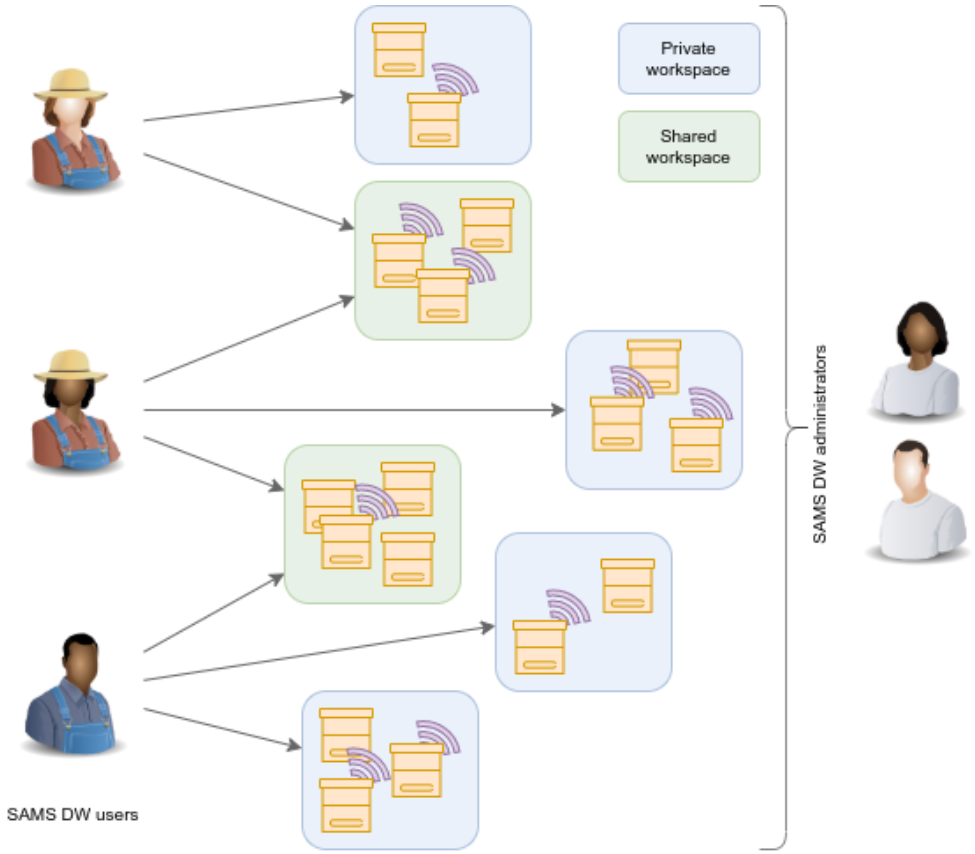

Figure 5. Concept of SAMS user workspaces

2.2 Needs of SAMS users and usability of SAMS data for them

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. It should be mentioned that not every party can use datasets in a raw format. To some users, only analysed data will be of value. These main users were identified based on the applied User Centred Design (UCD) approach, during internal partner discussions and during consultations with SAMS advisory board members. This chapter describes the needs and benefits of SAMS data of the four main target user groups, including the type of data format they can use. Some more detailed information about SAMS data consumers specifically in Indonesia and Ethiopia can be found in a Deliverable 3.6 – Report on Data Communication, where social, cultural aspects, characteristics and usability of SAMS data for identified target groups are summarised.

Table 2. Matching of needs and properties of SAMS data: who can make use of SAMS data?

Properties of SAMS Data	Beekeepers	Scientists	Economic community	General public
Bee colony temperature	X	X		X
Bee colony weight	X	X	X	X

Bee colony status	X			
Environmental parameters	X	X		X

2.2.1 Beekeepers

Beekeepers are considered end-users of the SAMS system and are for sure main users of SAMS data, as the data should support bee-management and shed a light on bee colony behaviour in examined regions. While Precision beekeeping approach is already widely used in Europe, USA and other developed countries it is concluded that In Indonesia and Ethiopia automated and remote bee colony monitoring is only on its starting point and temperature, weight monitoring not widely practised. There are only a few scientific articles on the digital monitoring of Apis Cerana parameters. This can be explained by the lack of knowledge in this field. SAMS data can help to remotely identify different states of bee colonies and detect abnormalities to prevent colony and income losses.

Beekeepers needs

In order to learn more about this user group, the SAMS project team has applied the UCD approach during which, beekeepers were interviewed with the aim to understand their specific needs in the target countries Ethiopia and Indonesia. This approach is especially suitable for system design processes as it leads to an interactive system development that aims to develop systems with optimal usability and usefulness for their users. The UCD approach strongly focuses on the users´ needs and requirements, by considering human factors/ergonomics and conducting intense user research through active involvement of end-users in the development process. It's a multi-step iterative process and includes analysing the context of use, specifying the user requirements, producing design solutions and evaluating them against those user requirements, if possible, with user participation (see Figure 6 below):

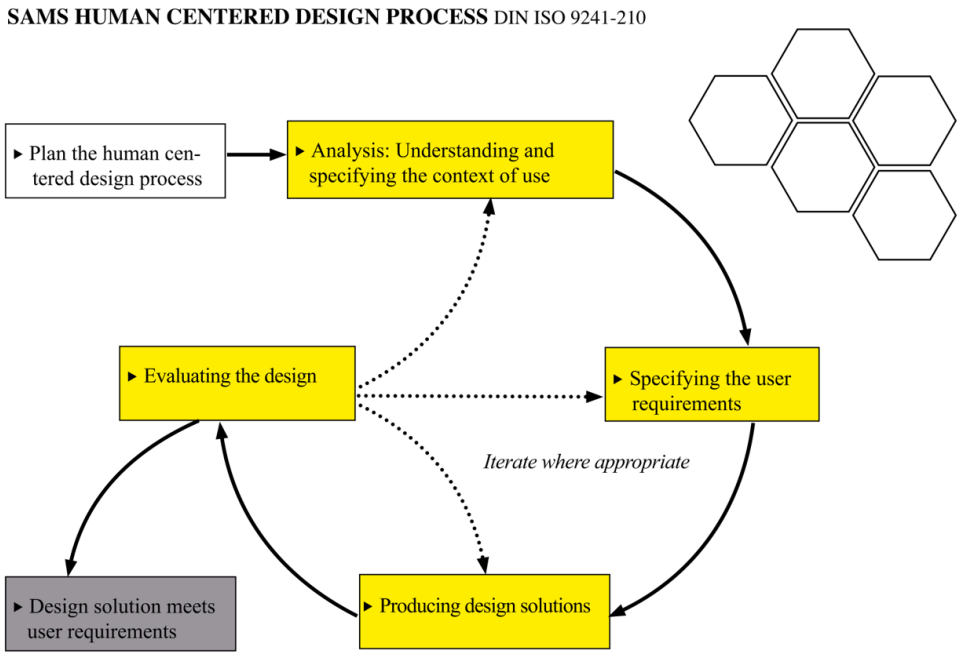

Figure 6. SAMS UCD process

After the conducted in-depth interviews with beekeepers, a list of specific user needs was created and evaluated by the SAMS consortium. The evaluation of those needs was based on the question if an automatic bee colony monitoring system like SAMS can tackle them.

