

ERA-NET Cofund on Sustainable Crop Production

- SusCrop -



Report on Task 7.3

Workshop on Data Collection and Harmonization

Pirjo Peltonen-Sainio

Natural Resources Institute Finland (Luke)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 771134.



www.suscrop.eu

Task 7.3. Data Collection and Harmonization

Content

Content			
1.	Introduction	2	
2.	Preparatory survey	3	
	2.1.Aim	3	
	2.2.Questionnaire	3	
	2.3.Respondents	5	
	2.4.Main outcomes	6	
3.	The workshop	12	
	3.1.Program and participants	12	
	3.2.Introduction and preparations	14	
	3.3.Workshop outcomes	15	
4.	Summary	16	

1. Introduction

With this workshop, the SusCrop ERA-NET Cofund wanted to offer an opportunity to its partners, stakeholders and project executers to acquire more insight in data collection and harmonisation, with the aim that the project partners would use this opportunity of data sharing in their current SusCrop projects and thereby, meet the FAIR Data Principles.

The FAIR Data Principles consist of a concise and measurable set of principles for data sharing and is already endorsed, e.g., by academia, industry, funding agencies, and scholarly publishers with the intent to act as a guideline for those wishing to enhance the reusability of their data holdings. The FAIR Data Principles put specific emphasis on enhancing the ability to find and reuse the data. FAIR refers to the words: findability, accessibility, interoperability and reusability of data (Fig. 1).

Box 2 The FAIR Guiding Principles				
To be Findable: F1. (meta)data are assigned a globally unique and persistent identifier F2. data are described with rich metadata (defined by R1 below) F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource				
To be Accessible: A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available				
To be Interoperable: I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data				
To be Reusable: R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards				

Figure 1. The FAIR Guiding Principles published by Wilkinson et al. (2016) with the title: The FAIR Guiding Principles for scientific data management and stewardship. Available online at (accessed April 2020): https://www.nature.com/articles/sdata201618.pdf

All the research proposals submitted to any of the SusCrop calls are required to prepare a data management plan, i.e., a tool for performing data management. This plan describes the data that the project executers are expected to acquire or generate during the course of a research project, how it will be managed, described, analysed and stored, and what mechanisms will be used at the end of the project to share and preserve the produced data.

In addition to having the data management plans, a preparatory survey was organised for all the coordinators and participants of the first set of the projects funded by SusCrop (i.e., the first call). The results of the survey served as a background information for planning of the kick-off meeting and the workshop.

The workshop was organised on September 11th, 2019, at the Flanders research institute for agriculture, fisheries and food (ILVO) in Melle, Belgium. It was part of a three days event, set back-to-back with the projects' kick-off meeting and the task 7.4 workshop. This report focusses on the preparatory survey and the workshop activities.

2. Preparatory survey

2.1. Aim

The aim of the preparatory survey has been:

- 1) to collect background information on ideas, views and general understanding of data sharing issues of the SusCrop project executers,
- 2) to identify main opportunities and obstacles for data collection and harmonisation activities within the SusCrop funded projects and
- 3) to compile already existing experiences of the most capable databases suitable for storing and sharing of data produced in the funded projects.

2.2. Questionnaire

The questionnaire on data collection and harmonisation was prepared by Pirjo Peltonen-Sainio, Natural Resources Institute Finland (Luke) in spring 2019. Before sending the questionnaire to the project coordinators and partners it was introduced and discussed a couple of times in SusCrop consortium meetings, in which additional activities were followed up. The final questionnaire was prepared in Webropol. It was semi-structured and contained 12 steps with open questions and those having alternative answer choices to be selected by the respondent. The recipients were informed that in order to be well prepared for the workshop, it is essential that all funded project coordinators and partners give their input. The questionnaire on data collection and harmonisation aimed to compile project executers' ideas about harmonisation of data collection on sustainable cropping to deliver a joint database that can be mined, e.g. to build models and to distil indices of sustainability and resilience. The questionnaire had the following 11 questions that followed the entering of the personal information of the respondent (name, email, project acronym, country, organisation or company):