As expected, not all of them are lying in the scope of such system and only the suitable ones were extracted. The full list and analysis of beekeeper needs will be in Deliverable 2.2 User Centered Design Results and Lessons Learnt (coming in October 2020).

Beekeepers needs that are more related to SAMS monitoring system are provided below:

Beekeepers are interested in nectar flow and start-end periods of the foraging and also to be informed about the possible harvesting time.

- This need can be solved by the SAMS monitoring system, as weight monitoring can provide exact insight on foraging activity.

Beekeepers are also willing to have new technologies that can help them to increase the productivity of the apiary thus leading to an increase of the income as well.

- Such a need can be satisfied by applying SAMS technology (monitoring system and DSS), where it can provide information about the current state (observations remotely reduces the number of manual inspections necessary, thus the bees are disturbed less) of the colonies and provide information about the nectar flow (provides information about weight changes, thus the beekeeper can timely plan to add more frames, supers).

Beekeepers are interested in bee behaviour and identification of abnormal situations (e.g., up-coming/ongoing swarming, colony strength) within the colony as well as detect any disease.

- The SAMS system can help with this, as bee colony main parameter (temperature and weight) monitoring together with data analysis can discover moments, when something unusual is happening in the colony and then beekeepers can go to the apiary to investigate the colony on-site.

Beekeepers are very interested in theft prevention, so that they do not have to worry about their colony safety.

- SAMS technology could partly assist in this need, as it can be possible to detect if theft is occurring through by various parameters measured by the system (weight, temperature inside the hive or if the whole colony is being stolen, then the non-transmission of data can be noticed).

Beekeepers want to be notified if it is too hot inside the beehive, so that they can move their hives to a more "shadowy" area.

- The SAMS monitoring system can provide temperature data inside and outside the hive, so this can be used to notify the beekeeper about hot conditions.

Beekeepers are mostly interested in knowing, when a specific flowering season starts, so they can put their hives in locations that can provide enough resources, that will also prevent absconding.

- This cannot be directly solved by SAMS system, but can be done by specific system, which can use information about the plant flowering seasons. The flowering calendar of important melliferous plant species in Java, Indonesia, is summarised in SAMSwiki for Indonesia (https://wiki.sams-project.eu/index.php/Flowering_Calendar_Indonesia) and for Ethiopia (https://wiki.sams-project.eu/index.php/Flowering_Calendar_Ethiopia)

Even if the SAMS system is not directly able to define a solution to all beekeepers´ needs, the SAMS project team acknowledges those as well and designed a different source of information to support beekeepers in understanding their bees and relevant factors that influence the bees: the SAMSwiki. The SAMSwiki is a living database on beekeeping knowledge from different regions around the world. It is continuously growing and developing through an active community of beekeeping experts who feed it with information, review content and share it to a public audience. In the long term it shall complement the SAMS monitoring system and provide additional information to beekeepers.

Usability and benefit of SAMS data for beekeepers

Beekeepers are considered as end-users of the SAMS system and direct users of SAMS data. They can make use of all four types of data which are being monitored in the SAMS system: bee colony temperature, bee colony weight, bee colony status and environmental parameters. In order to guarantee maximum usability of the data for the beekeepers, their needs were strongly considered in the development and design of the Data Warehouse. As most beekeepers cannot read raw data, different data analysis models are planned to be implemented to present and visualize interpretations of the collected bee colony data. Based on the defined models, different bee colony states, such as swarming, absconding, extreme temperatures etc. can be identified remotely. Moreover, beekeepers can save travel time, as the remote monitoring system will help to minimise the number of visual inspections of the colonies. Of course, it is possible for the beekeepers to see all the collected raw data as well if he/she is interested in it.

An important feature of the system is that it can be constantly updated and improved. If any new beekeeper need arises, it is possible to evaluate if it can be implemented, too. During the project time the UNILV team is responsible for the Data Warehouse updates, but after the project this can be done by the parties, who are hosting Data Warehouse on their own. Till the end of the SAMS project instruction on how to host the SAMS DW on different infrastructure will be developed, thus allowing any interested individual to host their own SAMS DW. For now, details of the SAMS data warehouse solution is summarised in the deliverable 4.3 - Evaluation of Responses and Support Services. There four (database, core, Web Api, front-end) main components of the DW are clearly described. It should be mentioned that deployment of the SAMS DW should be done by the IT specialist, as specific knowledge and IT skills are necessary to perform this task and definitely beekeepers will not be able to accomplish this task individually. That’s why within the SAMS new business opportunities are developed and SAMS DW deployment and support is one of such possible business ideas.

2.2.2 Scientists

Data scientists as well as environmental scientists can benefit from SAMS data for research purposes. They are primarily considered as third-party users of SAMS data but can also be considered direct SAMS data users if they connect a monitoring device to the SAMS DW. Within the SAMS project, four main groups of scientists were identified who could be interested in SAMS data: bee scientists, plant specialists and climate scientists. Nevertheless, beside these three groups for sure other scientists can use the collected data. Although this deliverable focuses on the selected groups, the use of SAMS data is not being excluded for other disciplines (e.g. food scientists, climate scientists or social scientists).

Scientists needs

Nowadays there are many scientific groups and individual scientists who are working in the field of Precision Beekeeping (PB). One of the PB tasks is collection of the bee colony related data to make conclusions on the bee colony states, identify bee behaviour issues, evaluate health of the colonies. To make a comparison of achieved results it is often needed to work with different data sets from different colonies from different geographical locations. Thus, collected SAMS data also can be used to facilitate such kind of a research. Scientists often need data of the location and environmental factors that influence the research subject to make better conclusions. SAMS hive monitoring systems can provide the environmental temperature and humidity, as well hive location can be communicated individually.

Within the SAMS project, a survey was conducted from April 29th to July 31st 2019, to learn how scientists of different disciplines could possibly use SAMS data (attached in the Annex 1.). The survey was conducted as part of the projects´ UCD approach and focused on the development of a user centred interface and interaction-design for the SAMS platform. It consisted of 18 questions related to the SAMS bee colony monitoring hardware and software, importance of various sensors and measurements, its intervals etc. In total, 43 scientists from 22 countries, mostly from EU, participated in the survey (see Figure 7). The scientists were from a broad range of disciplines including beekeeping, environmental sciences, agriculture, veterinary, entomology and more:

Table 3. Summary of respondents by the field of research

Fields of research	Researchers
Apidology, apiculture, beekeeping	10
Veterinary, bee health	8
Agriculture, agricultural engineering, farming systems	6
Bees, honey bees, bumble bees,	5
Entomology	4
Genetics	2
Bombiculture	1
Enviromental management	1
Climate change	1
Water resources	1
Policy making in science and technology	1
Marketing, honey marketing	1
Embeded IOT	1

Risk management	1

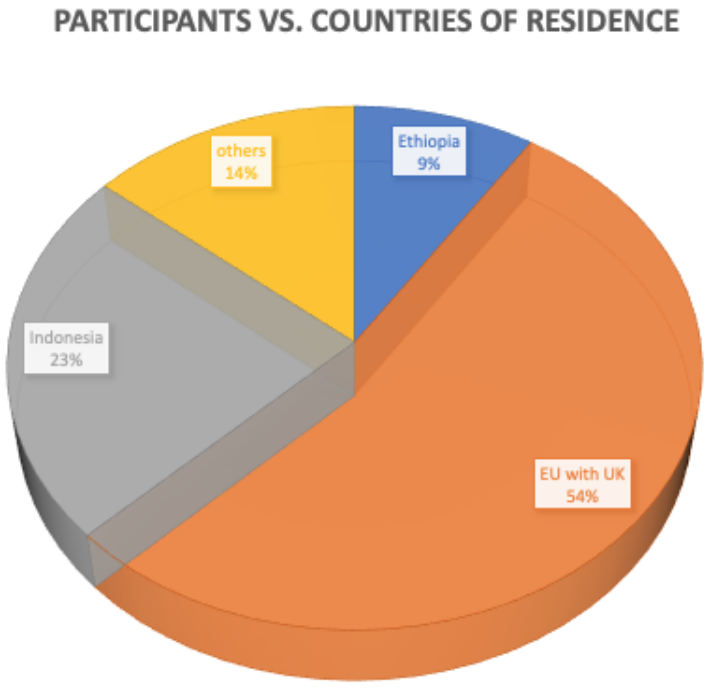

Figure 7. Participants vs countries of residence distribution

One main result of the study was related to data recording frequency. It is widely discussed how often measurements should be taken as requirements often vary depending on the purpose of the data use. For most participants a data recording frequency of 4h (as initially planned for first SAMS systems) was sufficient, but 21% require more frequent data for their work. Some remarks from the participants related to the data recording frequency are:

- Data recording during daytime will be nice if conducted every hour but during the night, it is okay if at 18:00, 0:00 (midnight) and 6:00
- The more the better but it depends on the purpose
- The needed data recording frequency depends on what you want to know and react on
- We would like to see continuous recordings as we are interested to study sound as a proxy for productivity (Nectar collection) of the colony
- Critical factor will be the power supply, if that is no issue a higher frequency may be useful

It is concluded that measurement frequency configuration mainly depends on the question – what is the purpose of the measurements and monitoring of the colony. Summary of possible bee colony states detection and measurement intervals are summarized below:

Table 4. Summary of possible bee colony states detection and measurement intervals

Event or state of the colony	Parameter	Recommended interval	Required interval
Start of the mass nectar flow	Weight	Each 4 hours	1 per day
End of nectar flow	Weight	Each 4 hours	1 per day
Swarming	Temperature	Each 1 min	Each 5 min
	Weight	Each 30 min	Each 1 hour
Broodless	Temperature*	Each 30 min	Each 1 hour
Absconding	Temperature*	Each 1 hour	1 per day
	Weight	Each 1 hour	Each 4 hours
CCD	Temperature*	Each 1 hour	1 per day
	Weight	Each 1 hour	Each 4 hours
Death	Temperature*	Each 1 hour	1 per day

* ambient temperature is also needed

Concerning data representation, survey results showed that even so quite a few scientists preferred to receive data in visualized graph formats (see Figure 8). Almost all of them also need raw or/ and table structured data.

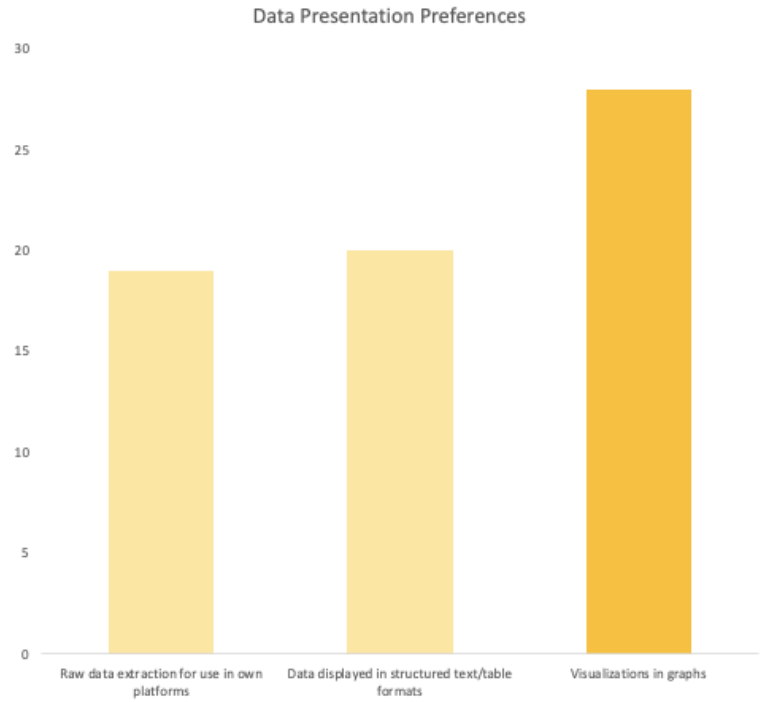

Figure 8. Preferred data representations by number of scientists

Usability and benefit of SAMS data for scientists

Scientists are primarily considered third party users of SAMS data, as they are not directly involved in Beekeeping. Scientist can be considered as skilled in data usage and processing and have an expected high computer literacy rate. Therefore, they can use raw SAMS data to develop their own representation. Focus interest of scientists is to receive data relevant for their individual research which they can read and integrate easily. In this perspective they have quite divers requirements. Apiologist for example are more interested in data of specific bee colonies, a climate researcher wants to analyse data from many beehives of one specific region. Scientists can presently make use of three types of data which are being monitored in the SAMS system: bee colony temperature, bee colony weight and environmental parameters, as well outputs of the models are available for them. Scientists can access the SAMS data contacting SAMS consortium or the beekeepers directly.