- 1) To ensure "findability", one of the four elements of FAIR Data Management Principles, we identified three types of database potential for SusCrop (for data acquisition and/or storing):
 - a. Prioritize which database you consider to be particularly valuable for SusCrop and should be further discussed in the workshop to create interlinkages: 1 the most important to 3 the least important
- 2) Are you familiar with any of these or comparable database? 1 yes, very familiar with to 3 not at all familiar with.
- 3) If yes, how did you use them and what was your experience? Free words.
- 4) Can you recognize any likely challenges, limitations or obstacles when benefitting from any of the existing database relevant for your SusCrop funded project? Free words.
- 5) Do you know any other comprehensive database that should be explored for possible interlinkages (like created by FACCE-MACSUR, COST Actions, EU-projects, another networks and initiatives)? Free words.
 - a. Potential database
 - b. Type of data usable for SusCrop
- 6) The first SusCrop call theme is Enhancement of predictive breeding technologies and development of new genotypes leading to new phenotypes and crop varieties for improvement of plant health, protection, production and resilience.
 - a. When considering data collection and harmonization what topics related to this theme should be of high priority and discussed in the workshop. Free words.
 - b. Any foreseen bottlenecks that may challenge data collection and sharing through joint database. Free words.
- 7) The second SusCrop call theme is Development and exploitation of novel integrated pest and crop management methods and practices.

- a. When considering data collection and harmonization what topics related to this theme should be of high priority and discussed in the workshop. Free words.
- b. Any foreseen bottlenecks that may challenge data collection and sharing through joint database. Free words.
- 8) The third SusCrop call theme is Improvement of resource-use efficiency of crops and cropping systems.
 - a. When considering data collection and harmonization what topics related to this theme should be of high priority and discussed in the workshop. Free words.
 - b. Any foreseen bottlenecks that may challenge data collection and sharing through joint database. Free words.
- 9) The fourth SusCrop call theme is Systemic research on agricultural crops as part of an ecosystem including interactions between plants and other organisms ("the plant as a meta-organism").
 - a. When considering data collection and harmonization what topics related to this theme should be of high priority and discussed in the workshop. Free words.
 - b. Any foreseen bottlenecks that may challenge data collection and sharing through joint database. Free words.
- 10) To enable mining of the data for future use, like building models and distilling indices for sustainability and resilience, the data needs to be harmonized.
 - a. What are the critical topics to discuss in the workshop to facilitate successful reuse of SusCrop dataset? Free words.
- 11) Any other comments on data collection and harmonization? Free words.

2.3. Respondents

The questionnaire has been sent out for all the SusCrop project coordinators and partners during spring 2019 by Nikki De Clercq, ILVO, Belgium. Several reminder messages have been sent to non-respondents. In total, 48 project executers responded, which corresponded to ca. 45% response rate. The response rate ranged a lot, from nine to only one project partner per project.

2.4. Main outcomes

The questionnaire started by asking the project executers to prioritise which database they considered to be particularly valuable for SusCrop and to be further discussed in the workshop to ensure "findability" of novel data for any possible future use. Plant phenotyping databases (e.g., <u>https://www.plant-phenotyping.org/dataset-home/</u>) have been considered to be the most relevant databases according to the respondents, followed by plant pests and diseases databases (e.g., https://gd.eppo.int) and databases on remote sensing and earth image analyses (e.g., https://earth.esa.int/web/guest/home) (Fig. 2). 63% of respondents considered plant phenotyping databases to be the most relevant for them, 23% plant pests and diseases databases while only 10% databases of remote sensing and image analyses.



Figure 2. The importance of different databases for SusCrop project executers, when the answer choices were 1=the most important, 2= quite important and 3=the least important.

In total, 39 respondents out of 48 were not at all aware of databases on remote sensing and image analyses, 37 out of 48 on plant pest and diseases database and 35 out of 48 on plant phenotyping database. Only two, two and six respondents were very familiar with them, respectively (Fig. 3).



Figure 3. Majority of the respondents were not at all familiar with any of the requested database. The answer choices were 1 = yes, very familiar with, 2 = quite familiar with and 3 = not at all familiar with.