The analysis of the survey results showed that data related to measurement of brood hive temperature, environmental temperature, weight of beehive and acoustics is seen as useful and relevant for scientists of different disciplines. 41 of the 43 participants rated the relevance of SAMS sensor data for their work (Figure 9) on a scale from 1 (not of interest at all) to 10 (very high interest):

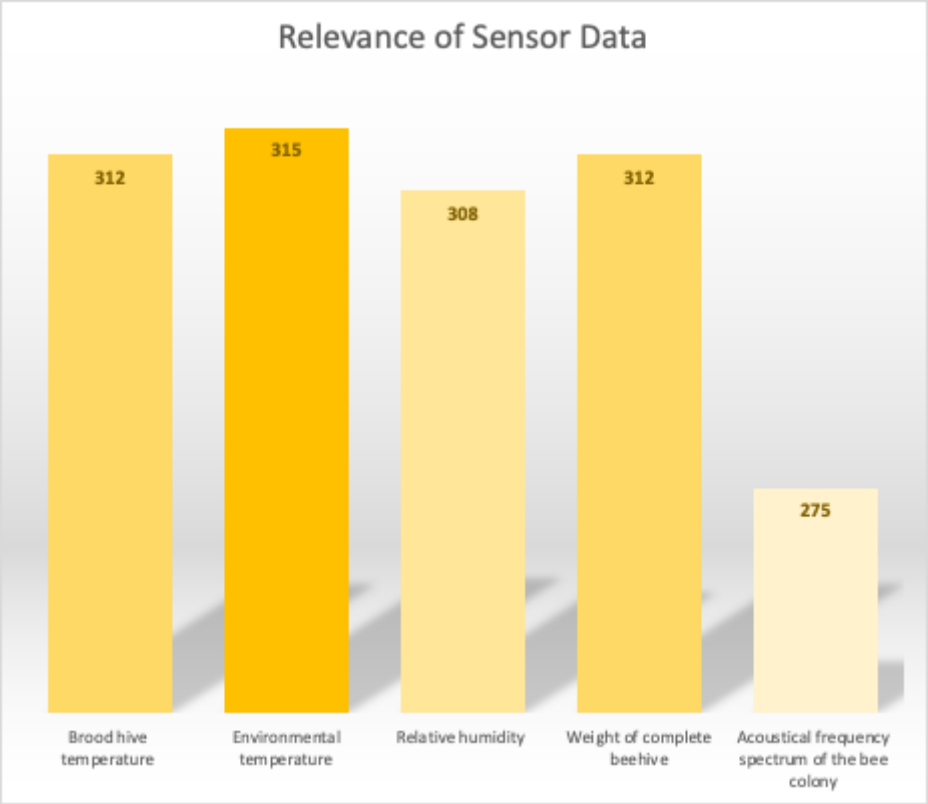

Figure 9. Rating of SAMS sensor data relevance to participants work

Brood hive temperature, environmental temperature and weight of beehive got comparable high ratings, the acoustical frequency spectrum was in general of lower interest. But 19 of the scientists rated acoustical frequency also 8-10. This acknowledges the fact, that temperature and weight are the main parameters of the bee colony, that should be constantly monitored. Environmental temperature should be fixed as well to make a right conclusion of the colony behaviour during various seasons.

Based on the results of the study and further literature research, the following scientific disciplines were identified as main beneficiaries of SAMS data:

- **Bee scientists:** The data will be of use for the scientists, to make an in-deep analysis of the colony behaviour and to compare it with colonies in other regions. This can help to identify the best regions for the bee colony placement. Analysed data can also give some insight of a potential correlation between diseases (known and unknown) and parameters monitored (temperature, humidity). Collected data can help the apiologists to better understand the behaviour of the colony and find the reasons for specific states using the SAMS data.
- **Plant specialists:** Based on weight data of bee colonies it is possible to conclude on flowering periods in different observed regions and evaluate/ differentiate the best honey plants - plants that provide the highest amount of nectar/ pollen. In such way plant specialists can gain new knowledge about plant specifics in different regions, which opens up new opportunities for collaboration with beekeepers/ bee scientists.
- **Climate scientists:** As environmental parameters are also measured, such data will be useful for climate scientists. Due to different placement of bee colonies, e.g. in forests or in places which are difficult to access (e.g. mountain areas in Indonesia) and were no climate stations are deployed, the data could provide valuable information on weather changes and influences on bees and agricultural aspects. Therefore, bee colony monitoring systems can also act as climate stations.

2.2.3 Economic society

Users of the economic society are mainly those stakeholders from the SAMS value chain and the honey value chain which could be potentially interested in SAMS data. These stakeholders include hive makers, apiarists, intermediary collectors, as well as stakeholders from the tourism sector, health sector and government.

With regards to SAMS products, a SAMS value chain was developed within the project to better understand the value of SAMS for the economic society. In addition to honey bee products, it also considers knowledgeable and new beekeepers, hive construction services including supplies (timber, smoker, packaging etc.), and technology (sensors, technical components etc.), particularly sensory hardware and software aspects (including data analysis). An example of a honey value chain mapping[1] that can also be applied to SAMS project context, is shown below (see Figure 10).

[1] Honey Value Chain (2015). URL:http://enpard.ge/en/wp-content/uploads/2015/05/HoneyValueChain-ENG.pdf (access: 12.05.2020)

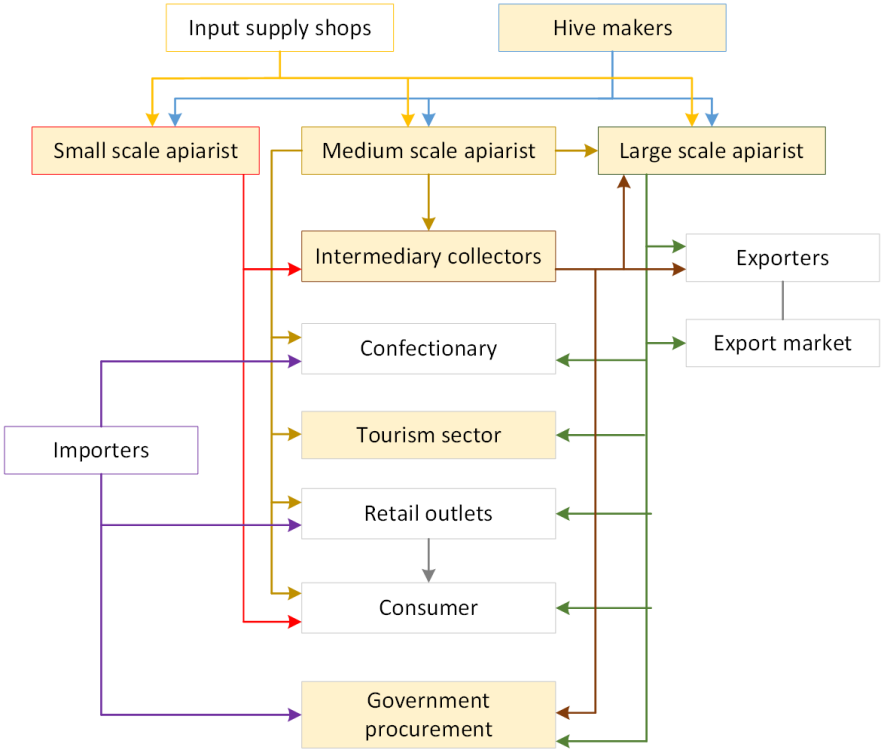

Figure 10. Example of honey value chain mapping[1]

Economical needs

Economical needs are more related to the honey value chain. There are several different stakeholders that can potentially use SAMS data in a direct or indirect way: hive makers, apiarists, stakeholders from the tourism sector and others. In general, ICT (technology and data science) can be considered as an instrument to improve production rate, efficiency which could result in higher income for some sectors. As well reduction of transportation costs can be achieved by using the remote monitoring of the colonies, as only in case of need apiary visits and colony inspections should be done. Using the SAMS data questions like – when will most honey be produced can be answered. This data indirectly could help to adapt the production chains.

Usability and benefit of SAMS data for the economic society

Stakeholders from the economic society are primarily considered as third-party users of SAMS data. They can make use of all types of SAMS data. To access the data contact with warehouse administrators should be established all they can contact beekeepers directly. Beekeepers then can share their workspace for data observation.

Several main groups of the honey value chain stakeholders are considered as potential data users:

- **Hive makers** – are linked to data indirectly, e.g., potentially can adapt hive for convenient sensor placement and technology integration.
- **Apiarists** – the same as beekeepers (described before in point Nr.1).
- **Intermediary collectors** – they can adjust their logistics knowing the dynamics of the bee colony weight data from different apiaries, making prediction of the honey amount. They can also implement a system like “Bee Kanban” to improve processes.

- **Tourism sector** – can propose to visit technologically advanced apiaries to demonstrate emerging technologies and its application for bee colony health monitoring in real environmental conditions.
- **Health sector –** honey quality is not directly related to the technologies used at the apiary, but beekeepers may market their product with the sign, that IT technologies are used and therefore colonies are less disturbed during their life cycle.
- **Government** – data can shed a light on overall bee colony mortality and related agricultural production rates in different regions. Furthermore, use of pesticides can be proven as harmful comparing those apiaries with apiaries from sides were no pesticides are used. This fact does also influence the health aspects linked to the apiculture sector/ products.

2.2.4 General public

For the General public, analysed data could provide insights about the honey production per season. They can be primarily considered as third-party users of SAMS data. Indirectly, application of ICT (monitoring systems and data analysis) in beekeeping (precision beekeeping) can raise awareness to the general public about the modernization in beekeeping and convince them that certain beekeepers might have higher quality standards regarding honey products. Representatives from various sectors can also use bee colony related data, e.g., biology teachers, using already analysed data examples for teaching purposes describing behaviour of bee colonies, identifying bee colony states and abnormalities. As well statisticians, teaching basics of data science e.g., by using raw bee colony data to identify statistically significant dependant variables, correlations.

Public needs

General public would be interested to know the honey production per season. For public, it would be very informative to know which beekeepers are using novel methods, technologies, because this is, in a way, an indicator of quality and production efficiency. Different motivations, behaviours and preferences of honey consumers in Indonesia and Ethiopia have been analysed in deliverable Results of Market Survey (D2.3).

Usability of data for the general public

The general public is considered as third-party users of SAMS data. They can make use of three different types of SAMS data: bee colony temperature, bee colony weight and environmental parameters. Data accessibility approach is the same as it is for the scientists and economic society through the contact to the beekeeper or to the warehouse administrators.

They can use SAMS data about bee colonies for educational purposes, to better understand the behavior and development of the bee colonies. Primarily educators can use SAMS data to better describe different processes that are happening within the bee colony. As well general public can use this data just for self-education, to better understand the bee colony living parameters.

Indirectly, knowing that certain beekeepers use ICT in their beekeeping practices, it is assumed that public might trust more to these beekeepers in regards of the quality of their products.

3. Data Security and ethical questions of data use

This chapter explains how the SAMS project addresses data security, reliability, ownership and ethical aspects of data being collected, used and shared within the project.

3.1 Political awareness on data security: laws and regulations in the EU, Ethiopia and Indonesia

This sub-chapter shortly describes data protection policies in EU and target countries (Indonesia and Ethiopia) and addresses the political support for the ICT. For the EU mainly GDPR is considered as the main reference document.

3.1.1 Data protection in the EU and governmental support for ICT

General Data Protection Regulation (EU GDPR) - is a regulation in EU law on data protection and privacy in the European Union and the European Economic. It applies to personal data. These are “any information which are related to an identified or identifiable natural person”. The GDPR does not apply to data from which you can identify an animal. Examples may include: Human Resources (HR) records, customer lists, contact details, CCTV recordings and computer records.

The GDPR also covers a further category of data, called ‘special category’ or sensitive personal data, which receives greater protection. This category includes data about an individual’s race, ethnicity, religious beliefs, health, trade union membership, sex life or sexual orientation and now also includes genetic data such as DNA. Some of the main principles of the GDPR are:

- Data must be processed lawfully, fairly and in a transparent manner.
- Data must be collected for a specified, explicit and legitimate purpose.
- Data processed must be adequate, relevant and limited to what is necessary.
- Data processed must be accurate and, where necessary, kept up to date.

3.1.2 Data protection in Ethiopia and governmental support for ICT

According to[2] the lack of an institutional framework to provide oversight on the collection and processing of data by public and private bodies is a significant challenge for Ethiopia. Currently, this mandate is partly undertaken by the Ministry of Communication and Information Technology (MCIT), which has taken the **initiative to draft a data protection law**. Nevertheless, the Ministry neither has a full mandate nor the required expertise to undertake this task. Key informants from within the Ministry expressed concerns that the Ministry of Science and Technology (MoST) and Information Network Security Agency (INSA) needed to do more on this issue. Hence, an independent institution with an extensive mandate on data protection and regulation of the concerned stakeholders would be essential.