Despite of the low share of respondents familiar with any of the potential databases, they had the following general ideas (in free words) about the usability of database and related challenges with them:

- Not very practical, too much *Arabidopsis*
- Phenotypic data can be used to classify the genetic diversity and the relationships between morphological data and molecular analysis (genotyping)
- To improve competitiveness of breeding programs
- EPPO database is useful to identify common and scientific name of a specific crop disease
- The European Potato Variety Database to compare differences between varieties for breeding purposes and selecting appropriate cultivars for experiments
- More generic databases are more useful database that link a lot of different data formats and supply a lot of analytical tools tend to be underutilized because designers had a specific set of applications in mind
- Plant and disease database to assess models
- To support crop modelling
- I have trouble with the plant phenotype database

In addition to the three types of databases shown as examples, the respondents were aware of and/or indicated the following potential database for consideration of SusCrop projects:

- www.ebi.ac.uk/ols/ontologies
- <u>www.gramene.org</u>

- <u>https://solgenomics.net</u>
- <u>https://urgi.versailles.inra.fr/gnpis</u>
- <u>www.kenomx.de</u>
- <u>https://triticeaetoolbox.org</u>
- <u>http://germinate.seedsofdiscovery.org</u>
- <u>http://edo.jrc.ec.europa.eu</u>
- <u>http://www.plantgdb.org</u>
- <u>www.wheat-expression.com</u>
- <u>http://snp-seek.irri.org</u>
- MACSUR, WorldClim, PET, NCBI

The project executors who answered the questionnaire indicated several challenges, limitations or obstacles when aiming to benefit from any of the existing databases relevant for SusCrop-funded projects (Fig. 4). The comments (in free words) included the following examples:

- Lack of uniformity
- Still few widely accepted rules, protocols, validation systems
- Data standardization: details, readability, protocols, scoring
- Poorly described protocols
- Ensuring flexibility: implementation of new features required by specific projects
- Ensuring that the information is correctly interpreted
- Limitation in background data (growth/soil conditions, weather)
- Understanding the applicability of the data
- Limitations in time (build competence to use and feed data)
- Limitations in budget (time is money)
- Access not user-friendly
- Ease of use, downloadability
- Overdesign of database
- Raw data as an output, not just maps (pests)
- Computing power
- Ensuring maintenance



Figure 4. The respondents delivered very detailed answers when they were asked to highlight some challenges, limitation and obstacles related to the use of any existing database. The words that have been most frequently mentioned were: access, ease, comparability, uniformity, conditions, interpretations, time, principles and standards.

After general considerations of the potential bottlenecks, the respondents were asked to identify more precisely what are the high priority issues and foreseen obstacles for each of the SusCrop-ERA-NET call themes. The themes of SusCrop's first call for research were: 1. Enhancement of predictive breeding technologies and development of new genotypes leading to new phenotypes and crop varieties for improvement of plant health, protection, production and resilience, 2. Development and exploitation of novel integrated pest and crop management methods and practices, 3. Improvement of resource-use efficiency of crops and cropping systems and 4. Systemic research on agricultural crops as part of an ecosystem including interactions between plants and other organisms ("the plant as a meta-organism"). For the first call theme, the respondents highlighted issues that are shown in Figs 5 and 6.



Figure 5. Word cloud presentation of most repeated words when considering potential obstacles for the first call theme. Corresponding example answers are shown in Fig. 6.

7. The first SusCrop call theme is Enhancement of predictive breeding

technologies and development of new genotypes leading to new phenotypes and crop varieties for improvement of plant health, protection, production and resilience

Number of respondents: 48

When considering data collection and harmonization what topics related to this theme should be of high priority and discussed in the workshop:	Any foreseen bottlenecks that may challenge data collection and sharing through joint database:	
Standardized phenotyping descriptors Methodologies for evaluating resilience at diffeent levels of complexity	Lack of harmonization in data collection.	
marker assisted breeding is a form of predictive breeding technology	breeding information is not easily shared between commercial breeders.	
how to harmonize data from different labs and locations.	unprecise data collection, storage and interpretation	
Collection and analysis of genome data	Accuracy of and cost related to genome data analysis	
harmonization of the assessment of microbiome data	Bioinformatics and IT	
This is not my area of expertise, sorry.	This is not my area of expertise, sorry.	
BARISTA is specifically focused on predictive breeding technology. We use phenotypic data (NUE, disease resistance, yield components, etc), extensive genotypoing data and deep info on specific traits/loci controlling key components for resiliance/sustainability to predict plant performace under expected climatic conditions. Considering the activity carried out in BARISTA, priority should be given to standardization of phenotypic traits (including the evaluation of disease response).	Once a common protocol has been decided a relevant issue is the training of the staffs. SusCrop is working with many research teams across many countries each of them with its own protocols. In my experience the training of a large and disperse community is difficult and to be effective we need to invest resources in training course and in promoting interaction, not at project level that is easy, at the level of students/workers (the staff that really collects the data).	
Should general, stably financed repositories be developed to host data from all crops, or should crop-specific databases be created to host genotype and phenotype data?	IP rights. Poorly designed interfaces. Lack of flexibility and project-specific features.	
Methods to create Open source websites to share coding to analyse genotyping or high throughput phenotyping data. Perhaps instructional videos to assist with this. Certainly standardizing phenotyping methods	Methods and standardized data collection work sheets. Significant could be security of these documents as perceived by firewalls and computer security software in various organizations	
The sharing of simple gene and/or QTL related to the main constrains in the framework of the sustainability and resilience in as many crops as possible should be the main priority	The bottlenecks will be the lack of willingness to share information among European researchers	
database linking genotypes, phenotypes and qtl/associations	no	
How can we share data globally with teh aim to enhance crop breeding and research. The case of genomic selection in animal breeding has shown how useful it is to combine large, global datasets to improve prediction accuracies. Can this be achieved in crop breeding?	intellectual property issues	

Figure 6. Examples of the answers of the first 12 respondents regarding the first call theme.

For the second call theme challenges and/or obstacles were related to, e.g., the following issues:

- Standardised methods to measure pest and disease damages
- Quantification of the level of infection is fundamental
- Impacts of environmental conditions and pathogen strains
- Spatial explicit data on pest outbreaks
- Spatial and temporal resolution
- Difficulties is organizing such complex data
- Rotational data needed, not just a single crop
- Exact identification of pest unsure
- Lack of meteorological data to link with
- Differences in experimental setups
- Pre-publication protection

For the third call theme very similar types of challenges were found as previously shown, but in addition to those the following challenges were reported:

- Similarities with general challenges of previous themes
- Many methods (resource use efficiency) exists and may be difficult to standardize
- Rotational data
- Differences in scope of projects will lead to different priorities on scale (e.g., plant to landscape scale
- The importance of roots: challenges of data collection
- Variety differences between locations and regions
- Differences in experimental arrangements

For the fourth call theme on system level the most frequent answer was that the respondent was not an expert on this subject. When it was asked what are the critical topics to discuss in the workshop to facilitate successful reuse of SusCrop datasets, the respondents shared a lot of ideas (Fig. 7), and thereby, it was apparent that the questionnaire as a whole was a good way to orientate and motivate to project executers for the workshop.

11. To enable mining of the data for future use, like building models and distilling indices for sustainability and resilience, the data needs to be harmonized

Number of respondents: 48

What are the critical topics to discuss in the workshop to facilitate successful reuse of SusCrop datasets:				
Ways of harmonizing data, as it may be difficult that all partners from diferent projects take the measurements in the same way. And even if they will, probably it will not be possible, as the equipments and facilities are not the same everywhere.				
the FAIR principle. understand that meta data are often indispensable to allow reuse of data.				
unprecise data colleaction, storage and interpretation				
Define parameters which need to measured during the timecourse of a field trial in order to precisely describe the environmental conditions. Agree on minimal standards, which need to be fullfilled during a field test, for example, plot size, number of plants, number of replications, parameters evaluated, both regarding test-plants and environmental conditions.				
as many projects will work on microbiome related issues as mentioned I see a need for harmonization of data assessment in this field starting from extraction of nucleic acids up to bioinformatics analysis, data stroage and metadata generation				
data formats, platforms, timing of data access				
The key point is the standardiziation of the data, data coming from different experiments might have been collected with different protocols.				
Formatting, metadata, storage, data visibility, sharing of analysis methods and pipelines.				
Make it open source and straightforward to access.				
The organization of big data could be a large constrain for reusing SusCrop datasets				
visibility of these data (FINDABLE)				
who is going to maintain this database after SusCrop has finished?				
All topics.				
Tools for dissemination within and uptake by the scientific community				
That all data contains geographic references and information on soils, climate and management				
How the data will be stored. Minimum set Agree of the data dictionary Sharing agreements Compatibility in the data management plans of each project.				

Figure 7. Some examples of the answers of the first 16 respondents regarding the critical topics to discuss in the workshop to facilitate successful reuse of SusCrop datasets (free words).

3. The workshop

3.1. Program and participants

In total 45 participants were registered to the workhop on data collection and harmonisation. They had either a role as a SusCrop Consortium member (25), a SusCrop funded project partner (13) or stakeholder (7) and they were in total from 16 countries and represented 37 different organisations (Fig. 8). No participants were present from SusCrop funded projects ROOT, NETFIB and SUSTAG.