Further, it is worth noticing that private entities that control and process data, particularly in the ICT sector, are growing in number in recent years. However, their policies and practices on data protection are not subject to any oversight by the government or any of its agencies. The

[2] Privacy and personal data protection in Ethiopia in September 2018

Ministry of Communication and Information Technology, which is partly mandated to undertake this regulation, has very minimal contact with the private sector.

The Ethiopian Communication Authority issues competency certificate for ICT related companies providing the services of manufacturing, wholesaling, retailing, maintenance, importing, exporting and installation of telecom cabling and exchange equipment's, telephone, mobile, computer and related equipment's, Voice and data communication equipment's and spare parts, software, networking, database works, and consultancy service[3].

For more information on Data Guidance for ET, please visit:
https://www.dataguidance.com/notes/ethiopia-data-protection-overview

Ethiopia has several laws that relate to privacy and data security, including the 1995 Constitution of the Federal Democratic Republic of Ethiopia, the 2005 Criminal Code of the Federal Democratic Republic of Ethiopia, the 1960 Civil Code, the Computer Crime Proclamation No. 958/2016 and the Freedom of the Mass Media and Access to Information Proclamation No. 590/2008. Though Ethiopia has not enacted a specific law to address personal data collection and processing issues, the country's scattered legislative framework is understood to require that personal data be collected and processed with due care and only for an intended lawful purpose[4]

3.1.3 Data protection in Indonesia and governmental support for ICT

Data protection in Indonesia regulates the primary regulations governing the protection of personal data in Indonesia. In principle, every person is entitled to the confidentiality of their personal health information that has been provided to or collected by healthcare providers. Financial Services Authority also prohibits financial services providers from disclosing customer data and/or information to third parties, unless there are written consent. Other law, regarding telecommunications prohibits the tapping of information transmitted through telecommunication networks. The Minister of Communication and Informatics is responsible for monitoring and regulating data protection.

The personal data protection regulations do not restrict international data transfers. However, some law requires that the transfer of data overseas be done coordination with the ministry. This coordination entails:

1. Reporting the plan to transfer personal data. At a minimum, the report must include:
    - The name of the destination country,
    - The name of the receiving party,
    - The data of the transfer,
    - The reason/purpose of the transfer.
2. Requesting the assistance of the Minister of Communication and Informatics for the transfer (if required).
3. Reporting the result of the transfer.

However, the Ministry never provided procedures to implement this coordination. Very few business players have, of their own volition, attempted to comply with the coordination

[3] Welcome to Ethiopian Electronic Services, (2020). URL: www.eservices.gov.et (access: 14.08.2020)
[4] https://www.dlapiperdataprotection.com/index.html?t=law&c=ET

obligation. The Ministry has also never sought to enforce the rule, which for the time being is widely disregarded in practice.

Obtaining the consent of data subjects, via consent to a privacy policy, is still the most common way to transfer data abroad. Although there are certain lawful bases aside from consent to process personal data, the application in practice are seldomly seen.

The Indonesian House of Representatives is currently in the process of finalizing the Draft Data Protection Bill, which has been included in the National Legislation Program for 2020. The National Legislation Program is a compilation of the top 50 draft bills that aims to ratify during the current five-year term of the House of Representatives members. However, despite being included in the National Legislation Program, there is no certainty as to when the Draft Bill will be passed into law. President Joko Widodo approved the most recent version of the Draft Bill, which was updated in December 2019, on January 24, 2020. In a recent statement, The Minister of Communication and Informatics, Johnny G. Plate, indicated that the government was “optimistic” that the Draft Bill would be ratified this year.

The current Draft Bill heavily resembles the EU’s General Data Protection Regulation (GDPR). In brief, the Draft Bill is intended to clarify the scope of personal data, the roles and responsibilities of data controllers, data processors and data protection officers, and acknowledges most, if not all, of the rights of data subjects under GDPR, and the general principles on consent to data processing.[5] The elucidation of the Electronic Information Law provides that the protection of personal data is a part of privacy rights, and defines privacy rights to encompass the following: the right to enjoy a private life, free of any disturbance; the right to communicate with other people without any espionage and others (https://www.dataguidance.com/notes/indonesia-data-protection-overview).

3.2 Data collection by SAMS monitoring systems

The developed SAMS monitoring system gathers data only about honey bee colonies and only this data is being stored and processed in the SAMS DW. Such data is 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 (see Figure 11). 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). Nevertheless, it is planned that after SAMS project end there shall be no more administrators of the DW (see chapter 2.1).

[5] Indonesia: Data Protection Laws and Regulations 2020. URL: https://iclg.com/practice-areas/data-protection-laws-and-regulations/indonesia (access 16.08.2020)

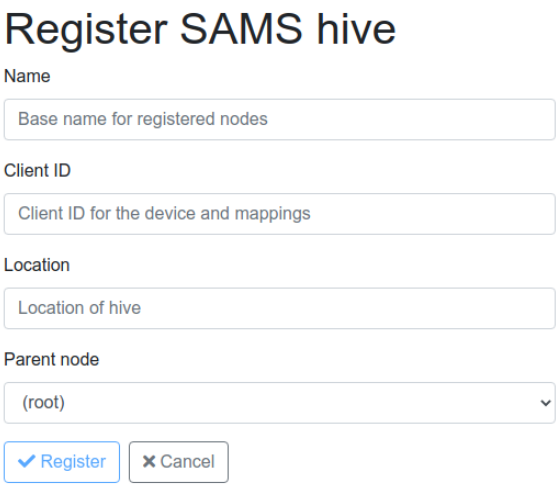

Figure 11. SAMS hive registration within DW

To access the bee colony data stored in the DW, an individual needs to log in to the system. This process is being handled by Auth0 (https://auth0.com/) 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 can be found here: https://auth0.com/privacy). 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.

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. No personal information is gathered within the SAMS bee colony data collection process.

No harm should occur to beekeepers by sharing the bee colony monitoring data, except when real-time bee colony data is shared. This is done together with the location of the apiary (this is solely up to the beekeeper/ owner of the data). In case the location of the bee colonies is somehow made publicly available by the beekeeper himself, there are some threat possibilities, e.g., thefts of the hives or competition for honey resources (e.g., other beekeepers can place their hives alongside). Hence location sharing of the hives together with the data is not foreseen. Nevertheless, for scientific purposes an area of some amount of km² around the hive shall be provided to enable e.g. climate scientists to know to which region/ area data

applies. Users can freely enter any location name to reference the location of their hives – this information can be very specific or simply the name of the country.

Regarding ethical misuse of data, currently there should not be such a 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) is 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.

3.3 Risks related to open source software

The SAMS DW source code will be licensed under MIT license and published in a version control repository service (e.g., GitHub).
Example of the MIT license:

Begin license text.

Copyright <YEAR> <COPYRIGHT HOLDER>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

The software is provided "as is", without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. in no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.

End license text.

Any interested party can clone the source code, adjust and commit new changes and improvements. Committed code changes will not directly be integrated in the master branch instantly, but later after a pull request is issued and reviewed by trusted developers. This also ensures quality control of the official repository.
The published source code will not include any personal data (no e-mail addresses or locations). The source code is what "creates" the system (SAMS DW or SAMS HIVE). In SAMS

DW case, when any interested individual clones the source code, builds and runs it, the DW will be fully operational, but empty - it is like a "shell" without any bee data.
Regarding misuse of SAMS HIVE system or the SAMS DW, it is possible that after the source code is publicly available, an individual can adjust this code for different intentions, like storing microphone recordings instead of only extracting frequency spectrum (but it requires additional programming skills). Since such adjustments are only possible after the project SAMS has ended this issue will be out of the control of the SAMS project team, and SAMS is not responsible for any modifications and adjustments.

3.4 Risks concerning quality and validity of data

Data quality refers to the state of qualitative or quantitative pieces of information. There are many definitions of data quality, but data is generally considered high quality if it is "fit for [its] intended uses in operations, decision making and planning"[6,7]. There are five main key points[8,9] associated with the term “data quality”:

- accuracy – determines if the measured parameter shows the according value, e.g., if the temperature is measured within a beehive, in normal conditions it should never be over 40°C. In the future it is possible to implement more data accuracy rules and check individual parameter values. For example, validate if temperature and weight is within limits. Currently the SAMS monitoring systems uses sensors that have been tested in the lab and corresponds to the given precision, stated in the sensor datasheets.
- consistency - determines if the measured values correspond to the context in which they were acquired. Consistency check is performed manually and it is concluded that within SAMS, data is consistent.
- completeness - determines if all the values within the environment of interest have been accumulated: if there are no extra gaps (missing values) within the measurement series. Unfortunately, it can be concluded that SAMS data does not fulfill this requirement, since there are different issues with SAMS hive monitoring systems (internet network stability, data transfer issues, system implementation problems etc.).
- timeliness - determines if the sensor values have been recorded “within a reasonable time frame”10. Timeliness of data recording procedure was checked manually and it is observed that sensor data is recorded within appropriate time intervals
- validity - In computer science, data validation is the process of ensuring data have undergone data cleansing to ensure they have data quality, that is, that they are both correct and useful. It uses routines, often called "validation rules", "validation constraints", or "check routines", that check for correctness, meaningfulness, and security of data that are input to the system. The rules may be implemented through the automated facilities of a data dictionary, or by the inclusion of explicit application

[6] Redman, Thomas C. (2013). *Data Driven: Profiting from Your Most Important Business Asset*. Harvard Business Press. ISBN 978-1-4221-6364-1.
[7] Fadahunsi, Kayode Philip; Akinlua, James Tosin; O'Connor, Siobhan; Wark, Petra A; Gallagher, Joseph; Carroll, Christopher; Majeed, Azeem; O'Donoghue, John (2019). "Protocol for a systematic review and qualitative synthesis of information quality frameworks in eHealth". *BMJ Open*. **9** (3): e024722.
[8] Loshin D. Characteristics of IoT data quality (2016) URL: https://blogs.sas.com/content/datamanagement/2016/04/27/characteristics-iot-data-quality/ (access: 06.08.2020)
[9] Rachel Levy Sarfin. Data Quality Dimensions: How Do You Measure Up? (2020). URL: https://www.precisely.com/blog/data-quality/data-quality-dimensions-measure (accessed: 08.2020)
[10] Loshin D. Characteristics of IoT data quality (2016) URL: https://blogs.sas.com/content/datamanagement/2016/04/27/characteristics-iot-data-quality/ (access: 06.08.2020)

program validation logic of the computer and its application. At this moment there are several basic data validation algorithms implemented to the data warehouse. At first, it is checked if the data is coming from authorized devices. SAMS DW also checks if measured parameters have values and not only timestamp. It is also checked if the sent data payload is not empty. It must be mentioned, that data validity check is also performed on the device (measurement unit) side: for example, sensors, like DS18B20, provides a CRC (cyclic redundancy check) calculation, to ensure, that data from the sensor has been correctly read. The CRC on the sensor side is considered as a method for data validation.

3.5 Risks for users of the SAMS app

In general, the SAMS collected monitoring data is available only to the owner of the bee colony monitoring system and if it is needed/wished data can be shared with any other users by sharing the warehouse workspace (as described in Chapter 1.2). Secondly, data can be exported from the Data Warehouse in the form of a ".csv" file and then used by others in any way, therefore the user can control what kind of data is the owner willing to share (e.g., data columns from the ".csv" file can be removed by the user, after the data have been exported). Sharing and data export can only be done by the data owner. Currently, in the event that the owner shares the workspace with other users, full access rights are given to the invited user, including configuration of the hive monitoring system, reports etc., but in the future different access rights will be defined. This will secure that system configurations cannot be changed by anyone and enable the data owner to decide which rights, e.g. only reading access, will be provided to the invited user.

Regarding user data, no specific user information is being saved and stored within the SAMS DW. Users access their hive data by using Auth0 service (see information under chapter 3 for more details).

4. Mainstreaming and further adoption options of the DSS

After the project end SAMS DW will be ready for public usage either using the infrastructure and support of the Latvia University of Life Sciences and Technologies or deploying an own Data Warehouse. UNILV will widely disseminate information about the ready solution to the beekeepers and other interested stakeholders. A close collaboration between interested parties is possible to set requirements and evaluate infrastructure availability.

Bringing the DSS into the beekeeping mainstream will require some effort and proper promoting, that could start with the beekeepers currently implementing SAMS system, using it and sharing their experience to fellow beekeepers privately or within social media channels.

Additional workshops or/and trainings would be needed to educate beekeepers and raise awareness about the Precision Beekeeping at all, and remote colony monitoring with data observation in particular.

It is unlikely that private beekeepers will host/use their own data warehouse, most likely that they will use this as a service from any third party.