The program of the workshop is shown in Figure 9. Johannes Pfeifer opened the workshop on data collection and harmonization and together with Pirjo Peltonen-Sainio introduced the workshop agenda to the participants. To motivate and encourage the audience to data sharing issues Hendrik Poorter, Research Centre JÜLICH, Germany, gave a keynote speech of about 30 minutes where he shared his views, ideas and experiences on need and benefits of having open data shared by the researchers. The title of his presentation was "MetaPhenomics: Analysing the response of plants to their environment by means of data from many small and independent experiments". Thereafter, Pirjo Peltonen-Sainio made a short summary of the main outcomes of the preparatory Webropol-survey. This included mostly information that are shown above, under subheading, "Preparatory survey".



Figure 8. The registered participants of the workshop on data collection and harmonization were from 16 countries and from 10 SusCrop projects: for LegumeGAP there were three project participants, for RYE-SUS two and for the rest of the projects one, except none for ROOT, NETFIB and SUSTAG.

Day 3, September 11 (09:00 - 12:30)

Meeting venue: VAC Gent (21.04 – Jacob van Artevelde – meeting room), Koningin Maria Hendrikaplein 70, 9000 Gent

	Time slot	Activity	Speaker		
	08:30 - 09:00	🗳 ARRIVAL + REGISTRATION			
WORKSHOP 2 "Data collection and harmonisation"					
1.	09:00 – 10:15	 PART 1: Introduction Introduction to the workshop MetaPhenomics: Analysing the response of plants to their environment by means of data from many small and independent experiments. Main highlights from T7.3 Questionnaire 	Johannes Pfeifer (BLE) Pirjo Peltonen-Sainio (Luke) Hendrik Poorter (JUELICH) Pirjo Peltonen-Sainio (Luke)		
• HEALTH BREAK					
2.	10:45 – 11:45	PART 2: Focus groups Discussion between groups of projects, groups according to field of expertise and/or relevant databases	Moderator (+ rapporteur)/group		
3.	11:45 - 12:30	Highlights from each focus group and concluding remarks	TBD by each group		
Ů LUNCH					

Figure 9. The program for the half-day SusCrop task 7.3 workshop on data collection and harmonization.

3.2. Introduction and preparations

Pirjo Peltonen-Sainio gave a general introduction to the workshop arrangements. The basic idea was that the participants were divided into three focus groups (FG) in such a way the FG1 participants were working at genomics and plant breeding and FG2 participant from crop to system scale. In addition to these, FG3 focused on program perspectives, regarding data sharing issues in their group working. The first two FGs consisted mainly of the partners of the funded projects, while SusCrop consortium members and stakeholders participated mainly the third FG (Fig. 10). By this means there were quite equal share of participants in each of the FG. Each FG selected one moderator and the group agreed who was the rapporteur and presented the key-outcomes for the audience.



Figure 10. The general idea how the workshop participants were divided into the three focus groups (FG).

The questions differed dependent on FG. For FG1 and FG2, which were the ones that SusCrop-funded project executers participated (plus some volunteers from SusCrop consortium members and/or stakeholders) the questions to be discussed were as follows:

- 1) Joint data interests? Potential for data sharing between SusCrop projects (what, when, where)? Means to ensure success?
- 2) Ideas about use of existing database? What is needed to be successful (supervision, training, resources)?
- 3) Agreeing next steps to exploit potential for data sharing and reuse?

For FG3 having SusCrop consortium members and stakeholders the questions to be discussed were:

- 1) Specific targets to set by the program for data sharing
- 2) Means to support data collection and harmonization for future use by other projects

3.3. Workshop outcomes

The main responses to the questions of each FG are compiled below. For FG1 and FG2 the main findings were:

- FG1 project partners (on genomics and plant breeding) recognized real opportunities for data sharing and agreed to have joint protocols for data collection, thereby enabling future use of the SusCrop-funded data production
- FG1 partners agreed that they will progress independently (i.e., in a lowbureaucratic way) without need for any organized supervision (provided, e.g., by SusCrop WP- or task-leaders)
- FG2 project partners (from crop to system scale) again found that it was quite impossible to find sufficiently analogy in measurements between projects and hence, it was hard to see the need and opportunities for agreeing about joint protocols for data collection among SusCrop-funded projects
- FG2 partners should in any case make sure that produced data is stored in a way that enables its reuse in the future

One of the main outcomes of the FG3 dialogue was that the SusCrop consortium members and stakeholders were in general concerned when they learned how hard it seemed to be for the project partners to fulfil the requirement of data sharing issues as a part of FAIR Data Principles (Fig. 11). Many comments were addressed on possible means to support the data sharing issues as this requires efforts from project executers likely as an expense of research activities. The comments highlighted during the discussions were, e.g., related to financial support, support given by additional experts and support by available on existing databases. Creating a joint database for SusCrop was also discussed, but it was evident that for such a high variety of projects focusing on very different organisational levels from genomics to agricultural systems cannot benefit from a single uniform database.