One more option for the sustainability of the SAMS Data Warehouse and bee colony monitoring system would be achieved through the SAMS partnerships. One of the partnerships

is called "*International Partnership on Data Management and Utilization*" and it focuses on issues of data management and usability as well as mainstreaming of the SAMS Decision Support Service. Within the partnership it is planned to conduct research on bee colony data and to develop an international network that is interested in bee colony data management and within this group different question can be addressed and discussed.

SAMS DW being open source provides the opportunity to any interested individual or company to further adapt it, develop different UI (user interface) or create a mobile app, and update the DSS with additional data analysis rules and models. As well the already made developments and systems could give the opportunities to start new projects for further adaptions and improvements.

SAMS will provide user interface design for the native mobile devices. The design of the front-end is developed mainly by CV.PI in close collaboration with SAMS advisory board member Labtek Indie and all SAMS project partners. This design afterwards can be used to develop fully functioning mobile app for the beekeepers. Based on project content this could be firstly done for Indonesia and Ethiopia and extended to other countries around the world.

An additional further system adoption could be the localization of the content as beekeepers in Asian and African countries mostly will face difficulties with an English user interface.

New models and rules for data analysis still should be developed and adjusted for the SAMS data. The process of model implementation to the Data Warehouse is started and basic example model is implemented, which compares temperature inside the bee colony with environmental temperature. So, the procedure of model implementation is established. This is also due to the fact, that there is still missing information about the bee annual behavior and how this is reflected to the measured parameters inside the hive in regions like Indonesia and Ethiopia.

As indicated in chapter 2 the SAMS Data Warehouse can be adapted to use in other fields for different livestock monitoring (for example pigs, cows, chickens etc.) as it is developed as flexible and modular system.

Annex

Annex 1. Survey for Scientists

SAMS Beehive Data – Survey for Scientists

What is the SAMS Project and Why do we do this Survey

Welcome to our Survey

The SAMS project (Smart Apiculture Management Services) implements sensor technique in beehives of small-scale beekeeping operations in tropical regions.

SAMS allow active monitoring and remote sensing of bee health and beekeeping by developing appropriate ICT solutions supporting management of bee health and bee productivity and a role model for effective international cooperation

A first version of SAMS sensor technique will be implemented in Indonesia (West Java) and Ethiopia this year and will deliver the following data*:

- brood hive temperature from 1 up to 10 temperature sensors
- environmental temperature from 1 outside temperature sensor
- relative humidity from 1 outside sensor for air moisture
- weight of complete beehive
- acoustical frequency spectrum of the bee colony over 4 min (FFT- Fast Fourier Transform)

* recording intervals: 4x daily all parameters

* if connected to power supply system permanent recording is possible

All data will be open source and a decision support system (DSS) that combines sensor-based data-outputs with other information sources and predictive models to measure, analyse and describe different states of the bee colony such as health, vitality, production, etc. will be developed in the project.

Purpose of this survey is to learn how scientists of different disciplines would use SAMS data, in order to accomplish a user centered interface- and interaction-design approach for the DSS.

Please help us and answer 18 questions how SAMS data can be useful for your research. It will take about 15-25 min to complete the questionnaire.

Responsible for the survey: Latvia University of Life Sciences and Technology in cooperation with University of Graz and University of Kassel

Privacy Policy

Thanks for your interest in the SAMS Beehive Data – Survey for Scientists. Purpose of this survey is to learn how scientists of different disciplines would use SAMS data, in order to accomplish a user centered design approach for the DSS. In addition to your opinion, we will also collect some personal information about you, such as your

gender, research field and optional your email address to be able to contact you for further information. The survey is anonymous, to take part in the survey you do not have to enter your name and you do not have to register. We securely store this data until the end of the SAMS project by March 2021. We respect your trust and protect your privacy, and therefore will never sell or share this data with any third parties. The survey will be analysed by the SAMS project partners.You have the right to get more information or delete your personal data. Therefore, please contact

Dr. Aleksejs Zacepins via: samssurvey@llu.lv

Please select:

- I agree and want to take part in the survey.
- I will not continue.

List of question

1. Gender
   - Female
   - Male
   - Diverse
2. Country of Residence
3. What is your general field of research?
4. What are presently your main research activities?
5. Looking again at the list of data SAMS sensors can deliver for a beehive, please rate on a scale from 1 (not of interest at all) to 10 (very high interest) if these data are interesting for your research?

	Not of interest at all									Very high interest
Brood hive temperature from up to 10 temperature sensors.										
Environmental temperature from 1 outside temperature sensor.										
Relative humidity from 1 outside sensor for air moisture.										
Weight of complete beehive.										
Acoustical frequency spectrum of the bee colony over 4 min (FFT- Fast Fourier Transform)										

6. Records Interval

Are the recorded intervals of 4 times daily for all parameter sufficient for your needs?

- Yes
- No

If no, please describe what would work better for you.

7. Please check which type of data delivery is useful for your work.
- Data on Demand
- Data delivery in regular Interval

8. Please explain your selection of data delivery.

9. Please select which type of data presentation is useful for your work.
- Raw data extraction for use in own platforms
- Data displayed in structured text/table formats
- Visualizations in graphs
- Other

10. Please describe, how you would use the mentioned data types.

11. Which algorithms, rules or data analysis models would you need to process SAMS data for your research purposes?

12. Which data sources do you presently use for your research?

13. Which of your data sources could you combine with SAMS data?

14. How satisfied are you with the data sources you are using? Can you give a positive and/or a negative example for a usable and/or useful database?

15. What states of the bee colonies are important for you to detect?
- Swarming
- Brood Rearing
- Quinless
- Nectar flow
- Death
- Bee health status
- Other

Scope of SAMS data

SAMS delivers the following data

- brood hive temperature from up to 10 temperature sensors
- environmental temperature from 1 outside temperature sensor
- relative humidity from 1 outside sensor for air moisture
- weight of complete beehive
- acoustical frequency spectrum of the bee colony over 4 min (FFT- Fast Fourier Transform)

*recording intervals: 4x daily all parameters

*if connected to power supply system permanent recording is possible

16. Can you think of other data from beehives not yet in scope of SAMS that could be useful for you?

17. Do you have ideas how to improve the usability of scientific data sources?

18. Do you want to learn more about the SAMS project? Please leave your email address below. (you can also follow us on Twitter @SAMS_EU_H2020 and on https://sams-project.eu)
- Yes I want to receive more information about SAMS

▪ Yes I agree to participate in a follow interview to discuss my requirements and expectations concerning SAMS data usage in more detail ▪ No
Please leave your Email Address Email addresses of participants will be stored separately from other data. Email: .....
Thank you for completing this questionnaire! We would like to thank you very much for helping us. Your answers were transmitted, you may close the browser window or tab now.

Project website: http://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