FG4: SusCrop consortium members and stakeholders 1. Specific targets to set by the program for data sharing Mertin Thereare. Splind: (EC platform /D) 617 until the de stell them, show on serve HIT MIAPPE. the one landbe fubrew: nestivation and barriers to cleave date? Agreements EMPHASIS BRAPI + Lylvia : Aste which hat they would be willing to offer. Hendrik : are you happy with aggregated hale, or revolted pitting ? possible to poststhate oudstabese? Volaines: & Heulife: must bases and don't's for doudardisation reagle still own their hat and will be asked Hendork -Bigs : Alist will also enonage and promote collaboration. 3 Hille buffe: Ilog and discuss PM Por mid-tern weeting to legre Billi : EMBO represted Lin Lus Ergy from each other Claude: t burnering and boats variables Hendrich: Is there the orectatabase? No us ? Toovidet enough money and resources ? Jolian Pitit = we should support them ! * Pirjo: Experiences from MA(SUR-> manife effort and experi Baiba: use reserve for it. Eldur. Wageningen Luke plearmes: where is the link to sustainability enderglad receards Reulyile : there are different levels of harmonisation Sylving " volad are the benchmarks for alcororing provident built for mu

Figure 11. Some notes from lively conversations in FG3 around settingthe specific program targets in the SusCrop for data sharing.

The summary

A special task 7.3 is devoted to activities in data collection and harmonization in SusCrop with the aim that the project partners use the opportunity of data sharing in their current SusCrop project and thereby, meet the FAIR Data Principles. All the research proposals submitted to the SusCrop calls are required to prepare a data management plan, i.e., a tool for performing data management. To strengthen the commitment on data sharing issues, a preparatory Webropol-survey was organized for all the coordinators and participants of the first set of the projects funded by SusCrop-ERAnet. This served as a background information for planning of the kick-off meeting and workshop, which was organized on September 11th, 2019, at the Flanders research institute for agriculture, fisheries and food (ILVO) in Melle, Belgium. The workshop was part of a three days event, set back-to-back with the projects' kick-off meeting and the task 7.4 workshop. This report focusses on the preparatory survey and the workshop activities. The aim of the preparatory survey (questionnaire sent for all the project coordinators and partners) was to collect background information on ideas, views and general understanding of data sharing issues of the SusCrop project executers, to identify main opportunities and obstacles for data collection and harmonization activities within the funded projects and to compile already existing experiences of the most potential databases suitable for storing and sharing of data produced in the projects. In total 48 project executers responded, which corresponded to ca. 45% response rate. The response rate ranged a lot, from nine to only one project partner depending of the project. In general, the questionnaire proved to be a good way to orientate and motivate the project executers for the workshop and to support workshop and program preparations. In total 45 participants were registered to the workshop on data collection and harmonization. They had either a role as a SusCrop Consortium member (25), a SusCrop funded project partner (13) or stakeholder (7) and they were in total from 16 countries and represented 37 different organizations. No participants were present from projects ROOT, NETFIB and SUSTAG. SusCrop partners from genomics and plant breeding agreed that they will progress independently with data collection and harmonization (i.e., in a low-bureaucratic way) without showing any specific need for any organized supervision. The more heterogeneous group of experts (from crop to cropping system level) was, however, hesitant about the usefulness of data sharing as each of the project have very specific measurements, data etc. not likely relevant for other partners. Nonetheless, also these projects should make sure that produced data is stored in a way that enables its reuse in the future. One of the main outcomes of the conversation between SusCrop consortium members and stakeholders were that they were in general concerned when they learned how hard it seemed to be for some projects/partners to fulfil the requirement of data sharing issues and to meet the FAIR Data Principles. Many comments were addressed on possible means to support these processes as this requires efforts from project executers likely at the expense of research activities. Creating a joint database for SusCrop was also discussed but it was agreed that for such a high variety of projects focusing on very different organizational levels from genomics to agricultural systems cannot benefit from a single database.